UNIVERSIDAD DE GRANADA

Departamento de Ciencias de la Computación e Inteligencia Artificial

Programa de Doctorado en Tecnologías de la información y la Comunicación

Soft Computing y Visión por Ordenador para la Identificación Forense mediante Comparación de Radiografías

Tesis Doctoral

Óscar David Gómez López

Directores

Óscar Cordón García

Óscar Ibáñez Panizo

Pablo Mesejo Santiago

Granada, Enero de 2020

Editor: Universidad de Granada. Tesis Doctorales
Autor: Óscar David Gómez López
ISBN: 978-84-1306-434-5
URI: http://hdl.handle.net/10481/59546

UNIVERSIDAD DE GRANADA

Soft Computing y Visión por Ordenador para la Identificación Forense mediante Comparación de Radiografías

MEMORIA PRESENTADA POR

Óscar David Gómez López

PARA OPTAR AL GRADO DE DOCTOR

Enero de 2020

DIRECTORES

Óscar Cordón García

Óscar Ibáñez Panizo

Pablo Mesejo Santiago

Departamento de Ciencias de la Computación e Inteligencia Artificial

Título en Español: Soft Computing y Visión por Ordenador para la Identificación Forense mediante Comparación de Radiografías.
Título en Inglés: Soft Computing and Computer Vision for Comparative Radiography-based Forensic Identification.
Programa de doctorado: Programa de Doctorado en Tecnologías de la información y la Comunicación.
Doctorando: Óscar David Gómez López.
Directores: Óscar Cordón García, Óscar Ibáñez Panizo y Pablo Mesejo Santiago.

A mi familia, por todo.
アリス、愛してる

Agradecimientos

Al escribir estos agradecimientos siento como que termina un instante que ha durado cuatro años. Un instante que se me escapa entre los dedos y que no soy realmente capaz de comprender. Quizá dentro de unos meses, quizá años, o quizá nunca, pueda entender todos sus matices. Pero ese momento no es ahora. Aunque hay dos cosas que sí tengo claro. La primera que esta etapa ha sido muy especial para mí. La segunda que siento un profundo agradecimiento hacia un montón de personas.

En primer lugar a mis directores de tesis, Óscar Cordón, Óscar Ibáñez y Pablo Mesejo. No solo por guiarme en el camino, sino por enseñarme a construir mi propio camino. Por enseñarme mil cosas, tanto académicas como personales, que me han ayudado a crecer y ver el mundo con una mirada diferente.

A Tzipi Kahana, por ayudarme a dar los primeros pasos en este largo viaje y enseñarme con pasión la importancia de la antropología forense y la comparación de radiografías.

A todos lo que han caminado a mi lado, Rubén, Andrea, Nacho, Enrique, Mari Asun, Juan Fran, Manuel, Carmen, Sergio, Jesús, Elena, Francisco, entre muchos más. También a los que empezaron el mismo camino bajo la brisa del mar hace ya diez años, Cristóbal y José Manuel. Entre todos habéis hecho el camino más divertido.

A mis amigos de toda la vida, María, Pedro Ángel, Juan Manuel, Isa y José David. Gracias por ayudarme a no olvidarme de mí mismo.

Y por supuesto a mi familia. A mis padres, por su apoyo continuo desde antes que tuviera uso de razón. A mis hermanos, por siempre estar ahí. A mis sobrinos, por sacarme una sonrisa cuando no tenía ganas de sonreír. También a los que ya solo viven en nuestros recuerdos e igualmente me siguen dando fuerzas.

Por último a Alicia. Mi mayor compañera en esta aventura, y en las que vendrán, la que me animó a iniciarla y a nunca olvidar que los sueños son más importantes que la realidad.

Contents

Resumen

1. Introdución al problema

La comparación de radiografías (CR) es una técnica de identificación forense basada en la comparación de estructuras esqueléticas, tales como huesos o cavidades, en imágenes radiográficas ante-mortem (AM) y post-mortem (PM) para determinar si pertenecen o no al mismo sujeto. Según las directrices del grupo de trabajo científico de antropología forense para la identificación personal (SWGANTH), CR es una técnica de identificación primaria que cuenta con una alta aceptación en la comunidad forense. Solo por aportar una cifra orientativa a este respecto, cabe la pena mencionar que se emplearon técnicas de CR en 193 identificaciones realizadas por el Laboratorio de Antropología Forense de la Universidad del Estado de Michigan (MSUFAL) entre los años 2002 y 2015. Es importante remarcar que las técnicas de identificación basadas en el esqueleto (SFI, por sus siglas en inglés, *skeleton-based forensic identification*), como CR, pueden suponer la última posibilidad de identificación en muchos escenarios, en donde otras técnicas de identificación (como ADN o huellas dactilares) no son aplicables por el estado de conservación del cadáver o la degradación del tejido blando. En este sentido, el esqueleto generalmente sobrevive a procesos de descomposición natural y no natural, como es habitual en escenarios de desastres masivos. En la literatura relativa a CR, numerosas estructuras óseas, como huesos y cavidades, se han mostrado útiles para realizar una identificación positiva o para *short-listing*, dependiendo de su singularidad. No obstante, todavía se trata de un procedimiento de comparación visual eminentemente visual y basado completamente en las habilidades y experiencia del experto forense. Como consecuencia, la aplicabilidad de la técnica de CR se ve reducida por la subjetividad y elevado tiempo requerido para su utilización. Mientras tanto, un número inabordable y vergonzoso de ciudadanos siguen sin ser identificados durante largos períodos de tiempo debido a la insuficiencia de medios humanos y tecnológicos para identificarlos. Además, los tribunales de justicia demandan la utilización de técnicas reproducibles y objetivas en los peritajes forenses, reduciéndose la aceptación de las técnicas de identificación basadas únicamente en un análisis subjetivo de los datos AM y PM. Por todos estos motivos, hay una necesidad de métodos asistidos por ordenador que ayuden a reducir los tiempos, aumenten la robustez y objetividad, y automaticen el proceso de identificación mediante el método de CR.

Esta tesis doctoral se enfoca en la automatización de la comparación de radiografías AM e imágenes 3D PM. Para su automatización es necesario abordar varias tareas manuales, lentas y tediosas: (1) la segmentación automática de la estructura esquelética bajo estudio; (2) la superposición de las imágenes AM y PM; y (3) la toma de decisiones en base a las superposiciones obtenidas.

2. Desarrollo realizado

En esta tesis doctoral se ha diseñado y validado un nuevo paradigma asistido por ordenador para identificación por medio de CR. Dicho paradigma incluye la automatización de las tres tareas mencionadas mediante el uso de *computer vision* y *soft computing*. En particular, se ha abordado la automatización de todos los procesos involucrados en la obtención de una radiografía PM "simulada", que reproduzca la pose y distorsiones de perspectiva de las radiografías AM. Estas tareas (segmentación y superposición) son el principal inconveniente de los métodos manuales basados en CR. Por tanto, proporcionar una solución automática para estas dos tareas es crucial para una mayor aceptación de las técnicas de CR por parte de la comunidad científica.

La tarea de segmentación ha sido automatizada utilizando redes neuronales convolucionales. Se han desarrollado 2 redes neuronales convolucionales, X-Net+ y RX-Net+, capaces de segmentar cualquier estructura esquelética en radiografías. X-Net+ se enfoca en obtener resultados de alta calidad, pero requiriendo una alta capacidad de cómputo. Por otro lado, RX-Net+ obtiene resultados con una precisión ligeramente mejor, pero requiere significativamente menos recursos computacionales. Estos métodos solamente necesitan alrededor de 200 radiografías para aprender a segmentar una estructura esquelética concreta. Se ha obtenido una precisión similar a expertos humanos en la segmentación de clavículas en radiografías de pecho, y ligeramente inferiores en senos frontales en radiografías craneales.

La tarea de superposición se ha automatizado haciendo uso de algoritmos evolutivos para el registrado de imágenes 3D-2D. Estos métodos buscan reproducir de manera automática los parámetros de adquisición de la radiografía AM en una proyección del modelo 3D PM. Este proceso de registrado será guiado por la silueta de la estructura anatómica utilizada en la radiografía AM (teniendo en cuenta también zonas donde ésta se encuentre ocluida, para identificar) y el modelo 3D PM de dicha estructura anatómica. El problema de optimización subyacente es altamente multimodal, ya que no se puede asumir una inicialización cercana y tampoco se puede depender del valor de intensidad de los píxeles (enfoque tradicional en imagen médica). Además, la evaluación de un escenario de adquisición determinado requiere dos operaciones computacionalmente costosas: la generación de una proyección 2D de la imagen 3D PM bajo un determinado escenario de adquisición; y la comparación de la proyección 2D contra la segmentación de la estructura esquelética en la radiografía AM. Para abordar este complejo y computacionalmente costoso problema de optimización se ha realizado un análisis comparativo de diversos métodos de optimización numérica, así como de diversos algoritmos evolutivos. El mejor optimizador en términos de precisión, robustez y coste computacional es MVMO-SH.

Con el objetivo de validar el método de superposición automático para CR se han segmentado los senos frontales en 180 radiografías y 180 tomografías computarizadas (TACs). Cada radiografía fue comparada contra cada tomografía computarizada produciendo 32.400 comparaciones cruzadas. Los resultados obtenidos han sido analizados utilizando rankings. El método del ranking consiste en ordenar todas las comparaciones realizadas contra una radiografía determinada en función de su error de superposición. Los resultados obtenidos pueden ser considerados prometedores. El caso positivo ocupa el primer lugar (de 180 candidatos, el 0,5% de la muestra

total) en el 50% de las comparaciones cruzadas. Se clasifica en las primeras 10 posiciones en el 80% de las veces (5,5% de la muestra). Finalmente, para alcanzar un nivel de confianza del 100% de éxito, hay que considerar las primeras 50 posiciones (27% de la muestra). En consecuencia, el método actual reduce considerablemente el número de candidatos que deben ser revisados por parte de los expertos forenses, convirtiéndose así en un instrumento útil para la selección de candidatos. Por último, estos resultados se obtienen utilizando una versión preliminar de un sistema de apoyo a la toma de decisiones. Por tanto, el método actual con una versión muy preliminar de un sistema de ayuda a la toma de decisiones (basado únicamente en el valor de la métrica *Masked DICE*), es capaz de filtrar el 73% de los posibles candidatos con una tasa de error 0 de forma completamente automática.

3. Conclusiones y trabajos futuros

En conclusión, en esta tesis doctoral se han automatizado y validado las tareas de segmentación y registrado del proceso de identificación forense mediante CR con resultados prometedores en términos de precisión y robustez. El principal problema del método de registrado es el tiempo requerido para obtener una superposición que, a pesar de haber sido reducido, es aún alto.

Éste es el primer trabajo que afronta la automatización de un sistema de identificación forense mediante CR. Sin embargo, aún queda trabajo por delante antes de que el método propuesto alcance la madurez científica y tecnológica. Por otro lado, con respecto a la tarea de segmentación automática, se planea el estudiar la capacidad de X-Net+ y RX-Net+ para la segmentación de un mayor número de estructuras esqueléticas en distintos tipos de radiografías. Por otro lado, con respecto a la tarea de registrado automático, se planea reducir el tiempo de ejecución mediante el estudio de métodos multi-resolución, funciones de evaluación subrogadas y la utilización de GPUs. Finalmente, con respecto a la tarea de tomas de decisiones, se planea desarrollar y validar de modo completo e integral el sistema jerárquico de toma de decisiones propuesto en esta tesis. Una vez las tres tareas hayan sido completamente automatizadas y validadas independientemente, se planea realizar estudios de fiabilidad del sistema completo utilizando diferentes estructuras esqueléticas.

Abstract

Comparative radiography is a forensic identification technique based on the comparison of the same skeletal structure in ante-mortem and post-mortem radiographic data to determine the identity of a deceased person. In particular, this PhD dissertation focuses on the automation of the comparison of ante-mortem radiographs and 3D post-mortem images (e.g. computed tomographies or 3D surface scans). To automate comparative radiography-based identification, several manual and time-consuming tasks have to be considered: (1) the segmentation of the anatomical structure under study; (2) the superimposition of the ante-mortem and post-mortem data; and (3) a decision making process based on the superimpositions obtained.

In this PhD dissertation, a novel framework has been designed to tackle these tasks using computer vision and soft computing techniques. In particular, we tackle the automation of every process involved in achieving the best superimposition of the ante-mortem and post-mortem images, i.e. the solving of both segmentation and superimposition problems. Providing an automatic solution for these two stages is crucial for a wider acceptance of comparative radiography techniques by the scientific community, since generating post-mortem radiographs is the main drawback of manual approaches, and the reason why some experts recommend to only use comparative radiography-based identification as a last resort.

The segmentation task has been automated using convolutional neural networks. We have developed 2 convolutional neural networks, X-Net+ and RX-Net+, able to segment any object within a radiograph with outstanding results. X-Net+ is focused on yielding a high performance, requiring a significant amount of computer resources, while RX-Net+ yields a slightly lower precision but demanding significantly less resources. These approaches only require 200 radiographs with their respective segmentations to be trained. We get human-competitive performance in the segmentation of clavicles in chest radiographs, and high-quality segmentation results in the challenging segmentation of frontal sinuses in head radiographs.

The superimposition task takes advantage of evolutionary 3D-2D image registration methods. The method searches for the ante-mortem acquisition set-up from the forensic object's silhouette considering occlusions. The underlying optimization problem is highly multimodal since a close initialization cannot be assumed and the image intensities are not reliable or not captured. Furthermore, the evaluation of a particular set-up requires two computationally expensive steps: the generation of a 2D projection of the post-mortem 3D image under a particular set-up, and the comparison of the 2D projection and the ante-mortem 2D segmented radiograph. To tackle this complex and computationally expensive task, a comparative analysis of several numerical optimization methods and real-coded evolutionary algorithms has been performed. The best optimizer is MVMO-SH in terms of precision, robustness and computational cost.

To validate the superimposition method for comparative radiography-based identification, the frontal sinuses were segmented in 180 radiographs and 180 computed tomographies. The results were analyzed using rankings, and then, each radiograph was compared against each computed tomography, resulting in a total of 32.400 crossed comparisons. The ranking method consists of ordering all the comparisons performed against a particular ante-mortem radiograph according to the superimposition error. Promising results have been obtained. The positive case ranks in the first position 50% of times (0.5% of the sample). It ranks in the first 10th positions 80% of the times (5.5% of the sample), and ranks in the first 50th positions 100% of the times (27% of the sample). Consequently, the current framework with a very preliminary version of the decision making stage, based only on the value of the Masked DICE metric, is able to filter out 73% of the possible candidates with 0 error rate in a completely automatic way.

Chapter I

Introduction

"No death, no doom, no anguish can arouse the surpassing despair which flows from a loss of identity." — H.P. Lovecraft

Forensic identification [TB06, MBB14, CC17] is the act of unravelling the identity of an unknown deceased. From missing people to mass disaster victims, every unidentified body represents a denied closure to grieving families and friends. Forensic identification is not only crucial for the grieving ones, but it also resolves serious legal and social predicaments. The "uniqueness" of the human body allows us to determinate the identity even in the event of death with reasonable certainty. Both inherent biological indicators (such as DNA, fingerprints, etc.) and acquired indicators (such as medical implants, dental intervention, trauma, etc) that do not change with time are reliable means in the forensic endeavour [TB06].

According to the Interpool [Int18], the primary and most reliable means of identification are fingerprints, comparative dental analysis, and DNA analysis. Unique serial numbers from medical implants are also reliable identifiers in terms of proving identity. During the last two decades, techniques like DNA or fingerprints have been employed in many identification scenarios. However, the application of these methods fail when there is not enough information available, ante-mortem (AM) or post-mortem (PM), due to the lack of data (second DNA sample) or due to the state of preservation of the corpse (e.g. the soft tissue is degraded or is lost). When primary methods are not able to secure a verifiable identification, secondary methods may provide sufficient information to make identification in selected cases. Secondary methods include several forensic anthropology identification techniques, such as the comparison of skeletal structures in AM and PM medical images, and craniofacial superimposition, as well as the analysis of medical records and pathologies.

The skeleton usually survives both natural and non-natural decomposition processes (fire, salt, water, etc.) and thus skeleton-based forensic identification (SFI) techniques, such as comparative dental analysis (primary), represent the last chance of identification in many cases [KH97]. The traditional SFI procedure is depicted in the Fig. 1. According to the scientific working group for forensic anthropology's guidelines for personal identification [fFAS10], SFI methods for positive identification include the comparison of surgical implants and comparative radiography (CR). Between them, identification using surgical implants is the easiest and most powerful. The method typically involves locating and identifying the manu-

facturer's symbol along with unique serial number from the device. Unfortunately, it can only be employed in the few cases presenting implants. Meanwhile, CR [KH97, Kah09, RLM16] traditionally involves the visual (side-by-side) comparison of AM radiographs (2D) of the suspected deceased and radiographs (2D) of the PM remains that simulate the AM radiographs in scope and projection. The radiographs are compared looking for consistencies and inconsistencies in the skeletal structures (e.g. morphology, trabecular patterns, skeletal anomalies, dental features, pathological and trauma conditions, etc.). Another reason for the importance of CR is their lower price and time required in comparison to DNA, which is a crucial factor in mass disasters identification scenarios.

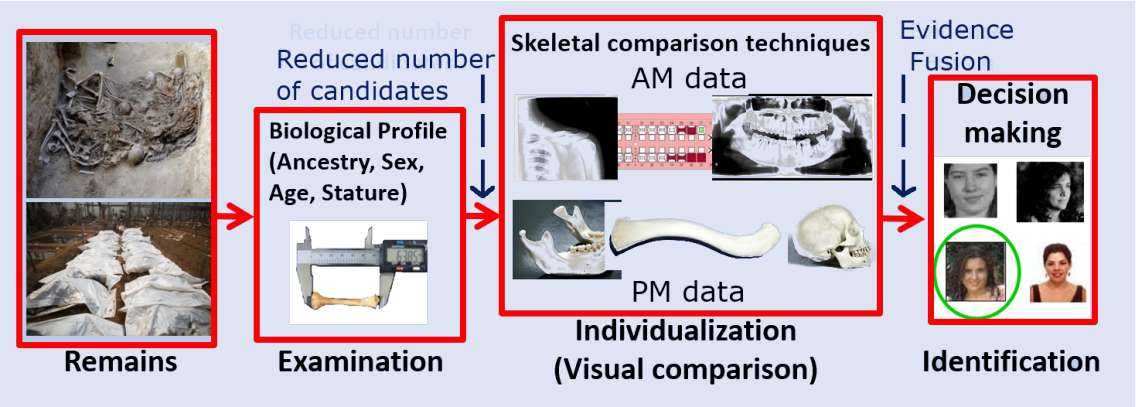

Figure 1: Skeleton-based forensic identification procedure. The usual procedure utilized by forensic experts is the following: (1) a biological profile (sex, age, stature, etc) is obtained based on the PM remains of the deceased; (2) the candidates that do not match the biological profile are discarded; (3) all the possible AM records and medical images of the candidates are gathered; (4) the PM remains are compared to the AM data through skeletal comparison techniques; and (5) an identification decision is taken based on the results of as many available identification techniques as possible.

In the literature, CR is used for identification or just for shortlisting depending on the skeletal structure(s) considered [PTB11]. Several bones and cavities have been reported as useful for positive identification in CR based on their uniqueness [KH97]. The most common ones are located in the skull, chest, and abdominal areas. In the skull, the most frequently used are teeth [Pre01], frontal sinuses [QFS+96], and the cranial vault [MR14]. In the chest and abdominal areas, clavicles [SWCT11], and vertebral features [KGH02] are those most considered. There are also a few bones outside these areas that are commonly used, such as the bones of the hand [KSF05] and the patella [NSGF16].

The advance of medical imaging techniques naturally led to the comparison of skeletal structures using all kind of medical images [TBV02] (e.g. the comparison of a skeletal structure in an AM computed tomography (CT) (3D) and a PM CT (3D) [RBC+16b, GCC+19]). As a consequence, the term CR has no longer "strictly" covered the comparison of radiographs but all possible identification scenarios based on the comparison of skeletal structures. In fact, terms such as comparative radiology or skeletal structure analysis seem more suitable but do not count with the support of the forensic community. Despite of the terminology, in all cases the AM and PM information must be precise and informative. Ideally, both AM and PM images are

of the maximum possible dimensionality, i.e. CTs or 3D scans[1] are preferred over radiographs, since 3D images retain more information about the skeletal structure than 2D images. Specifically, CT images are preferred as these can be rendered to match almost any AM medical imaging examination [HDC+14], but sadly few forensic labs can afford CT scans. Meanwhile, an increasing number of forensic labs are relying on 3D scanners nowadays [DCI+11], thanks to their great availability and relative low cost. However, the availability of 3D AM data is scarce compared to the number of AM radiographs available (especially when people who disappeared a long time ago are involved) limiting the applicability of methods based on the comparison of 3D images. That is especially true in underdeveloped and developing countries, where forensic identification is usually a major need due to a larger number of violent crimes and mass disaster events [Kah09]. As a consequence, most comparisons are performed against AM radiographs. Meanwhile, the PM remains can be scanned with any available acquisition device (e.g. CT scan or 3D scanner). Thus, in practice, the most common identification scenarios are based on the comparison of AM radiographs against PM radiographs or 3D images. In order to quantify the importance of radiographs, it is important to notice that 2.02 million chest radiographs were performed in 2015/16 by the National Health Service of United Kingdom [Eng16] and that 150 million are annually acquired in the United States alone [LLC+18].

In this PhD dissertation, as well as in many forensic works, the term CR refers to any identification method based on comparison of AM radiographs and PM images[2] of the remains, even if the current PhD dissertation is focused on CR based on the comparison of AM radiographs and PM 3D images. The following subsections are devoted to the justification of the work performed, the presentation of the objectives and the structure of this PhD dissertation.

I.1 Justification

The application of a CR procedure is still based on a manual comparison of AM-PM data through a time consuming and error prone visual inspection process that completely relies on the forensic expert's skills and experience. As a consequence, its utility is reduced because of the time required and the errors related to the analyst's fatigue. Meanwhile, an unapproachable and shameful number of citizens continue unidentified for long periods due to the insufficient human and technological means to properly analyze and compare them. In addition, in recent years there has been a shift within the courts of law from analysis of evidence based upon the skill and judgment of the expert witness to one based upon independent judicial assessment of the reliability of a particular methodology, demanding objective and reproducible approaches [Bow01].

There is thus a need of (semi) automatic CR identification methods. The automation of the classical CR scenario based on the comparison of AM and PM radiographs requires the manual acquisition of PM radiographs simulating the acquisition set-up of the AM ones. These PM radiographs are performed by a forensic expert

[1] 3D scans are only possible for the PM image since these require direct access to the skeletal structures without soft tissue in between.

[2] The term "image" is used in its broadest sense including both 2D and 3D images.

within a trial-and-error process. This acquisition process relies completely on the skills and experience of the expert who tries to mimic the available AM radiograph acquisition conditions. However, the lack of information about the X-ray acquisition set-up and precise parameters make this task a subjective and time consuming process. These approaches are based on the comparison of the AM and PM skeletal structure's morphology using geometric morphometrics techniques (such as elliptical Fourier analysis [CBS17]). In particular, these have been employed to compare radiographs of frontal sinus [Chr05b], cranial vault [MR14], and teeth [JC04, NAM05]. Apart from requiring the manual acquisition of a PM radiograph replicating each AM one, all these methods also require the manual segmentation[3] of the skeletal structures in both AM and PM radiographs. Meanwhile, CR methods based on the comparison of AM radiographs and PM 3D images can avoid these drawbacks since they automate the search of the AM acquisition parameters. Consequently, this operation mode allow us to obtain the best possible PM simulated radiograph from a 3D image for each AM radiograph. These methods reduce the subjectivity within the CR process and achieve a greater degree of reliability. However, there are just a few computerized approaches for this scenario. In particular, the existing methods are focused on the comparison of clavicles [SAT$^+$14] and patellae [NSGF16]. Both methods are based on the acquisition of 3D surface models with a 3D laser range scanner of the clavicles/patellae but the final decision still involves a comparison of a silhouette (again, a elliptical Fourier analysis descriptor) in a set of predefined 2D projected images obtained through the 3D model rotation. Furthermore, this also requires the manual segmentation of the skeletal structure's silhouette in the AM radiograph.

In summary, to cope with an automatic identification system based on CR via the comparison of AM radiographs and PM 3D images, the following challenges/issues should be taken into account:

- The skeletal structure's silhouette should be automatically segmented in AM radiographs (see Chapter V).
- The acquisition parameters of the AM radiographs should be automatically calculated/searched. Some authors recommend to only use CR techniques as a last resource because of this problem [ARG$^+$10], thus solving it is of crucial importance for a wider acceptance of the CR techniques by the scientific community (see Chapter VI).
- The projective transformation underlying any kind of radiograph (e.g. regular and angled posteroanterior radiographs, lateral radiographs, etc.) has to be reproduced (see Chapters VI and VII).
- The intensity level depicted could have changed between the AM images and the PM images since the time of the AM radiograph acquisition. The bone density changes within the individual through time due to factors as aging [RSFI15], osteoporosis, and the PM interval [BSJ96]. Besides, its representation in an image is sensitive to the acquisition device, and some old radiographs show a low quality. Therefore, automatic superimposition methods should rely on other features, such as the skeletal structure's silhouette (see Chapter VI).

[3]Segmentation consists of partitioning an image into regions (i.e. sets of pixels) [PXP00], each of them with a different semantic meaning (e.g. segmenting a frontal sinuses in a skull radiograph).

- PM 3D images of skeletal structures obtained using a 3D scanner do not have intensity information. This is of special importance because an increasing number of forensic labs are relying on these scanners nowadays [DCI+11], thanks to their great availability and low cost, in comparison to CT scans that are only affordable by a small number of forensic labs (see Chapter VI).

- The proposed methods should be efficient and scalable. Currently, CR is utilized mostly in verification scenarios (one-to-one comparison) [JH96]. Verification consists of comparing the AM and PM data to determine whether they belong to the same person or not. It is a 1:1 comparison problem. Meanwhile, identification involves exploring a database of AM data for finding to whom the PM data belong. It is a 1:n comparison problem, where n is the number of AM cases in a database (although some AM cases can be directly discarded based on criteria such as the biological profile). The applications of CR to identification scenarios [JLK06] are strongly limited by the size of the database, since an identification process usually takes, at least, n times longer than a verification. Thus, the efficiency and scability of the methods are crucial to enable the utilization of CR in identification scenarios with a significant number of possible candidates (see Chapter VII).

- All methods should be utilizable and accurate for CR identification based on any skeletal structure (frontal sinuses, clavicles, patellae, etc.) (see Chapter VI).

I.2 Objectives

The main objective of this PhD dissertation is the **development and validation of a novel computer-aided automatic framework for CR-based forensic identification**. It will automatically compare the available AM and PM images of skeletal structures and support the expert in the decision making process in an objective, fast, robust and reproducible manner. Computer vision and soft computing techniques will be employed to provide the theoretical and practical means to achieve this objective. This main objective can be divided into the following research lines or subobjectives:

- **Automatic skeletal structure segmentation**: The goal is to develop a common image segmentation (IS) framework for any skeletal structure in radiographic images. To that end, deep learning techniques [GBC16], and in particular convolutional neural networks (CNNs), will be employed. A neural network will be trained for segmenting each particular skeletal structure. Unlike other approaches, CNNs are expected to enable the automatic approximation of complex, non-linear, data-driven functions with minimal human intervention, avoiding or minimizing the deficiencies (in robustness and accuracy) offered by other methods.

- **Automatic skeletal structure superimposition**: The objective is to develop a novel framework for the superimposition of PM 3D images and AM radiographs of a skeletal structure. It will be based on a 3D-2D image registration (IR) paradigm [MTLP12, OT14] which will automatically, objectively

and precisely search for the AM acquisition parameters, applying them to the 3D image, and analyzing the match.

- **Validation and reliability studies**: In order to fulfill Daubert's rules for evidence admissibility in court testimony [FH99], objective validation and reliability studies are required. Thus, an important goal of this PhD dissertation is to objectively validate the developed methods over a significant number of cross-comparisons using real cases. At the same time, these validation studies will serve as a measurement of the accuracy of the methods.

The resulting computer vision and soft computing methods are expected to serve to reduce the intra/inter observer variability thanks to more objective and repeatable measurements and analysis. They will facilitate to address the worldwide human identification challenge in a more effective, cheaper and faster manner while accomplishing the necessary standards (Daubert's rules for evidence admissibility) for the acceptance of forensic evidence in court testimony.

I.3 Structure

This PhD dissertation is divided into three parts (fundamentals, proposals, and final remarks) apart from the current introduction. The first part, fundamentals, is composed of two chapters. Chapter II reviews the basics, the history, and the state-of the art of CR-based methods. Chapter III reviews the theoretical backgrounds of the computer vision and soft computing techniques utilized in this PhD dissertation to automate CR-based identification processes. Meanwhile, the second part, proposals, is formed by four chapters. Chapter IV introduces a novel computer-aided automatic framework for CR-based forensic identification. The contributions of this PhD dissertation to the design and implementation of this framework is the development of all processes involved in achieving the best superimposition of the AM and PM images: segmentation (Chapter V) and generation of the PM radiographs (Chapters VI and VII). Chapter V develops and validates an IS framework for segmenting skeletal structures in radiographs. Chapter VI develops and validates an IR framework for generating PM radiographs simulating the acquisition set-up of the AM-ones. Chapter VI performs a detailed comparative analysis of several state-of-the-art real coded evolutionary algorithms for improving the IR framework in term of robustness and computational time. Lastly, the third part, final remarks, is comprised by only one chapter, Chapter VIII, devoted to present a summary of the conclusions obtained and a discussion about possible future developments.

Part I

Fundamentals

Chapter II

Comparative radiography

"El que lee mucho y anda mucho, va mucho y sabe mucho." — Miguel de Cervantes

This chapter is devoted to review the basis of the forensic identification using CR, since the main goal of the dissertation lies in the development of new approaches to automatically tackle this task with the help of computational resources. Particularly, it will focus on: (1) most frequent types of images employed in the CR identification endeavour; (2) skeletal structures utilized for identification in the CR literature; and (3) identification procedure for performing a CR-based identification using images of skeletal structures.

II.1 Imaging techniques in comparative radiography

This section is devoted to a brief review of the kind of AM and PM images that are commonly employed for identification with the CR technique [CSG+18, HDC+14, TBD03]: radiographs, CT images, and 3D surface models. We will focus on the most relevant information with regard to the CR technique. There are many other imaging techniques (e.g. magnetic resonance images [OW08, TYS+03]) but their use for CR identification is scarce in comparison with the previous ones. Furthermore, some of these are not recommended for the study of skeletal structures (e.g. sonographies [FSF16], for their limited penetrative power that reduces the image quality in deeper skeletal structures).

II.1.1 Radiographs

Radiographs were the first images of modern medicine and opened up a new world of possibilities to all fields of medicine (see Fig. 2a). A radiograph is produced by the action of X-rays on an image receptor (see Fig. 2b for a geometrical scheme of the radiograph acquisition process) [BL13]. These X-rays pass through the object being attenuated as they pass through each internal structure. The higher the density of the internal structure, the higher the attenuation of the X-rays. As a result, each internal structure is depicted in the radiograph with a different intensity depending on its density. To sum up, a radiograph is basically a shadowgraph or a 2D projection

of a 3D object/subject where its internal structures are visible. Since their discovery, radiographs and their related technologies have evolved significantly: changing from analog radiographs to digital ones, allowing to share them among medical experts, developing post-processing algorithms to improve their quality and interpretability, etc.

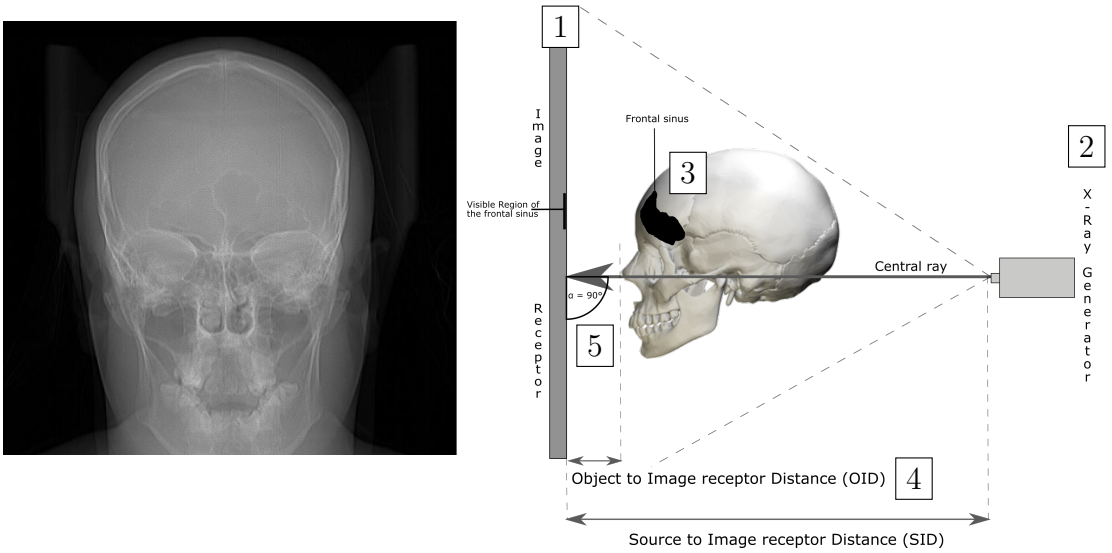

Figure 2: (Left) Posteroanterior radiograph of the head. (Right) Scene of the acquisition of posteroanterior radiograph of the head. The main elements involved in a radiograph are shown: (1) image receptor; (2) image generator; (3) bone (skull) and the target skeletal structure (the frontal sinus); (4) geometric distances (source to image receptor distance and object to image receptor distances) related with the perspective distortion on radiographs; and (5) geometric angles (central ray angel) related with the perspective distortion on radiographs.

Since the discovery of X-rays by Roentgen in 1895 [Roe95], forensic experts have made use of radiographic images as evidence in their endeavour (e.g. bullet analysis [Kir84], age estimation [Goo95], and forensic identification [JAM96, BL00]). During the first decades of the twentieth century, the use of X-rays as a method of positive identification gradually consolidated in scientific literature. In 1921, Schüller proposed the individuality of the frontal sinuses, visible on X-rays [Sch21]. Consequently, in 1927, Culbert and Law expanded the individualizing characteristics of the skull [CL27] and by the mid-1940s the use of radiography was extended to the postcranial skeleton in search of unique features for identification [Dut44, Sco45]. Later, in 1949, CR techniques played a crucial role in the identification of people involved in the Noronic ship's disaster, proving their importance for identification and, subsequently, being included in many mass disasters identification protocols [Sin51].

However, CR requires to acquire PM radiographs simulating the AM data. To do it, it is necessary to have some insight about the radiographs acquisition protocols. Acquisition protocols (some of them are depicted in Fig. 3) are designed to guarantee the radiograph quality (in terms of brightness, contrast, noise, etc.) by establishing several factors of the acquisition process. The most important parameters with regard to CR techniques are those related with the geometrical scene set-up [BL13] (see Fig. 2): (1) image receptor dimensions; (2) the X-ray source-to-image receptor distance (SID); (3) the object-to-image receptor distance (OID); (4) the anatomic

position of the body and its position with respect to the image receptor; and (5) the central ray angle (i.e. the impact angle of the ray that joins the X-ray generator with the centre of the image receptor). These geometrical parameters are set by acquisition protocols to reduce occlusions caused by the overlap among objects and perspective distortions that alter the objects' silhouette shown in a radiograph (e.g. magnification or silhouette distortion). In this sense, acquisition protocols establish the values of these geometrical protocols according to the following guidelines: (1) the SID should be as big as possible to minimise perspective distortion on the silhouette of the skeletal structure, since less perspective distortions occur at a greater SID than at a shorter SID. Furthermore, the bigger the target skeletal structure is the bigger the SID must be to avoid distortions (e.g. chest radiographs requires SID over 180 cm to be able to study the skeletal structure properly); (2) the OID should be as small as possible to minimise perspective distortion on the silhouette of the skeletal structure; (3) the target skeletal structure should be placed over the centre of the image receptor, since perceptive distortion has a greater effect as the distance to the centre of the image receptors grows; (4) the target skeletal structure should be parallel to the image receptor, otherwise perspective distortions occur; (5) the body should be placed in a pose where the overlap of the target skeletal structure with other body object is the minimum possible; and (6) the central ray angle should be perpendicular to the image receptor, i.e. 90°, since distortions increase as the ray is further from perpendicularity. The central ray angle should be only varied when it is needed to properly see a skeletal structure (e.g. Waters radiographs of frontal sinuses) or when the patient cannot cooperate fully.

Radiographs are not as informative as 3D imaging techniques, like CTs, but are still commonly employed for medical diagnosis purposes due to their low cost, high resolution and lower radiation dosages. Furthermore, they are historically the most frequent medical image modality acquired for diagnosis and thus it is still common that many AM candidates only have radiographs as AM data. For all these reasons, radiographs still have a crucial role in forensic identification, despite the difficulty of reproducing the previous factors of acquisition set-up limits their utilization [ARG+10].

II.1.2 Computed tomographies

The first computed tomography (CT) scanner was developed by Godfrey Hounsfield in the 1970s, revolutionising medical imaging techniques. A CT is a 3D volumetric image of an object and its internal parts. Volumetric images are acquired by a CT scanner (see Fig. 4a) or one of its variants (such as multislice CT, nano-CT, submicron CT, micro-CT, and cone beam CT [PIH+15]). Some variants are specially useful for forensic anthropologists since they allow them to analyze very small objects [CSG+18]. CTs are composed of a set of slices of the object (see Fig. 4b), where each internal part is represented with a different intensity level depending on its density. CTs are obtained using X-rays projections and image reconstruction techniques [Her09]. In a CT image, internal objects and skeletal structures are visible in 3D without suffering from overlap and perspective distortions, and thus CTs allow us to fully study their shape, opening a new world of possibility for

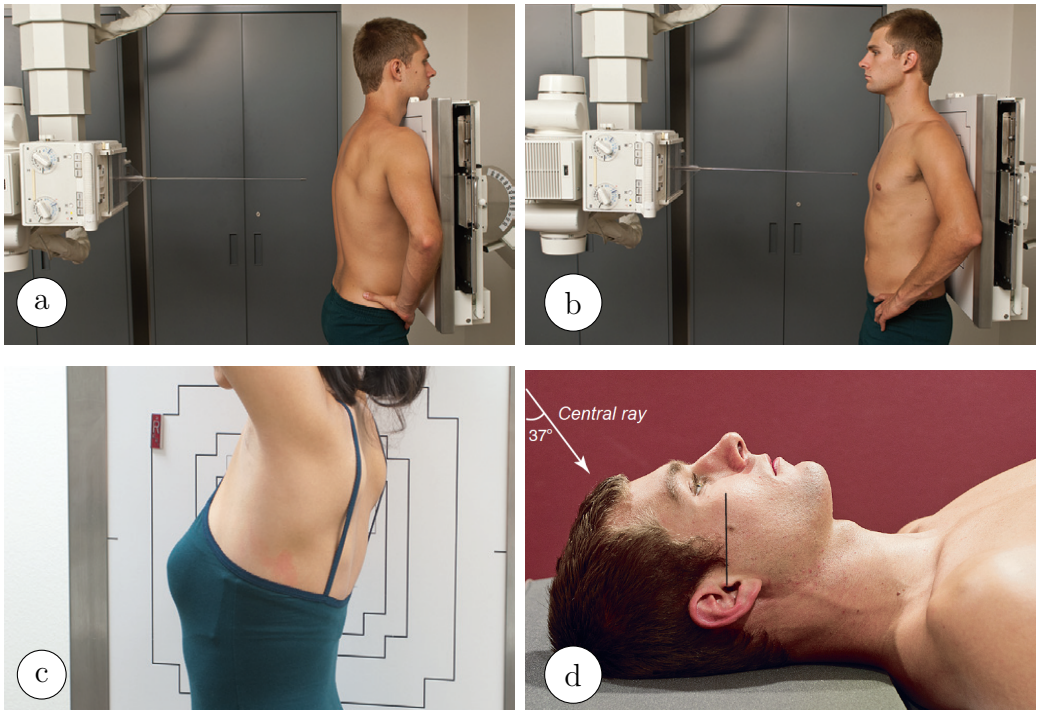

Figure 3: Acquisition protocols in radiographs can be divided into two main categories. Firstly, routine projections, which are procedures that are commonly performed for medical examinatios, such as posteroanterior projections (a), anteroposterior projections (b), and lateral projections (c). Secondly, special projections taken for a better visibility of certain anatomical parts, or when the patient cannot fully cooperate, such as anteroposterior axial projection (d). Images extracted from [BL13].

medical diagnosis [Her09], anthropology [FSF16], etc. However, not all CTs[1] have enough quality for their utilization in forensic identification [BMR+14]. If the slice thicknesses are higher than 2mm, the air-filled structures (such as frontal sinuses and mastoid processes), small features (such as sutures and the clear separation of dentition) and small bones, among others, are decimated [FD16]. Teeth are also fused into column-like structures. All these anatomical fine details are crucial in the identification of unknown individuals. Thus, a maximum slice thickness of 1.25 mm is recommended when skeletal structures are utilized for identification (and 0.05 mm when trabeculae in bones are analyzed [CSG+18]), otherwise it will result in a loss of fidelity of the skeletal structures [FD16].

CTs have a crucial role in the forensic procedures related to virtual autopsies (also called virtopsies) [DJV+06] as a complement to traditional autopsies or as an alternative when the integrity of the body must be preserved. CTs are also employed for sex estimation [UGK+05], anthropological measurement [REM+08], and age at death estimation [DBKK09], among others. CTs can be compared against a wide number of image modalities (such as radiographs, CTs, magnetic resonance images [CSG+18]) for identification purposes. When compared against AM radiographs, the forensic expert has to obtain simulated radiographs (digitally reconstructed radiographs [RRM+05], also called DRR, see Fig. 4c) from the CT simulating the acquisition set-up of the AM ones. This implies a tedious and subjective trial-and-

[1] In medical CT systems, the slice thickness generally ranges from 0.625 mm to 3.0 mm, while the slice thickness from high-resolution industrial scanners is commonly less than 0.045 mm [CSG+18].

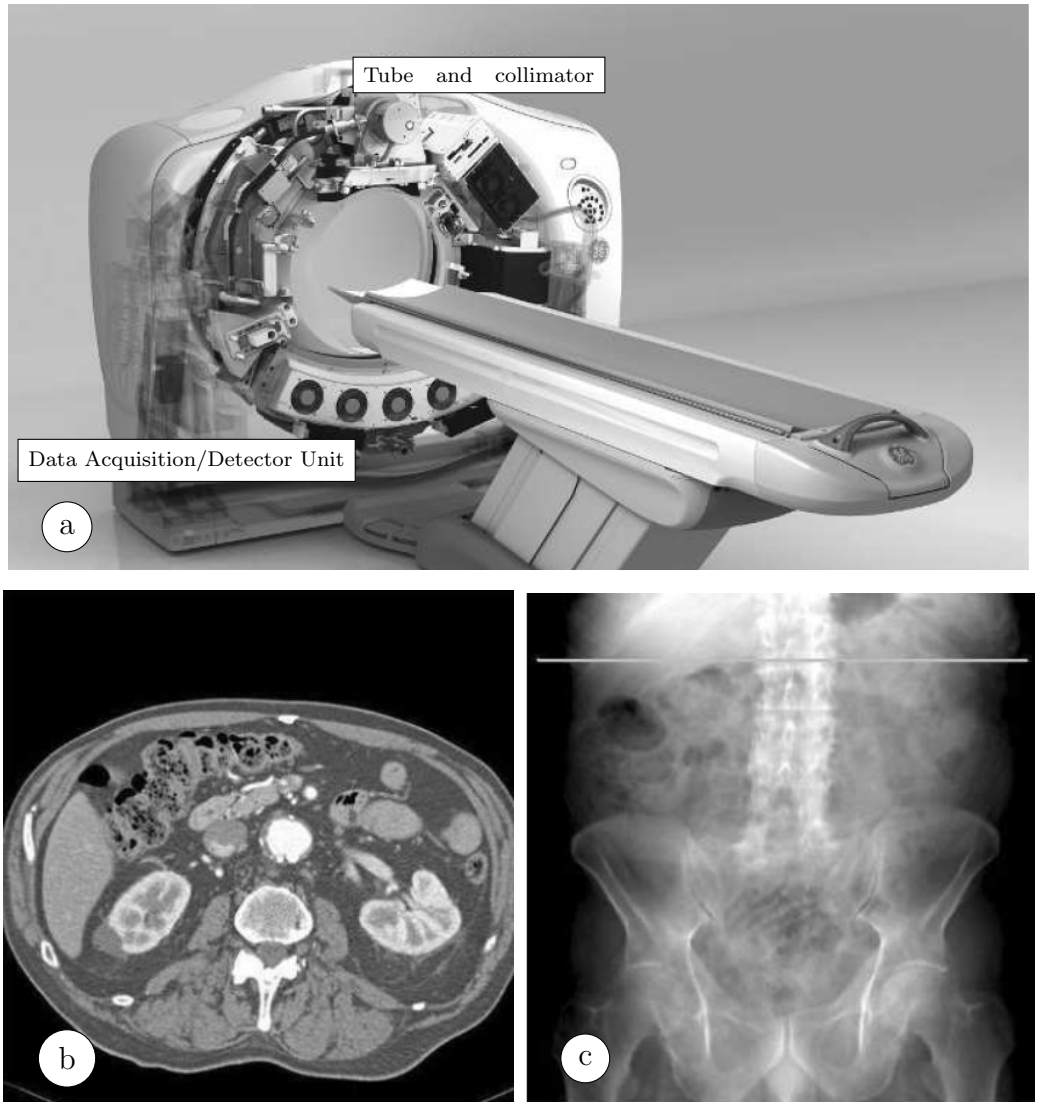

Figure 4: (a) A CT scan where the image receptor and the X-ray generator are visible. These two elements spin together obtaining several 2D X-ray images from which a 3D CT is reconstructed; (b) A CT slice obtained using image reconstruction techniques; (c) A digitally reconstructed radiograph (a simulated radiograph or DRR) of the CT with a horizontal line marking the location of the slice displayed in (b). Images extracted from [Her09].

error process. Meanwhile, the comparison using 3D medical images is performed via the comparison of anthropological measurements [TOA$^+$07, CDLB$^+$15] or via their superimposition [GCC$^+$19].

II.1.3 3D superficial models

Apart from medical acquisition scanners, there are two common approaches for obtaining a 3D image of a skeletal structure within the forensic community [FDGR11, MCA$^+$05, GDGS$^+$16]: laser surface scanners and photogrammetry (see Fig. 6). However, they can only obtain a 3D surface image of an object (i.e. its internal parts are not captured), and thus these approaches are only applicable when a PM "clean" bone (i.e. without soft-tissue) is available. Furthermore, none of them allows us to obtain 3D surface models of air cavities such as frontal sinuses.

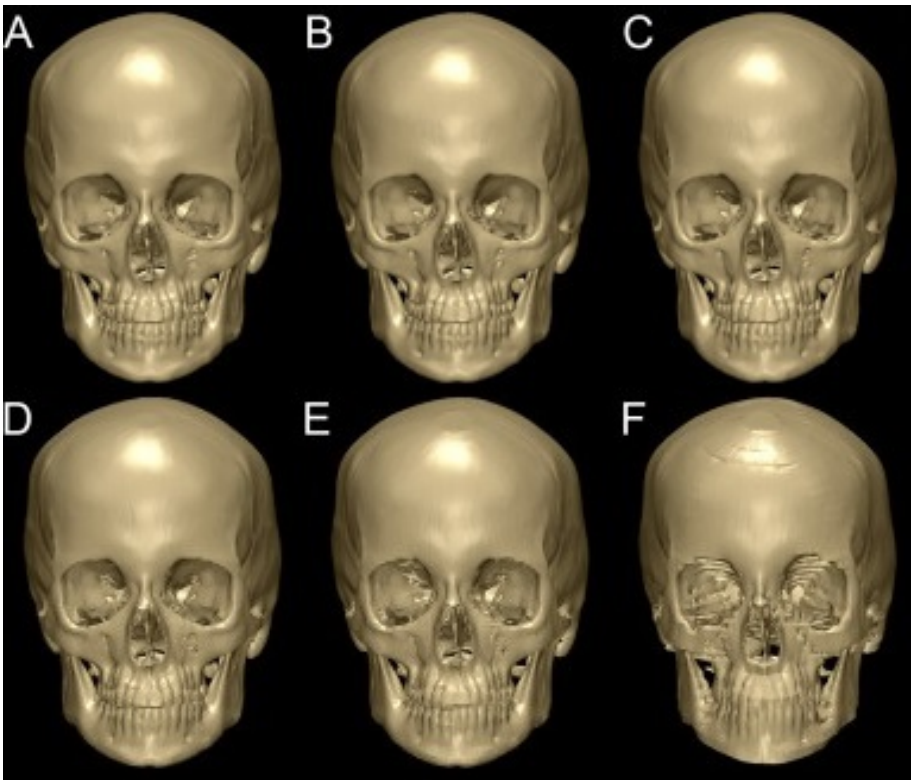

Figure 5: Effect of the slice thicknesses of a CT on three-dimensional reconstruction of anatomical structures: (A) 0.625 mm, (B) 1 mm, (C) 1.25 mm, (D) 2 mm, (E) 2.5 mm and (F) 5 mm. Images extracted from [FD16].

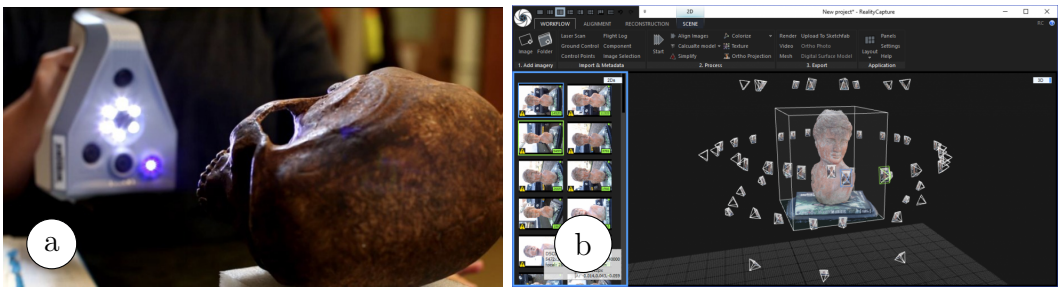

Figure 6: (a) Acquisition of a skull 3D partial view using the Artec Space Spider^{TM} laser range scanner of the Physical Anthropology Lab at the University of Granada (Spain). Image extracted from [ICD12]; (b) Screenshot of the software 3D Scan Expert showing a 3D image obtained from a series of photographs. Image extracted from [3SE].

II.2 Relevant skeletal structures for identification purposes

This section focuses on reviewing the most relevant skeletal structures for CR-based identification (the most important ones are depicted in Fig. 7). Particularly, it analyzes those bones and cavities that have been reported as useful for positive CR identification and with strong support by the forensic community, based on the individuality and uniqueness of their external morphology [CGC11, KH97, Kah09, CC17, WR10, SAT^{+}14, NSGF16] and/or their internal trabecular patterns [Man98, KHS98]. Most relevant skeletal structures are located in the cranial region, but

there are also other useful skeletal structures in postcranial regions (mostly in the chest and vertebrae regions).

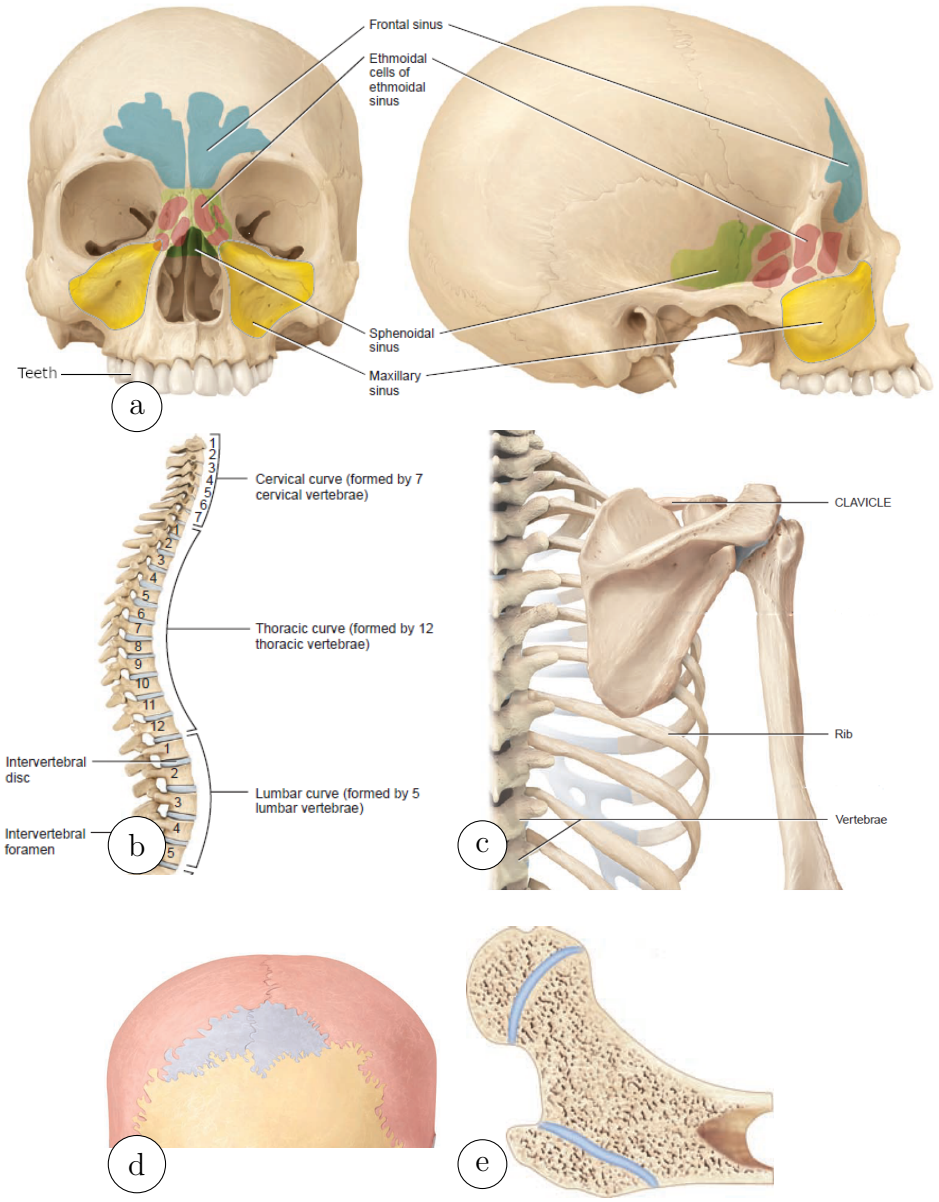

Figure 7: Most relevant skeletal structures for CR-based identification: (a) Paranasal sinuses; (b) Vertebrae; (c) Chest bones; (d) Cranial sutures; and (e) Bone 's trabeculae. Images extracted from [GJT08]

Notice that all skeletal structures are growing from infancy to adulthood and their utilization for identification is only recommended when their growth process has ended.

II.2.1 Cranial region

In the cranial region, the most reliable skeletal structure, and recommended for primary identification, is the teeth [Int18]. Dental comparison plays a crucial role in mass disasters and identification of decomposed and charred bodies [PMS12]. The combination of anatomic features, abnormalities, and dental treatments makes

the teeth patterns hardly similar among different subjects [MdlHVdDLB10, Ada03]. The comparison of these features among AM and PM images (e.g. intra-oral radiographs, extra-oral radiographs panoramic radiographs, and CTs) can provide useful insight from which a positive identification can be made [VR17] by comparing consistences and discrepancies. There are two types of discrepancies, those that can be explained (e.g. a tooth present in AM radiographs that is missing in the PM ones) and those that cannot (e.g. a tooth missing in AM radiographs but present in PM ones). Full mouth-radiographs are preferred over intra-radiographs. In full mouth radiographs, apart from a full view of the whole dentition, another various skull structures are visible (such as frontal sinuses, maxillary sinus, etc.) and can be used for improving the support of the identification decision.

Frontal sinuses are widely recognized as a useful and reliable method of identification [CH18], being one of the few skeletal structures for CR together with the teeth, fulfilling the Daubert criteria [CDLB+15][2]. Frontal sinuses start their development at 1 year, being visible in radiographs around the 5-6 years, and reaching their final appearance around 20 years. Furthermore, frontal sinuses are only absent in 4% of the population [dSPC+09]) and are maintained unchanged during the rest of the life [KWG02]. Although, rarely, some external factors such as traumatisms can change slightly their morphology [CFM+05]. Frontal sinuses are considered as a skeleton fingerprint. Their invariability along adult life and the wide variability in shape and size, number of cavities, intersinus septum, among other features, make them unique [KH97], being different even between homozygotic twins [PVD+07]. Their uniqueness was firstly assessed using an elliptical Fourier analysis [CBS17] of their contours in radiographs [Chr04]. Recently, their 3D shape variability has been quantified [GCC+19, NTG+18]. Their utilization for CR-based identification was first reported in 1926 by comparing their morphology in AM and PM radiographs [CL27]. Nowadays, CR identification based on frontal sinuses' pattern is widely accepted by the forensic community, and many works have reported their utility via image comparison to establish positive identification [YMSS87, dSPC+09, KWG02, CDLB+15, SAL+16, KLP+13, GCC+19]. Furthermore, there are standardised protocols to quantify frontal sinus size and shape and to improve CR results [Chr04, BR10, RLM16, NTGL18]. However, these works highlight that producing PM radiographs simulating the AM ones is an error-prone and complex process, that requires a lot of time, and it is only recommended when other techniques are not applicable. These drawbacks are avoided when comparing the shape of frontal sinuses in AM and PM CTs, increasing their utilization in the forensic endeavour [SAL+16, KLP+13, GCC+19].

Other cranial features, reported as useful for CR are cranial suture patterns [RA04], sphenoid sinus [RKG+12], maxillary sinuses [Sol11], mastoid air cells [CL27], hyoid [CFS+02], and the cranial vault [MR14]. However, the support of these skeletal structures for CR positive identification by the forensic community is inferior to teeth and frontal sinuses. Furthermore, the one with greater support (e.g. cranial suture patterns) are not usually visible in radiographs.

[2]The Daubert criteria [Vic05] determinate whether evidence is admissible in a court of law. An identification method fulfills the Daubert criteria when: (1) it is testable and peer reviewed; (2) it possesses known potential error rates; and (3) it is accepted by the forensic community.

II.2.2 Postcranial regions

In postcranial regions, multiples bones have been reported useful for identification or shortlisting. In the chest and vertebrae regions, clavicles [SWCT11], vertebral features [KGH02] and ribs [MWH77] have been reported as useful for identification. There are also a few bones outside these areas that are commonly used, such as patellae [NSGF16], pelvis [PVD+07], and bones of the hand and foes [KSF05], among others.

II.3 Methodological approaches

Once the imaging techniques and skeletal structures have been reviewed, we proceed to review methodological approaches utilized in the forensic literature for performing CR-based identification. Methodological approaches are divided into three groups according to the dimensionality of the data employed: 2D-2D, 3D-2D, and 3D-3D. The greater the dimensionality, the greater the accuracy and robustness of the methods. Within each of these groups, methods can be further classified into manual approaches and (semi-) automatic approaches. In manual methods, all the identification process is performed by forensic experts. Meanwhile, in (semi-) automatic methods, some tasks of the identification process are automatised.

II.3.1 2D-2D approaches for comparative radiography

In forensic literature, the comparison of AM and PM radiographs is the most extended approach. For instance, the Michigan State University Forensic Anthropology Laboratory (MSUFAL) performed 193 identifications using this approach between 2002 and 2015 [SF18] (see Fig. 8).

Manual approaches: The manual identification procedure [RLM16, Kah09] consists of acquiring PM radiographs trying to reproduce the skeletal structures' silhouettes of the AM radiographs (see Section II.1.1 for a brief review of the acquisition parameters related with the silhouette to be reproduced). Once the PM radiographs have been acquired, AM and PM radiographs are visually compared looking for consistences, explainable inconsistencies (e.g. resulting from the time elapsed among AM and PM radiographs, degenerative process, the effect of gravity on the body, perspective distortions resulting from errors reproducing the AM radiographs, etc) and inexplicable inconsistencies (e.g. point-to-point comparison of frontal sinuses using an standardised protocol [RLM16]). The most relevant factors reported in the literature are: skeletal structures' silhouettes, anthropological measurements, and the presence of pathologies or lesions. Direct measurements cannot be performed since physical units are unknown and are affected by perspective distortions [MVIA18]. Only factors, such as proportions among distances, deviations or asymmetries, can be utilized [dSPC+09, BR10].

(Semi-) automatic approaches: There are several works that (semi) automatically compare different skeletal structures (e.g. frontal sinus [Chr05b, Chr05a], cranial vault [MR14], and teeth [JC04, NAM07]) between AM and PM radiographs. These methods are based on the comparison of the silhouettes of skeletal structures in radiographs using geometric morphometric techniques (such as elliptical Fourier analysis [CBS17]). Then, the AM and PM silhouettes are compared using the el-

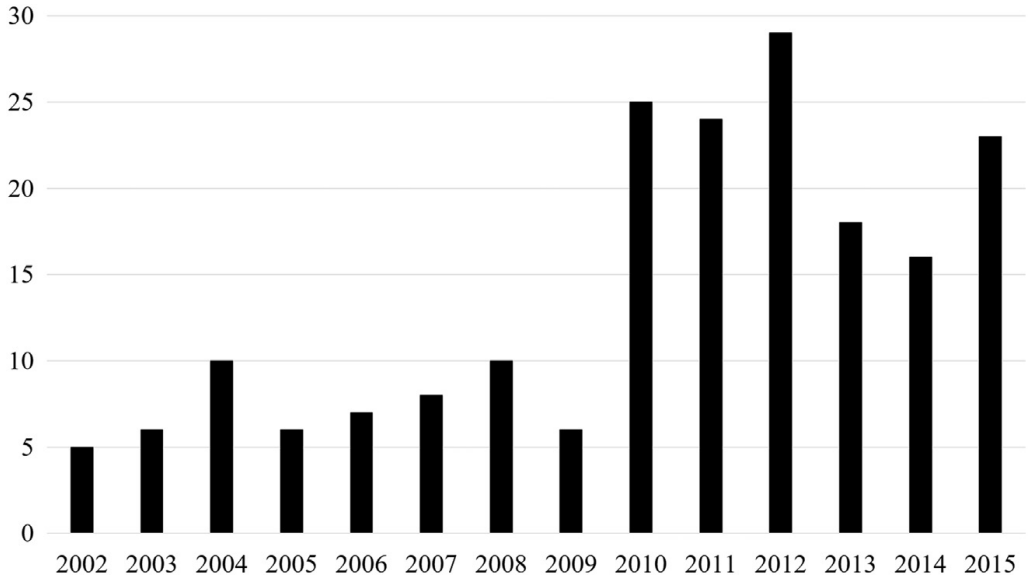

Figure 8: Number of CR-based identifications performed at the Michigan State University Forensic Anthropology laboratory per year, from 2002 to 2015 (n = 193). The most common radiographs employed were: chest/thoracic (32.2%), abdominal/lumbar (18.0%), ankle/foot (16.9%), and head (3.7%; this low percentage is explained due to the few AM records with AM head/skull radiographs). The most common projection was a posterior to anterior or anterior to posterior (77.2%), with lateral (19.9%) and oblique (3.0%) projections also represented. Image extracted from [SF18].

liptical Fourier analysis obtaining a short list of the most probable PM matches for each AM one. All the former methods require the segmentation of the skeletal structures in every AM and PM radiograph. Nevertheless, there are a few computational approaches that avoid the need of manual segmentations, either via using IS methods (e.g. automated dental identification system (ADIS) [DTM$^+$11, AD18] for teeth comparison [DHG18], or [TKWK08] for frontal sinuses segmentation, in both approaches using ad-hoc rule-based segmentation methods, see Section 10) or via the direct comparison of the intensities (e.g. computer-assisted decedent identification (CADI) [DHG18] for vertebrae comparison). However, the latter approach suffers from the elapsed time between AM and PM radiographs and the consequent change in the intensities of the skeletal structures (as stated in the introduction). CADI [DHG18] reduces its impact via the manual selection of a region of interest around each vertebra, the equalisation of the pixels within these areas (e.g. with a histogram equalization filter), and lastly the comparison of AM and PM vertebrae using the Jaccard similarity metric [Jac12].

Drawbacks of both manual and (semi-) automatic approaches: Acquiring PM radiographs simulating the AM ones is a complex and error prone trial-and-error process, since small changes in the acquisition parameters (e.g. SID, OID, central ray angle) result in great changes in the skeletal structures' silhouettes [SG14]. Furthermore, both the acquisition of the PM radiographs and their visual comparison against the AM ones rely completely on the forensic expert's skills and experience. As a consequence, the utility of the method is reduced because of the time required (2-8 hours per superimposition depending on the AM radiograph [SF18]) and the inherent subjectivity. Some authors [ARG$^+$10, CGC11] have recommend to only use this approach as a last resource in validation scenarios (confirming an identity or deciding among few possible candidates). Nevertheless,

many forensic researchers [RLM16, CH16, SF18, SDW+18] are currently working to provide standardized, quantifiable methods for radiograph-radiograph comparison. These standards minimize the effect of these drawbacks and establish a minimum number of matching points to make an identification or an exclusion), and thus have improved the reliability of the radiograph-radiograph comparison approaches.

II.3.2 3D-2D approaches for comparative radiography

Manual approaches: The comparison methodology is similar to the manual radiograph-radiograph comparison methodology but this requires the acquisition of PM simulated radiographs from a CT trying to simulate the AM radiographs [HDC+14, Kah09, PVD+07, SHNY17], instead of real radiographs. These simulated radiographs are manually obtained by the forensic anthropologist through a trial-and-error process using generic medical imaging software (such as Vitrea, see Fig. 9, or digital autopsy). Afterwards, AM radiographs and PM simulated radiographs are visually compared as in the radiograph-radiograph approach. Therefore, these approaches have the same limitations than the radiograph-radiograph comparison approach (i.e. time-required and subjectivity) but allowing to generate as many PM radiographs as necessary without the corpse after its CT scan. Another drawback of this approach is the higher cost of a CT scanner in comparison of radiograph acquisition device (composed by X-ray generator and an image receptor), as many forensic labs cannot afford them (as stated in Section II.1.2) [DCI+11].

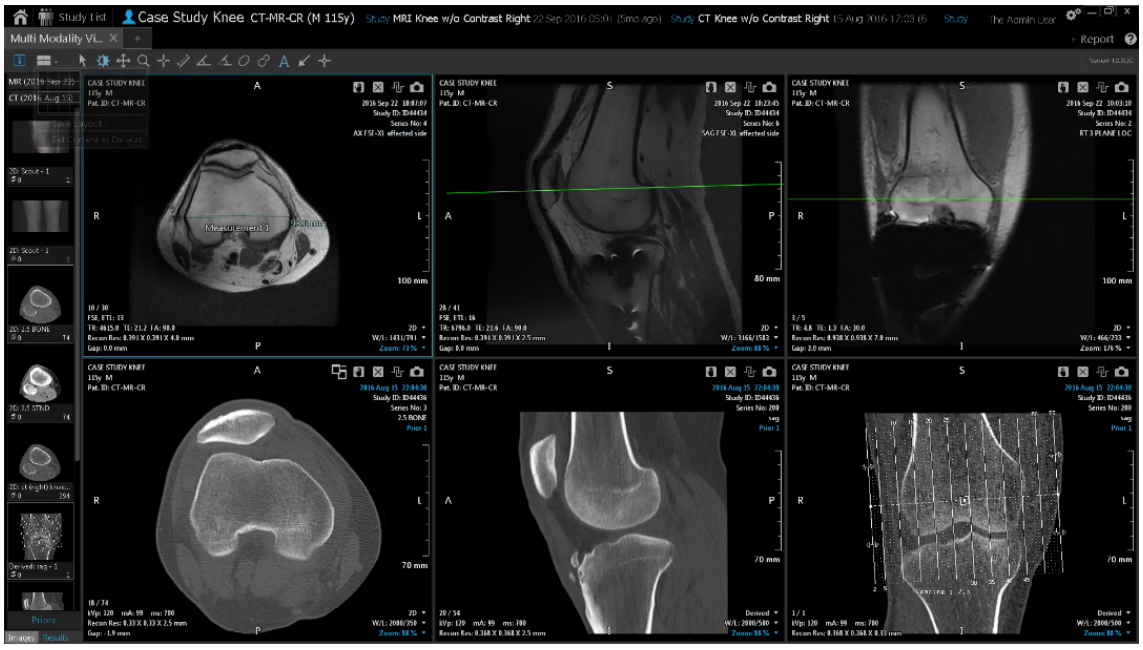

Figure 9: Screenshot of Vitrea® Advanced Visualization [Vit] generating a simulated radiograph from a CT. Image extracted from [Vit].

(Semi-) automatic approaches: Automatic 3D-2D CR approaches reduce the errors related to the manual acquisition of PM real or simulated radiographs (i.e. time required, subjectivity and errors caused by the fatigue of the forensic expert). However, there are just a few automatic approaches for the comparison of AM radiographs and PM 3D images [SAT+14, DGBS17, NSGF16]. These approaches are based on the acquisition of 3D surface models with a 3D laser range scanner of the

skeletal structure (clavicles in [SAT+14, DGBS17] and patellae in [NSGF16]). From these PM 3D surface models, a set of predefined 2D projected images are generated through the 3D model rotation. Notice that these 2D projections only contain the silhouette of the target skeletal structure. Finally, this set of PM projections is automatically compared to the manually segmented silhouette of the skeletal structure in the AM radiographs using elliptical Fourier analysis descriptors. However, these methods are limited by the set of predefined 2D projections and assume the value of the parameters related to perspective distortions. As far as we know, there are no approaches that completely automatize the search for the best possible 2D projection of the PM 3D surface model of the skeletal structure.

II.3.3 3D-3D approaches for comparative radiography

Manual approaches: Lastly, the CT-CT comparison approach is the most reliable one and does not have any of the previous limitations since the 3D shapes can be directly compared [IFY+16, HCO+17, RBC+16a, DDC+19]. The comparison is performed via visually comparing their 3D shapes [GCC+19], avoiding occlusions or perspective distortions, or via anthropological measurements [KLP+13], where the distances can be directly compared since CTs maintain the original physical units. Thus, when AM and PM CTs are available, this approach is recommended over the latter two due to its greater reliability and forensic potential [GPF+18].

(Semi-) automatic approaches: Few computerised approaches have been proposed for the comparison of AM and PM 3D data (such as [ZYF+11, ZYW+13, ZOZF16] with teeth, [GCC+19] with frontal sinuses, or [DF19] with lumbar vertebrae). These methods required the segmentation of the 3D skeletal structures in both the AM and PM CTs (although the PM data could also be acquired with a 3D laser range scanner), their automatic registration, and the measuring of the quality of the match. However, the availability of 3D AM data (such as CT) is scarce compared to the number of AM radiographs available (specially, when people who disappeared a long time ago are involved) reducing significantly their applicability.

Chapter III

Theoretical Background

"You see, but you do not observe."
— Sir Arthur Conan Doyle

This chapter is devoted to briefly review the basics and the state of the art of the techniques utilized in this PhD thesis to automate the CR problem. An extensive survey of these techniques is beyond the scope of this dissertation, so the analysis will only focus on the most relevant contributions in the literature. Particularly, it focuses on two classical computer vision tasks, **image segmentation (IS)** and **image registration (IR)**, and in two soft computing techniques utilized to tackle them, **deep learning (DL)** and **evolutionary algorithms (EAs)**, respectively. Both computer vision problems are considered middle-level tasks. In a nutshell, low-level tasks are related to primitive operations, such as smoothing, enhancement, and histogram transformations; middle-level tasks tackle the analysis of images; and high level operations focus on image understanding and give meaning to image analyses performed by low- and middle-level tasks. Thus, IS and IR allow us to analyse an image to gain an insight about it, which in turn enables to tackle more complex tasks as CR.

III.1 Image Segmentation

IS consists of partitioning an image into regions (i.e. sets of pixels) [PXP00], each of them with a different semantic meaning (e.g. segmenting a frontal sinus in a skull radiograph). Regions usually do not overlap, although there are some particular segmentation scenarios where a pixel can be part of multiples regions (e.g. segmenting lungs and clavicles in chest radiographs). Automated segmentation processes allow many applications to be executed in real time (manual segmentations are tedious and time-consuming), while reducing subjective and segmentation errors. Segmentation algorithms are nowadays a crucial part of many computer vision systems, both as a pre-processing stage and also as an end in itself. The range of practical applications go from medical imaging systems [SFB+15] to traffic control systems [SAANM03], ranging through many others such as human-robot interaction systems [JJLL15], video surveillance systems [KH02], etc. As a consequence, the importance of segmentation algorithms have grown exponentially and have become a hot topic of research (as can bee seen in Fig. 10).

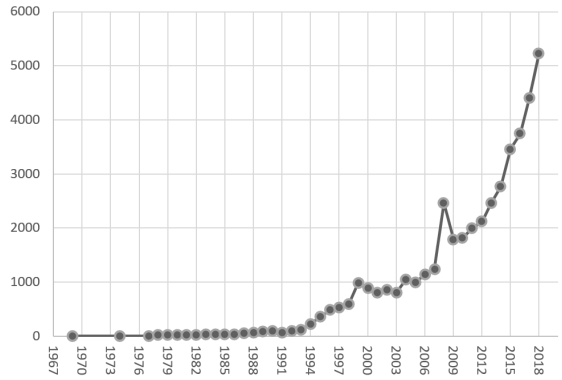

Figure 10: Number of articles related to image segmentation published from 1967 to 2018 according to Scopus. Search performed the 8th August 2019 using the keywords (TITLE-ABS-KEY ("image segmentation") AND (LIMIT-TO (DOCTYPE , "ar"))).

In medical imaging applications, IS algorithms segment medical images (such as radiographs) helping to the quantification and measurement of tissue volumes, diagnosis, locating pathologies, and the study of anatomical structures [Mes14]. For instance, clavicles are used to detect lesions, such as tumorous lesions [KTK08], or for forensic identification [SAT+14] via comparing their silhouette in AM and PM radiographs. Many different taxonomies can be presented to classify IS approaches [VGRV01, WK07, Kha14, SFB+15]. One possible classification could be the following based on the underlying computer vision techniques utilised:

- Rule-based methods, where the image is segmented by the application of a set of low-level and spatially blind rules [VGRV01]. The most important families of methods within this category are:
  - Thresholding methods [BDBP15] are based on grouping all pixels belonging to an intensity range into one group. Traditionally, only two groups (background and foreground) are considered, needing only one threshold value to separate them (lower values are background, greater values are foreground). However, since spatial information is ignored, segmentation results are not sufficient, unless the intensity levels of the segmented objects and the background are clearly separated.
  - Edge detection methods [NSH08] focus on contour detection. These methods usually work under the assumption that an abrupt change on pixels intensities likely represent an edge. To detect these changes and their directions, various edge operators can be utilized, most of them based on pixels' gradient information (e.g. convolutional filters such as Sobel or Canny). However, these methods usually fail with overlaping objects, fuzzy borders or noisy images.
  - Region growing methods [PSB12] partition an image into connected regions. To do so, several seed pixels are randomly selected or manually selected inside the object of interest. Regions are then grown from these seed pixels by grouping all neighboring pixels with similar features (e.g. pixel intensity, colour, etc). Adjacent regions are merged when some similarity criteria are accomplished (e.g. homogeneity or sharpness of region

boundaries). The main drawback of these methods is that segmentation results are not robust to the initial seeds.

- Deformable models [MICC16], where the segmentation is performed by matching a model that includes some sort of prior shape information to the image. Generally, these methods start from an initial boundary shape and iteratively modify the shape through several shrinking and/or expansion operations. These operations searches to minimize an energy function that measures how well the shape fits the boundary of the object to be segmented. To tackle this optimization process, numerical optimization methods and metaheuristics are usually utilized. The two main drawbacks of this family of methods are their sensibility to the initial shape and the risk of the optimizer being trapped in a local minimum due to the high multimodality of the search space. There are mainly two types of deformable models:
  - Parametric/explicit methods represent curves and surfaces explicitly in their parametric forms. These models incorporate information about the object to segment, such as the mean shape, shape variability, mean location, mean orientation or mean size. For example, some notable methods are: active contour models, active shape models, and active appearance models.
  - Geometric/implicit methods (also called level set methods) represent curves and surfaces implicitly without their parameterization via using curve evolution theories and level set methods. These models allow us to segment curves and surfaces that cannot be expressed in a parametric form.

- Atlas-based methods [COL+11], generally based on the registration of an atlas (i.e. an already segmented image) and a target image (see Section III.3 for further detail of a registration process). Once registered, a mapping function between both images is obtained. The segmentation is obtained by mapping the segmentations in the atlas image to the target image.

- Graph-based methods [PZZ13] that represent the image as a graph that is partitioned into a set of separated connected components (generally making use of techniques such as conditional random fields or Markov random fields).

- Machine learning-based methods [SFB+15], traditionally based on handcrafted features (e.g. SIFT) together with a classifier (i.e. k-NN or an artificial neural network) but, with the advent of Convolutional Neural Networks (ConvNets) [GGOEO+17], this paradigm has shifted towards end-to-end approaches where the ConvNet input is directly the image to segment and the output is the target segmentation.

Since each methodology has its own pros and cons, the best results are commonly achieved via hybrid approaches that combine two or more of the former strategies [CYRC18, TI11].

The automatic segmentation of anatomical structures in radiographs remains very challenging, despite the great clinical importance of radiograph understanding.

This is mainly due to the projective nature of X-ray imaging, which causes large overlapping of anatomies, fuzzy object boundaries and complex texture patterns. For instance, even among expert radiologists, minor errors in diagnosis are performed in *circa* 30% of studies [BWA15] and major errors in 3-6% [RWC+99, BLMM12]. As a consequence, the automatic radiograph segmentation has been extensively studied since the 1970s [TSNF73, WS77]. More than 150 research works dealing with this problem were already published during the twentieth century [VGRV01], raising the number to 388 at present, according to Scopus[1]. In general terms, in the case of radiograph segmentation, most approaches are either rule-based [VGRV01], shape-based [VGSL06], or machine learning-based [MHS17]. Given that the state of the art in radiograph segmentation are deep learning techniques, and ConvNets in particular, this category will be analized in Subsection III.2.2.

III.2 Deep Learning

As a general description, deep learning methods [GBC16, LBH15] learn high-level abstractions from data by using hierarchical neural architectures. These machine learning techniques have revolutionized many classical artificial intelligence domains, such as computer vision, natural language processing, semantic parsing, and many more. As a result, their popularity has grown exponentially, as did the amount of papers employing deep neural networks (see Fig. 11).

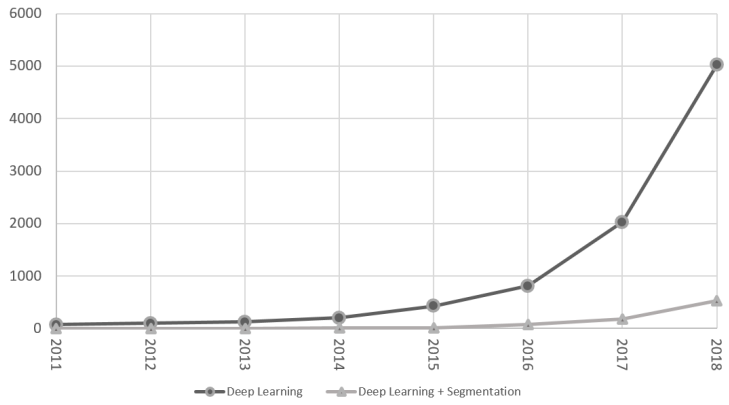

Figure 11: Number of articles related to deep learning (dark gray) and deep learning for segmentation (light gray) published from 2011 to 2018 according to Scopus. (Deep Learning) Search performed the 8th August 2019 using the keywords (TITLE-ABS-KEY ("deep learning") AND (LIMIT-TO (DOCTYPE , "ar"))). (Deep Learning + Segmentation) Search performed the 8th August 2019 using the keywords (TITLE-ABS-KEY ("deep learning") AND TITLE-ABS-KEY (segmentation) AND (LIMIT-TO (DOCTYPE , "ar"))).

Artificial neural networks (ANNs) [B+95, Sch15] can be considered as the seeds of deep learning, whose origin can be tackled back to the 1950s. ANNs are learning algorithms roughly inspired in biological neural networks. ANNs are comprised of neurons, and each neuron is composed of (see Fig. 12): (1) a set of n input signals x; (2) a set of weights w that quantify the importance of each input; (3) a linear

[1]Search performed the 8th August 2019 using the keywords (TITLE-ABS-KEY (chest AND X-ray AND segmentation) OR TITLE-ABS-KEY (chest AND radiograph AND segmentation) AND NOT TITLE-ABS-KEY (computed AND tomography))

aggregator $\sum$ which gathers all input signals weighted by the synaptic weights to produce an activation "voltage"; (4) an activation bias θ, that is a threshold utilized to shift the activation function; (5) an activation potential u, equal to the difference between the linear aggregator output (i.e. the activation "voltage") and θ (i.e. $u = \sum_{i=1}^{n} w_i \cdot x_i - \theta$); (6) an activation/transfer function g (e.g. sigmoid, hyperbolic tanget, exponential linear unit (ELU), rectified linear unit (ReLU), etc.) which limits the neuron's output to a range of values; and (7) an output signal y, that results of applying the activation function to the activation potential (i.e. $y = g(u)$).

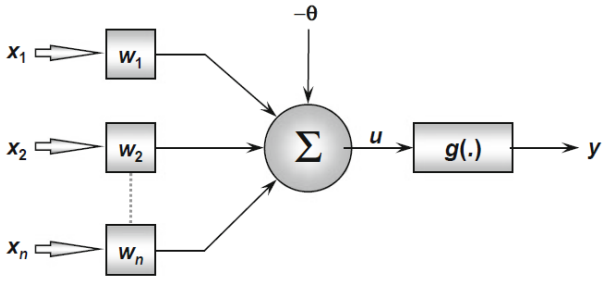

Figure 12: The artificial neuron. Image extracted from [DSSF+17].

There are numerous ANN architectures depending on how neurons are arranged and interconnected, but in general ANNs are divided into three parts, input layer, hidden layers and output layer (see Fig. 13). Hidden layers are responsible for extracting patterns and abstractions from the data, and perform most of the internal processing. When multiple hidden (intermediate processing) layers are introduced into an ANN, it is usually called a "deep" neural network, hence the term deep learning. ANNs can learn the relationship between the input data (e.g. a chest radiograph) and the output (e.g. whether there is a tumour or not in the input radiograph) by adjusting the weights and biases, i.e. the trainable parameters. Thus, the fine-tuning/training process is crucial for the ANN performance. The most relevant training approaches are the following:

- Supervised learning: the networks are trained with a dataset composed of pairs of input data and its desirable output (each pair is usually referred as training sample). The trainable parameters are adjusted to minimize the differences between the desirable outputs and the network's outputs. Within supervised learning approaches, the following subfamilies can be distinguished:
  - Offline learning (or batch learning approaches), which updates the trainable parameters after one epoch (i.e. when all the training samples have been presented to the network). Furthermore, there are some variants that update the trainable parameters after each n trainable samples. The value n is usually refereed as batch size. These methods are recommended with unchanged problems where a complete dataset is avialable from the beginning.
  - Online learning, which updates the trainable parameters after each training sample. These methods are recommended with problems that change quickly limited by the absence of a complete dataset from the beginning.

- Unsupervised learning: the networks are only trained with input data without knowing the desirable output. These methods are usually utilized to detect patterns on data (e.g. clusters).

- Reinforcement learning: the network output represents the action of an agent in an environment (e.g. the action utilized by a player in a videogame) and the trainable parameters are updated according to a certain reward metric, which measures the goodness of the action performed (e.g. the score achieved on a videogame).

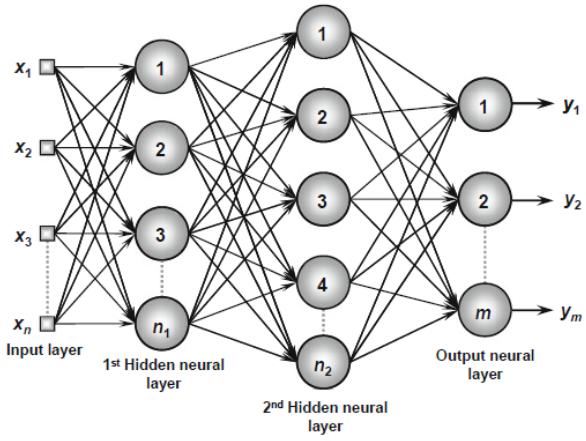

Figure 13: Example of a feedforward and fully-connected ANN with two hidden layers. Image extracted from [DSSF+17].

In the 1970s, backpropagation methods [LBOM12] were a big step forward for supervised training of ANNs. Backpropagation methods allow us to compute the gradient of each neuron for a given input and a desirable output. These gradients can be interpreted as an indicator of how each neuron's output (and consequently its trainable parameters) should change to increase the network accuracy with respect to a loss function (e.g. percent of true positives and true negatives in classification problems). These gradients are then utilized by a gradient-based numerical optimization method [NW06] to find the best update of the network parameters. This process is applied iteratively, improving on each iteration the network behaviour. The most common optimization method utilized for training ANNs is gradient descent, but other optimizers have also been utilized, such as the Adam optimizer [KB14] or the Levenberg–Marquardt optimizer [Lev44, Mar63]. Nowadays, backpropagation methods are still the basics for supervised learning of deep neural networks. However, these training methods suffer from the same issues of any complex optimization problem, such as local minima, as well as some ad-hoc problems, such as the vanishing gradient problem[2] or the exploding gradient problem[3]. Another

[2]The vanishing gradient problem consists of the difficulty of training the first layers of very deep neural networks since their gradients approaches zero. This is because gradients become smaller at each backward propagation, and thus gradients are smaller as the distance to the output layer increases.

[3]The exploding gradient problem is a difficulty found in training deep neural networks when gradients accumulate and have a large value, which leads to very large updates of the network at each iteration.

crucial problem of deep neural networks was the computational cost of training them. Thus, deep neural networks were considered hard to train efficiently for a long time [Sch15]. In the meanwhile, several relevant contributions were performed with a great importance in today's deep learning techniques, such as recurrent neural networks in the 1980s [Jor86], Boltzmann machine in 1985 [AHS85], restricted Boltzmann machine in 1986 [Smo86], convolutional neural networks in 1989 [LBD+89], deep belief network in 2006 [HOT06], or deep Boltzmann machine in [SL10]. Finally, in the 2010s, deep learning methods finally become feasible thank to the advances on hardware, particularly in GPUs, parallelization, the incorporation of several architectures changes (such as using ReLU activation function, dropout, etc), and the availability of very large labeled datasets. Particularly, the spark that started the deep learning revolution was a ConvNet proposed in 2012 called AlexNet [KSH12] (which was an evolution of LeNet [LBD+89, LBD+90] proposed in 1990 for handwritten numbers recognition) that obtained an improvement of $\sim 10\%$ with respect to state-of-the-art computer vision classification methods in the ImageNet competition [DDS+09] (see Subsection III.2.1 for further details of ConvNets, LeNet and AlexNet).

Nowadays, in 2019, the most relevant families of deep neural networks are the following [GBC16]:

- ConvNets: ANNs that use convolutional filters. ConvNets are mostly employed to solve computer vision problems. They are reviewed in greater detail in the following section.

- Recurrent neural networks: ANNs with backward connections among layers. These networks are mostly employed in domains where the order within a data sequence is crucial, such as in natural language processing, speech synthesis, and machine translation.

- Autoencoders: ANNs composed of a down-sample stage, followed by a central layer called code, and lastly followed by an up-sampling stage, all comprised by hidden neurons. As a general description, these networks are trained to copy the input into the output, obtaining a simplification or an extraction of the principal components of the input in the code layer. Autoencoders are typically employed for dimensionality reduction and to distinguish between real and fake data (e.g. in bank transitions).

- Generative adversarial networks (GANs): ANNs composed of two networks, a generator and a discriminator, which are trained simultaneously. These two networks compete with each other, since the generator is trained to generate realistic fake data while the discriminator is trained to distinguish between real and fake data. GANs are mostly utilized for image generation and for style transfer among images.

The rest of this section is devoted to review ConvNets for general purposes (see Subsection III.2.1) and for IS (see Subsection III.2.2), as well as training and regularisation strategies (see Subsection III.2.3). In these subsections, only the most notable ConvNets and architectural innovations are presented, due to the immense number of works presented in this area in recent years.

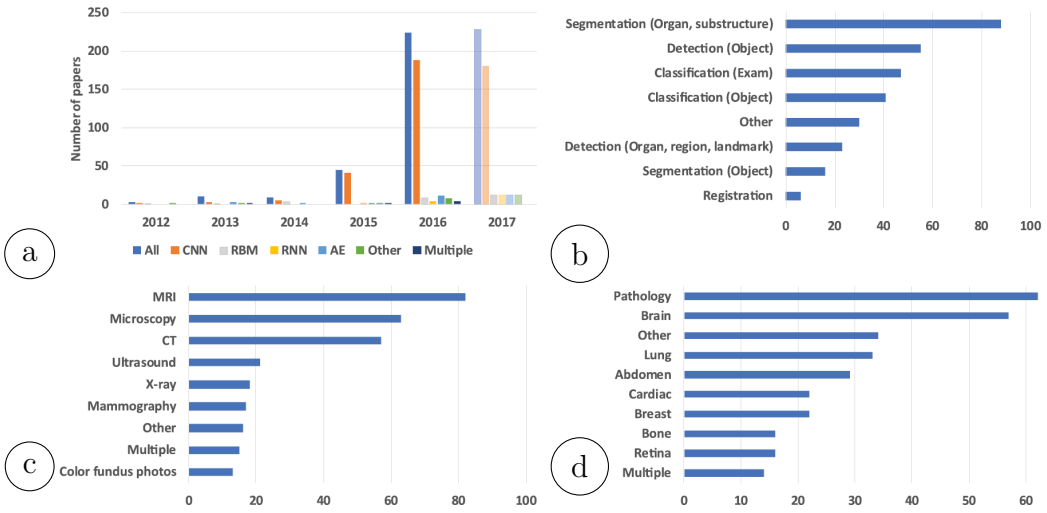

Figure 14: Summary of the papers related to deep learning and medical imaging published from 2012 to 2017 by year of publication according to: (a) type of ANN (CNN refers to ConvNets; RBN to restricted Boltzmann machine; RNN for recurent neural networks; and AE to autoencoders); (b) task addressed; (c) imaging modality; and (d) application area. Image extracted from [LKB+17].

III.2.1 Convolutional neural networks

ConvNets are a family of deep neural networks designed to directly process data with a grid-like topology, e.g. images, without any preprocessing or feature extraction stage. Particularly, in the computer vision field, ConvNets have solved numerous tasks with an increasing level of difficulty [VDDP18, GLO+16], such as object detection, motion tracking, action recognition, human pose estimation, and IS, among many others. As a consequence, ConvNets have also revolutionised the analysis of medical images (see Fig. 14), allowing us to automate crucial tasks [LKB+17], such as landmarks location, disease detection and classification, organ and lesion segmentation, and many others.

The typical pipeline of a ConvNet (see Fig. 15a) is composed of convolutional layers and pooling layers in the body of the ConvNet, and fully-connected layers in its end. These layers are detailed below:

- **Convolutional layers** (see Fig. 15b) are the key component of ConvNets. They are layers that employ a linear operation called convolution [Sze10]. The input layer is a tridimensional grid composed by two spatial dimensions (width and height) and depth (related to colour channels and to the feature maps[4] generated in the previous layer). In general terms, a convolutional layer convolves the input using various learned kernels of a given size. Kernels are tridimensional having a size of width×height×input depth size (e.g. 3×3×1), where width and height are hyper parameters, and input depth size is set by the layer input dimensions. Thus, each output neuron interacts only with a small set of input neurons (e.g. 9 neurons with $3 \times 3 \times 1$ kernels), instead of interacting with all input neurons as in traditional ANNs (see fully-connected layers), and the output dimension is equal to the number of learned kernels. Notice that the output of a convolution operation is not the neuron output

[4] A feature map is the result of applying a convolutional filter to a given input.

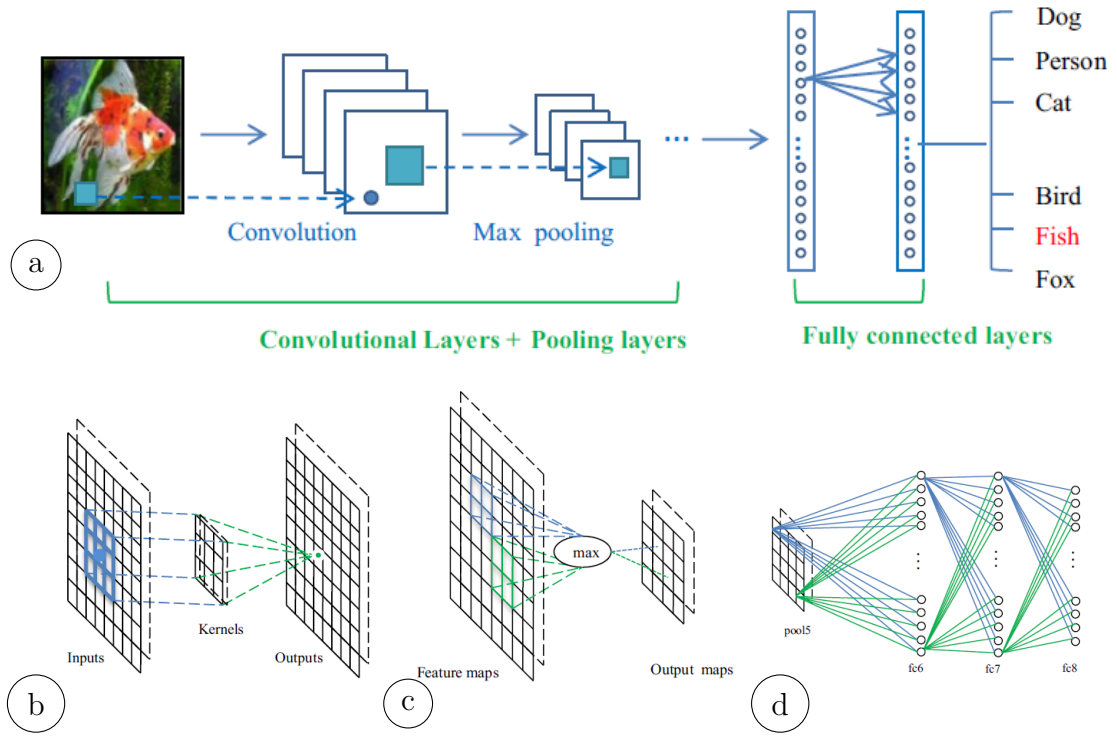

Figure 15: (a) The pipeline of a general ConvNet architecture for classification; (b) Convolution operation; (c) Pooling operation; and (d) Fully connection operation. Images extracted from [GLO+16].

but instead the input of the neuron activation functions. Furthermore, instead of learning an ad-hoc kernel for every output neuron as in traditional ANNs, kernels are shared among all input neurons, making ConvNets equivariance to translation (but not to other spacial transformation such as rotations or scale). These features reduce significantly the number of weights/parameters that have to be learned (e.g. for a input of 100×100×1 neurons with 1 convolutional kernel of 3×3×1, only 9 weights have to be learned instead of 90,000 weights), making easier to train ConvNets and reducing their memory requirements. In summary, the main hyper-parameters of a convolutional layer are the following:

- Number of kernels, the output depth dimension is equal to the number of kernels.
- Kernels size, composed of its height, width and depth. Width and height are hyper parameters that have to be chosen, although height and width are usually equal, and the depth is always equal to the depth of the layer input.
- Activation function, which limits the neuron output to a range of values, and basically determines whether a neuron should be activated or not, i.e. it determines whether the neuron's input is relevant for the prediction or whether it should be ignored. Examples include sigmoid, ELU, ReLU, etc.
- Stride, which is the step size of the kernel when moving through the input (e.g. convolution is applied on each pixel with a stride of 1, and

convolution is applied only to 1 out of every 2 pixels with a stride of 2). Thus, the stride affects directly to the output size, which is equal to input size divided by the stride, i.e. the input is downsampled with any value different to 1.

- Padding strategy, which indicates how convolution is applied in the border of the input, i.e. when a position of the kernel is outside the limits. Padding strategies are either based on not using convolution operations on the borders reducing the output size, or on using convolution on the borders via assuming that the values outside the limits are equal to a certain value, typically this value is equal to 0 (zero padding strategy).

- **Pooling layers** (see Fig. 15c) reduce the spacial dimension (width × height) of the layer input by the pooling size (e.g. a pooling size of 2 reduces an input of 100×100×10 to 50×50×10), while depth is not reduced. Pooling layers are beneficial to ConvNets since they significantly reduce the computation and weights required by subsequent layers, reducing over-fitting and increasing the equivariance to scale transformations. Average pooling and max pooling are the most commonly used strategies.

- **Fully-connected layers** (see Fig. 15d) are regular ANN layers. In these layers, every output unit interacts with every input unit. These layers have high computational and memory requirements and have a significant amount of weights to be learned, and thus its utilization in ConvNets is usually only seen at their end when the input dimensionality has been reduced after several pooling layers. The most relevant hyper-parameters of these layers are the number of output neurons (e.g. in classification problems, the last fully-connected layer usually has the same number of output neurons as classes) and the activation function.

One of the first ConvNets was LeNet, proposed in 1990 [LBD+90]. LeNet is composed of two blocks, each comprised by a convolutional layer followed by a pooling layer, and two fully-connected layers at the end. LeNet was able to recognize digits in 32 × 32 images but due to its computational requirements it was hard to implement until the 2010s. In 2012, a new ConvNet, called AlexNet [KSH12], was proposed for classifying colour images of 224×224 pixels in the ImageNet competition [DDS+09]. The most relevant features of AlexNet with respect to LeNet were the usage of ReLU as activation function, the inclusion of a reguralization technique, called Local Response Normalization (see the subsection III.2.3 for further details about regularization), and a few architectural changes (particularly, AlexNet was composed of 5 convolutional layers, 2 max pooling layers, and 2 fully-connected layers).

Afterwards, numerous ConvNets architectures were proposed but the first one that significantly improved the AlexNet behaviour in the ImageNet competition was a ConvNet called VGG [SZ14]. The main contribution of VGG was showing that the depth of a network is a critical component to achieve a better accuracy. The basic version of VGG, called VGG-16, was composed of 5 blocks, each of them composed of 2 or 3 convolutional layers and a max pooling layer, followed by 3 fully-connected layers (the last of them utilizing a softmax activation function instead of a ReLU).

The next relevant contribution was the incorporation of **inception modules/ layers**. An inception layer is composed of several convolutional layers computed in parallel, each with a different kernel size, whose outputs are concatenated in the depth dimension (see Fig. 16). Some variants reduce the computational requirements by decreasing the input depth by adding an extra 1x1 convolution before the convolutional layers. In general terms, inception layers allow us to capture local and global features simultaneously, and to increase the scale equivariance. GoogLeNet [SLJ+15] was the first ConvNet that incorporated inception layers. The utilization of inception layers, a regularization technique called batch normalization (BN) [IS15] and a deeper architecture with 22 layers, led GoogLeNet to be the winner of ILSVRC 2014 competition.

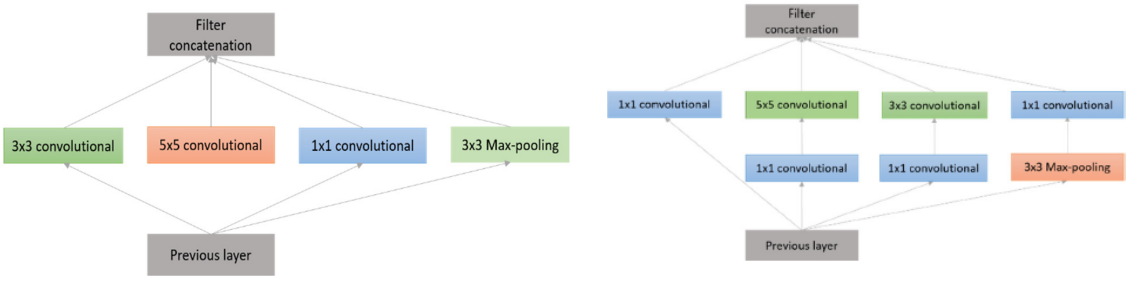

Figure 16: (Left) Example of a naive inception module. (Right) Example of an inception module with dimension reduction. Images extracted from [ATY+19].

However, despite increasing ConvNet depths can be beneficial for the performance [SZ14], the vanishing and exploiting gradient problems make these ConvNets hard and slow to train. **Residual connections** [HZRS16] lessen this problem allowing us to significantly increase ConvNets depth. Residual connections are basically feedforward connections between nonconsecutive layers (see Fig. 17), i.e. the input of a certain layer with a residual connection is the concatenation of the output of its directly previous layer and one or more other previous layers. The first ConvNet with residual connections was ResNet [HZRS16], which was the winner of the ILSVRC 2015 competition for classification and also obtained the best results with the COCO dataset [LMB+14]. Particularly, residual connections have allowed us to successfully train several variants of ResNet going from 34 to 1202 layers (e.g. ResNet-50 is composed of 49 convolution layers, 2 pooling layers and 1 fully-connected layer at the end of the network).

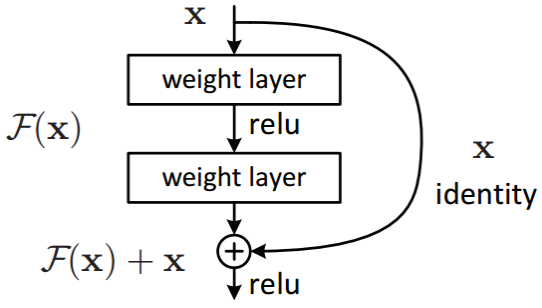

Figure 17: Example of a residual connection. Image extracted from [HZRS16].

Overall, numerous ConvNets have been proposed obtaining better performances in some particular computer vision problems but these are mostly variants and combinations of the ConvNets described below and/or their architectural innovations. Furthermore, these are the basis of ConvNets designed for IS, although there are also several ad-hoc architectural innovations in ConvNets for solving the IS problem (e.g. atrous convolution [CPSA17]), which are reviewed in Subsection III.2.2.

III.2.2 Convolutional neural networks for image segmentation

The first ConvNet trained end-to-end for IS was a Fully Convolutional Network (FCN) [LSD15] with a loss function based on the pixel segmentation accuracy. FCN (see Fig. 18) is mostly based on VGG-16 [SZ14], sharing its 5 convolutional blocks (each of them composed of 2 or 3 convolutional layers and a pooling layer), but without the last 3 fully-connected layers. Instead of the fully-connected layers, FCN appends 3 convolutional layers (the first with convolutional filters of 7×7 and the last two with filters of 1×1; the last of them with the same number of filters as classes to be segmented). Notice that 5 pooling layers are utilized in total and, thus, the input images are down-sampled by a factor of 32, decreasing significantly their quality. To minimize this drawback, a last up-sampling layer with a residual connection with the input is utilized. **Up-sampling** layers increase the dimensionality of their input by a given factor f, which allows us to restore the size loss by the down-sampled layers (notice that down-sampled layers cannot be removed since it will hugely increases the memory requirements of the ConvNet and allow us to obtain spatial information). Up-sampling layers are based either on **interpolation** functions (e.g. bilinear interpolation) or on **deconvolution** filters. Deconvolutional layers are basically convolutional layers with a fractional stride of $1/f$, that allow us to learn nonlinear up-sampling functions, instead of using fixed linear interpolation functions, thus achieveing a better performance. However, the global performance of FCN was insufficient in comparison to other state-of-the-art methods in the VOC2012 competition [EEVG+15] for IS.

U-Net [RFB15] is an evolution of FCN composed of 2 parts following an **encoder-decoder architecture** (see Fig. 18) with residual connections between the two parts. The first block, or down-sampling part, has a FCN-like architecture composed of 4 blocks, each with 2 convolutional layers with filters of 3×3 and 1 pooling layer, followed by 3 last convolutional layers. This part reduces the feature maps dimensionaly decreasing the memory and computation requirements while extracting features and spatial information. Meanwhile, the second block, or up-sampling part, is composed of 4 blocks, each comprised by 2 convolutional layers with filters of 3×3 and 1 up-sampling/deconvolutional layer, and one final convolutional layer with filters of 1×1 with the same number of filters as classes to be segmented. This second part progressively reduces the number of feature maps while increasing their height and width. Furthermore, the residual connections allow us to avoid loosing pattern information in the down-sampling and up-sampling processes while dealing with the exploiting and vanishing gradient problems. It overcomes the results of FCN on the VOC2012 competition.

Both FCN and U-Net, as well as their variants, are based on the same layers of ConvNets designed for other computer vision tasks. The most relevant ad-hoc

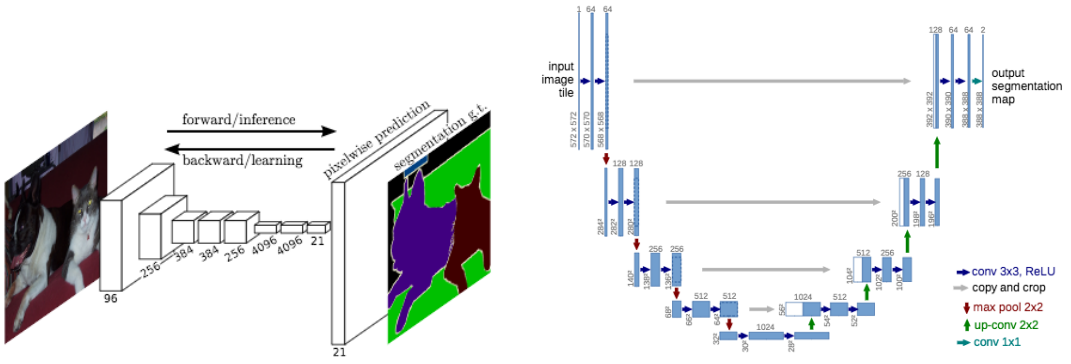

Figure 18: (Left) Fully Convolutional Networks, image extracted from [LSD15]; (Right) U-Net, image extracted from [RFB15]

architectural innovation for IS is atrous convolution. **Atrous convolution** (also called dilated convolution) is a convolutional operation that introduces a spacing between the values in a kernel (the number of spaces between values is called rate or dilated rate) (see Fig. 19). This allows us to adjust the filter's field-of-view and capture multi-scale context information without reducing the spatial dimensions of the feature maps (i.e. a 3×3 kernel with a dilation rate of 2 will have the same field-of-view as a 5×5 kernel, while only using 9 parameters and without down-sampling). However, this is computationally expensive and takes a lot of memory, as a consequence its use is normally preceded by a few pooling layers to make the features maps computationally approachable. Atrous convolution was firstly introduced in [YK15], which is a variant of VGG-16 without the last 2 pooling layers (leaving 3 pooling layers and thus reducing the input only by a factor of 8) and with atrous convolution in the last 2 blocks (with a rate of 2 and 4 in the 4th and 5th blocks, respectively) and a final 1×1 convolutional layer similar to the one of FCN or U-Net. This ConvNet with atrous convolution outperformed FCN and U-Net, but also other advanced ConvNets, such as DeepLab V1 [CPK+14], which is a variant of VGG-16 with a final fully-connected conditional random field (CRF) layer [KK11].

DeepLab V2 outperforms previous ConvNets and all state-of-the-art methods in terms of performance in the VOC2012 competition. The main innovations of DeepLab V2 were: (1) adopting ResNet-101 instead of VGG-16, which by its own improved significantly the performance; (2) the utilization of the only 3 first pooling layers of ResNet-101, reducing the input image only by a factor of 8, instead of 32; (3) the combination of inception modules with atrous convolution resulting on **atrous spatial pyramid pooling** (ASPP) modules (see Fig. 19), which allows us to obtain even more spatial context of the input feature maps reducing the effect of only down-sampling by 8; and (4) maintaining the last fully-connected CRF layer of Deep Lab V1.

Afterwards, Deep Lab V3 rethinked [CPSA17] the utilization of atrous convolution by combining **cascade modules**, which are composed of consecutive atrous convolution layers with an increasing rate size (e.g. 5 layers with rates of 1, 2, 4, 8 and 16), and deeper ASPP modules, boosting its performance over its previous version without using fully-connected CRF layers, which is no longer needed since detailed spatial context is already introduced through atrous convolution based

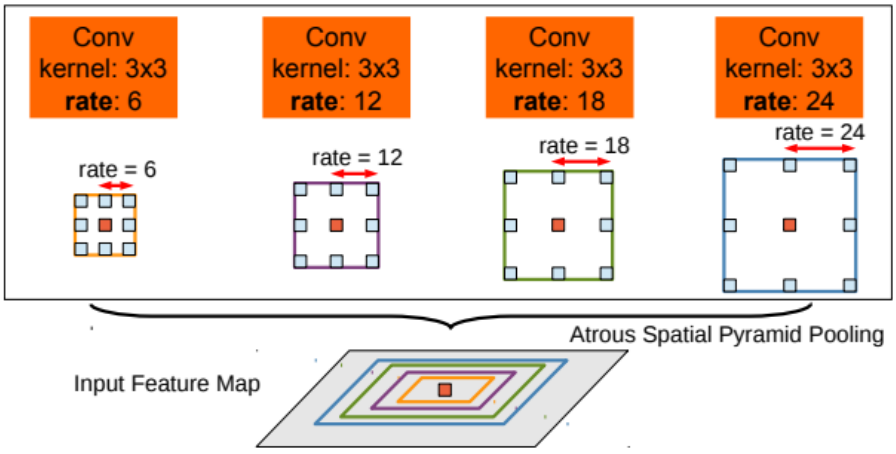

Figure 19: Several atrous convolution filters with rate sizes of 6, 12, 16, and 24. The combination of these filters with an inception module results on a atrous spatial pyramid pooling (ASPP) module. To classify the centre pixel (orange), ASPP exploits multi-scale features by employing multiple parallel filters with different rates. The effective Field-Of-Views are shown in different colours. Image extracted from [CPK+17]

modules. Lastly, Deep Lab V3 is extended by Deep Lab V3+ [CZP+18], introduced in the second semester of 2018, by adopting a U-net like architecture where the encoder part is basically Deep Lab V3 and the decoder part is similar to the one in U-Net but with some architectural changes (see [CZP+18] for further details). Furthermore, there is also a "light" version of Deep Lab V3+ with a lower computational complexity, while having a similar performance, which replaces standard convolution and atrous convolution layers by depthwise separable convolution and depthwise separable atrous convolution layers, respectively (see Fig. 20). **Depthwise separable convolution** factorises a standard convolution operation into 2 consecutive operations: (1) a depthwise convolution layer (see Fig. 20a) or depthwise atrous convolution layer (see Fig. 20c), which performs a convolution operation for each depth level independently; and (2) pointwise convolution, which is a 1×1 convolutional operation that combines the output from the depthwise convolution operations.

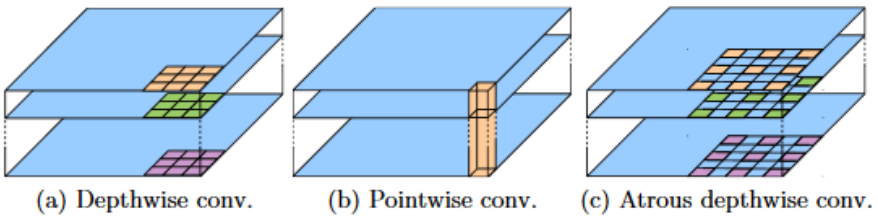

Figure 20: (a) Depthwise convolution; (b) point wise convolution; and (c) depthwise atrous convolution. Image extracted from [CZP+18]

However, despite the success of Deep Lab V3+ in computer vision competitions, U-Net (and some variants, such as InvertedNet [NLM+18]) is still frequently utilized in many application domains. U-Net is designed in a way that can be trained with only a few hundreds of images, due to its low number of parameters to be

learned. Therefore, U-Net is suitable for IS of medical images since, in the medical domain, there are rarely large datasets available. Furthermore, it was published (code included) in 2015 in MICCAI (Medical Image Computing and Computer Assisted Intervention), which is a main medical conference. It propitiated its use for the segmentation of medical images in many application domains.

III.2.3 Regularization strategies for training convolutional neural networks

A crucial problem on training ConvNets, specially in very deep ones and/or with a large number of learnable parameters, is achieving a similar performance with training data (which was used to train the ConvNet) and with test data (which has never been seen by the ConvNet) avoiding overfitting. In other words, ConvNets should learn to generalize from the training data to perform well in new data, instead of just "memorizing" the training data. **Regularization strategies** [GBC16] increase generalization (and in consequence also reduce overfitting) improving the results on testing data, at the expense of slightly increasing training errors. The most remarkable, and commonly employed, regularization strategies utilized for training ConvNets are the following:

- **Data augmentation** [GBC16]. The simplest strategy to improve generalization is to train ConvNets with more data but this is not always possible, specially with medical data. An intermediate solution is to create fake training data from the real training data through image transformations (e.g. translations, rotations, flip, crop, zoom, etc.) and add them to the training dataset.

- **Early stopping** [GBC16]. When training ConvNets, the training error decreases steadily over the epochs. However, if the ConvNet is validated after each epoch with new data (also called validation data, which is usually a split of the training data that is only utilized for validation), the validation error starts to increase after a certain number of epochs. An early stopping strategy consists of stopping the training process after a certain number of epochs without improvement on the validation error, and then returning the learned parameters that achieved the lower validation error rather than the latest parameters, which provided the smallest training error.

- **Dropout** [SHK+14]. The dropout technique reduces co-adaptation, which is a learning problem involving that some neurons have a greater importance in the ConvNet's predictions than others. This increases overfitting, that it is usually caused by learning all the weights of the ConvNet simultaneously. Dropout consists of randomly dropping or disabling neurons, along with their connections, from the ConvNet during training. Each neuron is disabled randomly with a fixed probability p, independently of the rest of neurons, following a Bernoulli distribution (regular dropout) or a Gaussian distribution (**Gaussian dropout**). On each epoch, a different set of neurons is dropped and, consequently, a different set of weights is employed and "learned".

- **Batch normalization (BN)** [IS15]. BN allows us to reduce the training time of ConvNets and to improve generalization, via reducing covariate shift.

Covariate shift refers to the change in the distribution of ConvNet activations due to the change in their parameters during training. These distribution changes force the subsequent layers to adapt to those changes, which slows down the learning process. As a general description, BN reduces covariate shift by normalizing the ConvNet activations of a layer across spatial locations (in the case of convolutional layers) and the batch. Furthermore, if the batch size is equal to 1, BN normalizes the activations only across spatial locations (a variant called instance normalization [UVL16]). Instance normalization has a better performance than regular BN in some ConvNet architectures and problems.

As a brief summary, designing or choosing a ConvNet, choosing its hyperparameters and its training strategy for solving a particular problem is far from trivial, and there are few guidelines [LMAPH18], since all design choices are strongly interconnected and the high computational cost of exploring all possibilities prohibit an exhaustive finetune.

III.3 Image registration

Image registration (IR) [MTLP12] is the process of integrating two images into a common coordinate system, where one of the images is fixed and the other is transformed. IR has multiple applications [OT14], specially in remote sensing, medical diagnosis and image guided surgery. Thus, numerous IR approaches and applications have been proposed in the literature through the years, as can bee seen in Fig. 21.

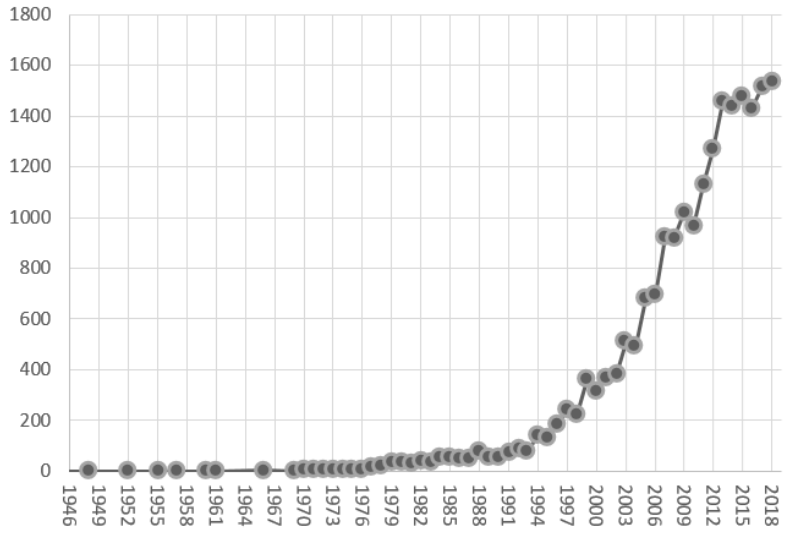

Figure 21: Number of articles related to image registration published from 1949 to 2018 according to Scopus. Search performed the 8th August 2019 using the keywords (TITLE-ABS-KEY (“image registration”) AND (LIMIT-TO (DOCTYPE , ”ar”))).

There is not a universal standard for any IR method because several considerations of the particular application must be taken into account. Nevertheless, IR methods usually require the four following components (see Figure 22): (1) the fixed

model image and the scene image to be transformed; (2) the registration transformation that relates the scene and the model images; (3) a similarity metric, which measures the resemblance between the fixed model image and the transformed scene image; and (4) an optimizer, which looks for the best parameters for the transformation to minimize the error of the similarity metric. These components are further detailed in the following subsections.

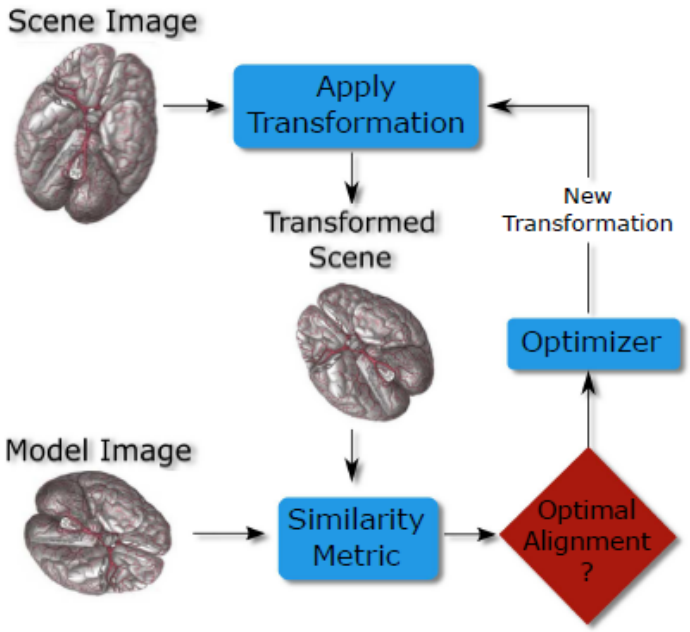

Figure 22: Basic schema of an IR process where the scene image is transformed through an optimization process to minimize its error with respect a fixed model image. Image extracted from [Ber18]

III.3.1 Nature of the images

According to the dimensionality of the images, IR approaches are categorised as **2D-2D**, **3D-3D**, and **3D-2D**. The choice of one over the others depends on the data available. 3D-3D IR approaches are the most informative and robust, and they are the focus of a large amount of research [Ber18] (both with CTs [MVM+18] and 3D surface scans [GD18]) because of their utility in many scenarios, specially in the medical domain. However, the availability of 3D images, although becoming more popular in developed countries, is still scarce compared to the number of 2D images. When one of the images is forcefully a 2D image, 3D-2D IR approaches show the best performance since these are robust to object's pose in the 2D image. Even so, finding the best 2D projection of the 3D images with respect to the fixed 2D image is complex and computationally expensive [TLSP03]. Lastly, 2D-2D IR approaches are the less informative and are sensitive to the pose of the object in the images. Furthermore, another relevant factor in IR approaches is that the images have been obtained with the same acquisition device (**monomodal IR**) or with different ones (**multimodal IR**).

Regardless the dimensionality, IR approaches are classified into **intensity-based**

and **feature-based** according to the information that guides the optimizer. Intensity-based methods compare intensities [MTLP12] (approaches are based on mean squares, normalized correlation, pattern intensity, or mutual information) of a 2D projection of the volumetric image with the 2D image. In medical imaging, intensity-based methods are more extended because they do not require segmentation, that usually involves subjectivity and errors (e.g. they have been successfully applied to the 3D-2D IR of CTs and radiographs [MTPL08]). Feature-based methods minimize the distance between geometrical features (i.e. isolated points or point sets, contours, or surfaces) to be extracted in both images. The main use of feature-based approaches in medical domains are in those scenarios where the modality of the images is different or in those where the intensities between the images cannot be related.

III.3.2 Registration transformation

A registration transformation is a mapping function between the space of the scene and the model images. There are two main categories: linear transformations and elastic transformations [Ber18]. **Linear transformations** modify the entire image but preserving geometrical features, such as distances, lines and angles. There are several subcategories within linear transformations depending on the geometrical information preserved by the mapping functions: i) **rigid** transformations, which alter translation and rotation preserving lengths and angles; ii) **similarity** transformations, which alter translation, rotation and scale (in the same proportion in all axis) preserving the aspect ratio and angles; and iii) **affine** transformations, which alter translation, rotation, scale (in a different proportion in each axis) and shear preserving only parallelisms. Meanwhile, **elastic** transformations deformate locally the image using "internal" and "external" forces without preserving any geometrical feature. These transformations alter the scene image preserving their dimensionality and thus are utilized for both 2D-2D and 3D-3D IR approaches.

Meanwhile, 3D-2D IR approaches are based on **projective** transformations, which are a kind of liner transformations that only preserve collinearity. A projective transformation describes a mapping from 3D to 2D coordinates. Projective transformations are classified according to the type of camera that they model into: **orthographic projections** that model an orthographic camera (see Fig. 23a); and **perspective projections** that model a pinhole camera (see Fig. 23b) [HZ03]. A pinhole camera is composed of 6 extrinsic parameters (3 translations and 3 rotations of the camera), related with the position of the 3D object in the world, and 5 intrinsic parameters (1 focal distance, 1 pixel aspect ratio, 2 principal point coordinates, and 1 skewness), related to perspective distortions. The 11 degrees of freedom (DoF) of a pinhole camera allow us to reproduce the Even so, projection of any photograph or radiograph. Although the pinhole camera is often simplified by assuming that the principal point is in the centre of the image, that the aspect ration of the pixels is square and that there is no skewness. This simple pinhole camera with 7 DoF (6 extrinsic parameters: 3 translations and 3 rotations of the camera; and 1 intrinsic parameter: focal distance), which considerably reduces the complexity of the search space, while allowing perspective distortions to be reproduced within most photographs and radiographs. Furthermore, the pinhole camera is further simplified in some applications by assuming the value of the focal distance, reducing the DoF

to 6. Meanwhile, an orthographic camera is a particular case of a pinhole camera located at the "infinity" and thus it does not model perspective distortions. Furthermore, the orthographic camera is more mathematically tractable (only 6 DoF: 2 translations, 3 rotations, and 1 scale) and the constraint of the translations and the scale does not require expert knowledge.

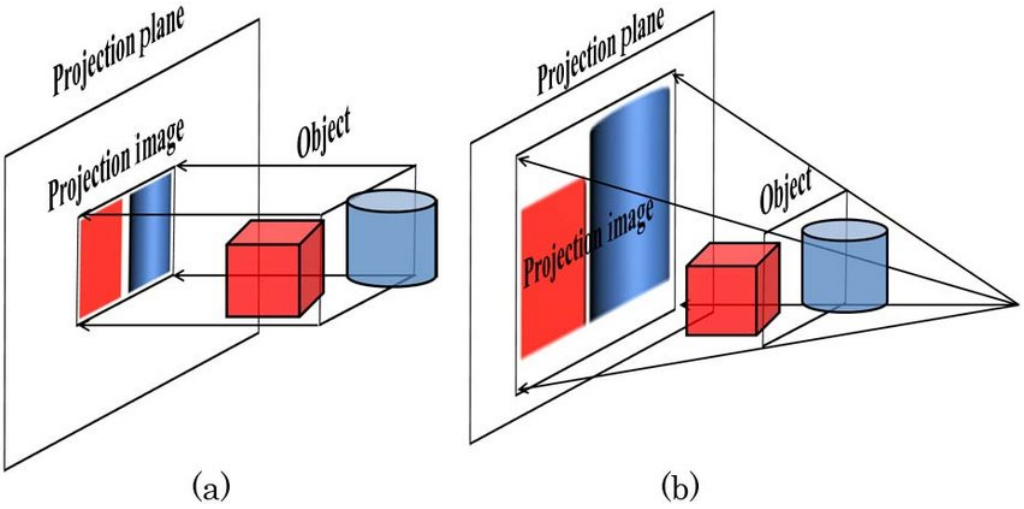

Figure 23: (a) Orthographic projection. (b) Perspective projection. Images extracted from [JLJW14].

III.3.3 Similarity metric

Similarity metrics asses the quality of a registration transformation by measuring the matching between the fixed model image and the transformed scene image. Several similarity metrics have been proposed due to its importance in the field of computer vision [VH01] and in its application to medical imaging [MTLP12]. Again, similarity metrics are classified into two distinct categories depending on the information that guide the optimization process:

- Intensity based metrics measure the similarities between the intensity distributions of the pixels or voxels of two images. Common intensity based metrics are: sum of squared differences [OSP02], normalized correlation [Son11] and mutual information [PMV03]. Their suitability depends on whether the intensities between the images can be correlated or not, e.g. with monomodal or related modalities as radiographs and CTs.

- Feature based metrics measure the similarity of geometrical features between the images. The most common metrics for measuring the similarities between two sets of 3D (or 2D) corresponding points are mean square error [WB09] and median square error [SCDI09]. Meanwhile, the most common metrics for comparing two 2D silhouettes are Hausdorff Distance (HD) [BTE98], Jaccard Index (JI) [Jac12], and Dice Similarity Coefficient (DICE) [Sør48].

III.3.4 Optimization

The key idea of the IR process is finding the best registration transformation of the scene image to model image with respect to a similarity metric. Many factors affect to the structure and complexity of the search space tackled, as the registration transformation and its DoFs or the similarity metric. Since IR search spaces are often highly multimodal with numerous local minima, even in their simplest versions, exhaustive search methods are unsuitable. Furthermore, IR problems are **computationally expensive** since the evaluation of the fitness of a candidate solution requires two computationally expensive operations: the transformation of the model image and the measurement of the matching between the scene and the model images. Numerical optimization methods, both linear search methods (Nelder-Mead, BFGS, LBFGS) and trust region methods (Levenberg-Marquardt, BOBYQA), have been employed for IR [Mod04]. However, these methods have only obtained a good performance in those problems where either a good initialization can be assumed or the registration transformations and their parameters can be significantly constrained. These drawbacks have been overcome by IR methods based on real-coded evolutionary algorithms (RCEAs), also called evolutionary IR methods, in several IR problems [DCS11, SCD11, VBDC18, Ber18]. RCEAs [BFME97, YG10, MLH18, ZYQ19] have improved the results obtained by traditional methods in many IR problems [CDS07, InBC$^+$09, Ber18]. RCEAs are global optimization techniques with a robust performance that enables them to tackle complex medical IR problems. As a consequence, several image registration problems have been tackled using RCEAs over the last years [DCS11, SCD11, VBDC18] (see Fig. 24).

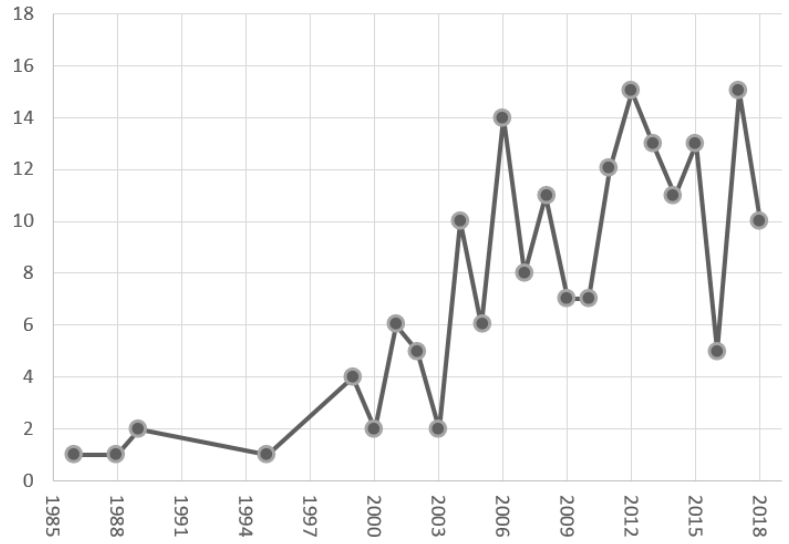

Figure 24: Number of scientific contributions related to evolutionary image registration published from 1985 to 2018 according to Scopus. There is a clearly growing trend, specially after the year 2000 and, currently, between 10 and 20 contributions are published every year, making a total of 559 works published until the year 2018. Search performed the 2nd October 2019 using the keywords (TITLE-ABS-KEY ("image registration") AND (TITLE-ABS-KEY ("evolutionary algorithm") OR TITLE-ABS-KEY ("genetic algorithm") OR TITLE-ABS-KEY ("evolutionary algorithm") OR TITLE-ABS-KEY ("evolutionary") OR TITLE-ABS-KEY ("metaheuristic") OR TITLE-ABS-KEY ("metaheuristics") OR TITLE-ABS-KEY ("stochastic optimization") OR TITLE-ABS-KEY ("stochastic search") OR TITLE-ABS-KEY ("heuristic search"))).

III.4 Real-coded evolutionary algorithms

RCEAs [BFME97, YG10, MLH18, ZYQ19] are global optimization techniques inspired by biological evolution for solving optimization problems over **continuous search spaces**. RCEAs are population-based optimizers, i.e. RCEAs maintain a population of candidate solutions (also called individuals). Each candidate solution is composed of an array of parameters coded as real numbers (each parameter is called gene and the array of all the genes is called chromosome). As a general description, RCEAs usually contain the following operations: (1) the population is randomly **initialized**; (2) several **offsprings (i.e. new candidate solutions) are generated** by varying the candidate solutions of the population using operation such as selection, crossover, mutation, etc; (3) the "goodness" of each candidate solutions is measured using a fitness function; (4) a new population is formed with candidates solutions from the previous population and the offsprings according to a certain criteria based on their fitness (e.g. the best ones); (5) repeat the operations 2 to 4 (these operations together are known as a **generation**) until some **stop criteria** are reached (e.g. a maximum number of generations or a fitness threshold). Through these evolutionary processes, the candidate solutions evolve in **parallel** to meet the the criteria defined by the fitness function without any assumption about the underlying fitness landscape. Furthermore, these evolutionary processes have to look for a trade-off between **exploration** (i.e. exploring new regions of the search space, related with the population diversity) and **exploitation** (i.e. improving the current candidate solutions, related with local optimization) [Mar91] of the search space to avoid local minima and to find the best possible solutions. Lastly, RCEAs are **stochastic** optimizers since they involve several random operations in their initialization and in the generation of the offsprings. As a consequence, RCEAs do not guarantee to find the same final solution in each execution of the algorithm. It is recommended to perform several runs of a RCEA to study their robustness.

RCEAs have been widely applied in many real-world problems because of its robustness, fast convergence, and the reduced number of parameters to set in some variants [DS11]. They have been successfully applied to optimization problems including non-linear, non-differentiable, non-convex and multi-modal functions [Cha08].

III.4.1 Promising real-coded evolutionary algorithms for image registration problems

Differential Evolution (DE) [SP97], a classical RCEA, has shown an outstanding performance on global numerical optimization problems, as demonstrated in the IEEE CEC competitions [QL13]. DE is widely praised by its reduced number of parameters to fine tune [DS11], its robustness and its fast convergence. Furthermore, several self-adaptive DE approaches yielded better results than the classical DE in many different problems [DS11]. Among them, a self-adaptive DE approach with a linear reduction of population and an external memory of elite solutions (to enforce diversity in the mutation), called L-SHADE [TF14], has shown a very significant performance. L-SHADE ranked on the first positions at the IEEE CEC2014 competition on real-parameter single objective optimization [TF14]. In this competition, L-SHADE's results outperformed other state-of-the-art DE variants.

Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [HMK03], another

classical RCEA, has also demonstrated an great performance on global optimization competitions [Los13]. CMA-ES has advantageous convergence properties and performs well with small populations, which makes it even more promising when it comes to improve the computational time. Several modern CMA-ES variations have yielded better results than the classical CMA-ES in many different problems [Los13]. Among them, a restart CMA-ES with two interlaced restart strategies (one with an increasing population size and another with varying small population size) called BI-population-CMAES (BIPOP-CMA-ES) [Han09a, Han09b] has outperformed the classic CMA-ES and other modern CMA-ES versions in the BBOB-2009 function testbed [Han09a, Han09b].

As stated, both DE and CMA-ES variants have shown a superb performance in numerous optimization problems. Thus, the choice between them mainly depends on the problem to tackle and its fitness landscape. For instance, some publications [RKP05, HK04] have shown that DE and its variants face significant difficulties on non-linearly separable functions, and are outperformed by CMA-ES. With regards to their performance in IR problems, both DE and CMA-ES variants have already shown an excellent performance [SDGTC12, DFDCMT08, InBC$^+$09].

Recently, a powerful and versatile RCEA called Coral Reef Optimization with substrate layers (CRO-SL) was proposed in [SSMBV17]. CRO-SL is inspired on the formation and reproduction of coral reefs. CRO-SL simulates the different phases that corals undergo during their lives, such as reproduction, larval settlement, or fight for a space in the reef. Furthermore, CRO-SL simulates the substrate layers in coral reefs. Substrate layers affect to the growth and development of the coral. These layers are modeled by using different exploration operators (e.g. DE search, Gaussian mutation, etc.) on different regions of the coral reef. Their simulation mixes very different exploration operators within the competitive evolution rules of the coral reefs, providing a competitive grid-based co-evolutionary strategy to CRO-SL in just one population. Lastly, CRO-SL also improves the best solution using a local search (LS) method with a limited number of evaluations, making it become a powerful memetic algorithm [OLC10].

There is a lot of controversy with the proposal of new bio-inspired algorithms [Sör15] and their justification must be based on their actual performance beyond the natural metaphor. CRO-SL presents a high novelty since it provides an excellent exploration-exploitation trade-off and robustness as results of the combination of all the previous mentioned features, specially for its competitive environment and the incorporation of multiple search patterns. In addition, CRO-SL usually converges quickly to high quality solutions even in multi-modal seach spaces, being suitable for computationally expensive optimization problems both satisfying quality and computation time constraints. However, its performance varies significantly depending on the CRO's parameters and the different substrates included in the simulated reef. In particular, CRO-SL has outperformed both classical and state-of-the-art evolutionary IR methods in 3D-3D medical IR problems [BCD$^+$18], making it a really promising RCEA for CR with the only drawback of the complex tuning of its parameters.

The best RCEA for solving computationally expensive optimization problems according to the IEEE CEC competitions is the mean-variance mapping optimization (MVMO) optimizer [EVW10]. MVMO has ranked in top positions in expensive optimization competitions, such as IEEE CEC 2013 [RE13], 2014 [ERWS14], 2015

[RE15], 2016 [RTE16], and 2018 [RE18], showing an excellent performance and robustness. MVMO is a novel single-individual RCEA that considers a best solution archive, but its novelty lies within a new mapping function employed for mutating the offspring. This mapping function is based on the mean and variance of the best solution archive. MVMO has been numerically compared to other enhanced RCEAs showing a better performance in many problems, especially in terms of convergence speed. For instance, a powerful variant called MVMO-SH (the "S" refers to the offspring approach based on single parent and multi-parent crossover, and the "H" for the hybridization of MVMO with the use of LS) improves the global search performance of the classical MVMO. MVMO-SH considers a set of solutions (i.e. particles of a swarm) instead of just one, each having its own best solution archive and mapping function, and allows the exchange of information and dynamic reduction of the swarm size.

Motivated by the analysis of the literature, the RCEAs to be studied in this dissertation are as follows: (1) DE; (2) L-SHADE, one of the best self-adaptive variants of DE; (3) CMA-ES, a classic RCEA that has outperformed DE in many problems; (4) BIPOP-CMAES, one of the best modern variation of CMA-ES; (5) CRO-SL, a powerful RCEA that is the state-of-the-art method in 3D-3D IR problems but is complex to finetune; and (6) MVMO-SH, a novel RCEA that has obtained groundbreaking results in many prestigious competitions such as those held within IEEE CEC, especially in costly optimization problems [ERWS14]. They are briefly introduced in the next six subsections.

III.4.1.1 Differential Evolution

DE [SP97] is a variant of an evolutionary strategy [Bey13]. It begins with a random initialization of a population of n candidate solutions. Afterward, DE searches for better solutions by combining the candidate solutions' parameters using a crossover operator along a limited number of generations. The crossover operator combines the parameters of three random candidate solutions from the previous generation (detailed equations can be reviewed in [SP97]). Lastly, DE also has an elitism mechanism which maintains the best candidate solution so far into the next generation. In summary, DE has the following parameters: the population size p, the differential weight F, and the crossover probability P_c.

III.4.1.2 L-SHADE

L-SHADE is a self-adaptive DE approach proposed by Tanabe et al. in 2014 [TF14] based on a previous adaptive DE optimizer called SHADE [TF13]. Its main addition was a linear reduction of the population size (which is initially set to p_{init}) through the generations. L-SHADE maintains the automatic adjust of the F and P_c parameters in each generation of SHADE. To this end, it keeps a historical memory with H entries for both F and P_c. Furthermore, it also conserves its mutation strategy, to-pbest/1, where the greediness is adjustable using a parameter pb, and the use of an external archive for maintaining diversity, its size equal to p_{init} plus r^{arc}. The goal is to adjust the optimizer behaviour during the first generation to promote the search space exploration and subsequently to reinforce its exploitation. To sum up, the parameters to be tuned for L-SHADE are: p_{init}, H, pb, and r^{arc}. Their recommended ranges are reported in [TF14].

III.4.1.3 CMA-ES

CMA-ES [HMK03] has been largely considered as the state of the art in RCEAs and outperforms DE and its variants in many optimization problems, as stated in Section II. CMA-ES is based on updating the covariance matrix of the multivariate normal distribution along the algorithm's generations to focus the exploration on the most promising regions. Afterward, CMA-ES performs the following two steps in each generation: (1) λ candidate solutions are generated according to the multivariable normal distribution, the covariance matrix, and the step size σ; and (2) the distribution centre and the covariance matrix are updated based on the μ best candidate solutions and σ is updated based on the improvement achieved (detailed equations can be reviewed in [HMK03]).

CMA-ES only requires to set three parameters μ, λ (number of best solutions considered to update the distribution center and number of individuals of the population, respectively) and initial step size σ. Their default value in function of the number of decision variables n according to the authors is: $\lambda = 4 + \lfloor 3\ln(n) \rfloor$ and $\mu = \lambda/2$. However, some works have shown that larger λ and a modification of the value of μ can lead to make CMA-ES more robust and/or exploitative on multimodal problems [InBC$^+$09].

III.4.1.4 BIPOP-CMA-ES

BIPOP-CMA-ES [Han09a, Han09b] is a restart CMA-ES with two interlaced restart strategies, that modifies the values of λ and μ in each restart. The first restart strategy consists of increasing the population size λ by a factor of 2. Meanwhile, the second restart strategy involves decreasing the population size λ based on the previous and the default values of λ (detailed equations can be reviewed in [Han09a]). In both restart strategies, the new value of μ is obtained by halving the new value of λ. Performing the first or second restart strategy depends on which restart strategy's budget value is smaller. Nevertheless, the first and last restarts always utilize the first strategy. Lastly, the maximum number of restarts that can be performed is nine. To sum up, BIPOP-CMA-ES requires to set the three same parameters than CMA-ES (λ, μ, and σ). The only difference is that the values of λ and μ given to BIPOP-CMA-ES are only their initial values since they are adapted in each restart.

III.4.1.5 CRO-SL

CRO-SL [BCD$^+$18] is based on natural processes occurring in coral reefs. The coral reef R is represented as a bi-dimensional grid of p positions (population size), where each position stands for solutions to the current optimization problem. At the beginning, p_0 positions (given as a percentage of the total population) are randomly initialized with candidate solutions to the problem tackled while the rest are empty, reserved to allow other corals to grow. For each generation, the following stages will be applied to the coral reef sequentially (these stages are further detailed in [BCD$^+$18]): (1) Broadcast spawning: it consists of generating new larvae from a pair of candidate solutions using a crossover operator; (2) Brooding: new larva are generated via a mutation mechanism that is applied to a fraction of corals 1 - F_b; (3) Larvae setting: each larvae will try to set in a random position of the coral reef, they will only set if it the location is free or the larvae has a better fitness value than

the solution occupying that position; (4) Depredation: a fraction (F_d) of the corals with the worst fitness are removed from the population with very small probability (P_d).

CRO-SL is an extension of the basic algorithm that also simulates the substrate layers in coral reefs. It divides equally the coral reef R into several substrate's layers and the crossover operator of the step 2 will vary depending in which layer the larvae falls. The choice of the crossover operators (or substrate layers) to be used has a significant effect in the optimizer's behaviour. In particular, the crossover operators considered for IR in [BCD+18] are: Harmony search, DE, Gaussian mutation, Cauchy Mutation, SBX, and BLX-α. Furthermore, CRO-SL (as stated in Section II) also has a LS to improve the larvaes with the BOBYQA optimizer [Pow09] using a maximum of n_{LS} evaluations.

To sum up, the parameters to be tuned for CRO-SL are as follows: reef size p, number of coral reef positions initialized p_0, number of generations g, number of LS evaluations n_{LS}, deprecation fraction F_d, deprecation probability P_d, asexual reproduction proportion F_a, mutation fraction F_b, mutation probability, the set of substrate layers utilized, and the parameters from the crossover operators (e.g. F for DE, and δ for harmony search).

III.4.1.6 MVMO-SH

MVMO-SH [RE13] begins with a initialization stage where the p particles (candidate solutions) of the swarm are randomly generated. The particles are normalized to the range [0, 1], which is a necessary condition to the latter mutation via mapping function (a key element in MVMO) and are only des-normalized for their fitness evaluation. Afterward, the following steps are performed for each generation (these are detailed in depth in [RE13]): (1) LS optimization of the particles with a probability p_{LS}; (2) If a particle founds a better solution in terms of fitness than those in its solution archive, the new solution is added to the particle's solution archive (notice that if the archive has reached its maximum size A_s the solution archive's worst solution is removed); (3) Particles are sorted and divided into two groups according to their fitness value, the GP best ones are classified as "good particles" and the rest as "bad particles" (GP is adapted along the process taking values between the 20% and 70% of p). The good particles are modified via a custom single parent crossover operation based on local best [ERWS14] and bad particles via a custom multi-parent crossover operation based on a subset of good particles [ERWS14]; (4) the particles are mutated using a mapping function. This mapping function is based on the mean and variance of each particle's solution archive and a scaling factor f_s that modulates the function's shape. The scaling factor usually begins with a small value f_{start} and progressively increases until reaching its maximum value f_{end} to progressively increase the algorithm's accuracy.

To sum up, the parameters to be tuned are: number of particles p (the recommended value is 15*$number\_variables$. If the number of particles chosen is equal to 1, MVMO-SH will perform as the standard MVMO), LS probability p_{LS}, archive size A_s, scaling factor start (f_{start}) and end values (f_{end}), initial value of the shape of all the variables at the beginning of the optimization d_r (values around 1-5 are suitable to guarantee good initial performance. In practice, it is usually set to 1), and parent selection method (random, neighbor group selection in single step or block steps, or sequential selection of the first variable and the rest randomly).

Part II

Proposal

Chapter IV

Computer-based framework for comparative radiography

'Over and over, we begin again."
— Banana Yoshimoto

This chapter is devoted to present **a novel computer-aided automatic framework for CR-based** forensic identification. The proposed framework tackled the CR-based identification based on the comparison of a 2D AM image (i.e. a radiograph) and a 3D PM image (i.e. a CT or a surface model) with any non-articulable skeletal structure. This framework overcomes all the issues of current state-of-the-art approaches (enumerated in Section I.1 and detailed in Chapter I), while reducing subjectivity and time. It automatically compares the available AM and PM images of skeletal structures and supports the expert in the decision making process in an objective, fast, robust and reproducible manner, shifting from current observational methods to computer-aided ones.

The automation of a CR-based identification procedure can be divided into three consecutive stages (see Fig. 25) related with the different tasks performed in manual approaches (see Chapters I and II):

1. **Skeletal structure segmentation.** The delimitation of the skeletal structures silhouette in 2D and 3D images (not required with 3D surface scans).

2. **AM-PM Overlap.** The goal is to produce a PM radiograph that simulates the scope and projection of each of the AM radiographs.

3. **Decision Making.** Based on the superimpositions achieved, the identification is performed by comparing consistencies and inconsistencies in the bone or cavity morphology, together with other elements such as the quality of the AM radiograph, the visibility of bone or cavity, etc. Notice that, the use of computers aims to support the final identification decision that will always be made by the forensic anthropologist.

IV.1 Stage 1. Image segmentation

In this framework, the identification process is guided by the silhouette of the bone or cavity. Thus, it requires its segmentation within the 2D and 3D images. The

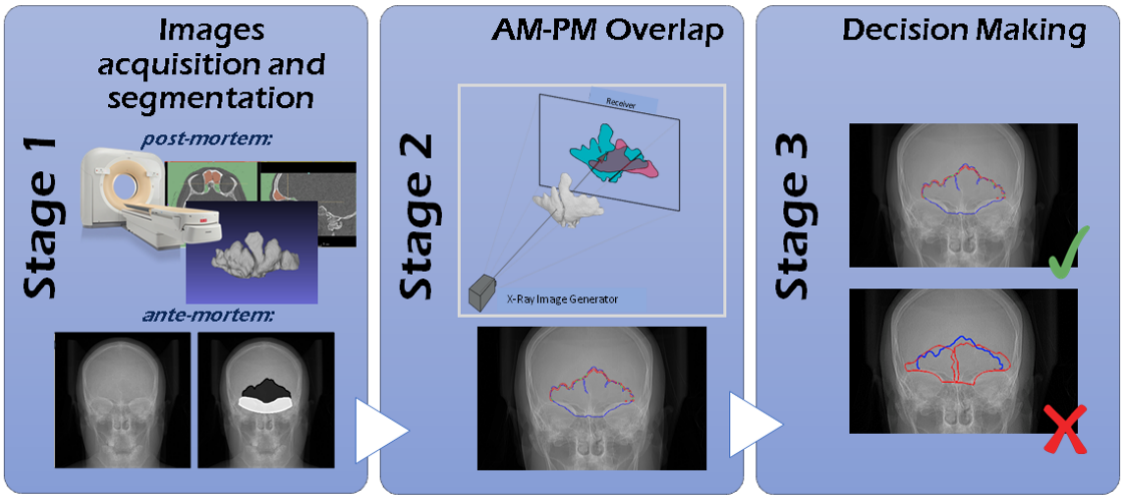

Figure 25: Three stages proposed for automating forensic identification using comparative radiography.

goal is to develop a common IS framework for any skeletal structure in any image. This first stage is required since the intensities are not reliable, because these could have changed between the AM and the PM images (see Chapter I). Besides, its image representation is sensitive to the acquisition device, and some old radiographs show a low quality. Furthermore, intensities are not depicted with 3D surface scans (see Subsection II.1.3). Therefore, automatic superimposition methods should rely on other features as the skeletal structure's silhouette (see Chapter I).

In the segmentation of 3D images, we can distinguish two scenarios depending on the image modality: CTs and 3D surface images. In **CT scans**, bones are directly segmented by thresholding the CT according to the corresponding Hounsfield units. When dealing with cavities like frontal sinuses, a further hindrance has to be addressed: by their nature, cavities can be connected among them and even with the external air. To overcome this problem, the cavity is first isolated by using one or several planes. These planes are horizontal or vertical and must go through a bone landmark (i.e. in frontal sinuses, it is a horizontal plane that goes through a clearly identifiable landmark, called the *crista galli*). Finally, the internal air of frontal sinus is selected by thresholding it with the particular Hounsfield units. **In 3D surface scans,** segmentation is not needed. However, internal cavities such as sinuses cannot be acquired.

Meanwhile, **in the 2D case (radiograph)**, the segmentation of the skeletal structure is more difficult, since its silhouette can be occluded by other structures. Therefore, it is also desirable and informative to segment the region where the skeletal structure is occluded or not clearly defined (e.g. due to fuzzy boundaries), called occlusion region. Furthermore, without the occlusion region, the projection of the 3D PM image under the AM acquisition set-up will be larger than the segmented regions, biasing subsequent tasks. To sum up, two regions have to segmented in radiographs (see figure 26): the silhouette of the skeletal structure and the region of occlusion. This task can be automated using IS techniques (see Section III.1). In particular, the state-of-the-art methods for segmentation of radiographs are **ConvNets** (see Subsection III.2.2).

In our IS framework, a different CNN is trained for the segmentation of each

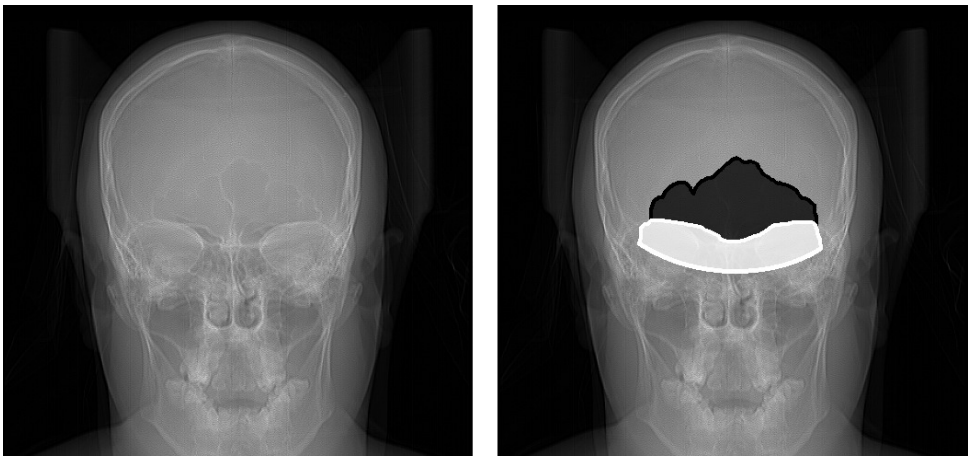

Figure 26: (Left) Skull radiography. (Right) Skull radiography segmented, with the frontal sinuses silhouette in black and the occlusion region in white.

skeletal structure, see Fig. 27. Each CNN can be trained from scratch or from the weights of another network trained to segment another skeletal structure (fine-tuning and transfer learning). Since these are data-driven learning methods, it is necessary a dataset of radiographs, together with the correspondent ground truth (GT) of the target skeletal structure to carry out the learning process. However, most public radiograph segmentation datasets are composed of just a few hundred of radiographs. Thus, the training process cannot only rely on the amount of data but also requires more sophisticated techniques, as ConvNets trainable using small data-sets, data augmentation or few-shot learning. Furthermore, one important goal in radiograph IS is to design a network able to work without any down-sampling or at least to reduce it to the minimum possible. The objective is to avoid upsampling the results since it causes a loss of detail in the final segmentation, which could be crucial for identification purposes.

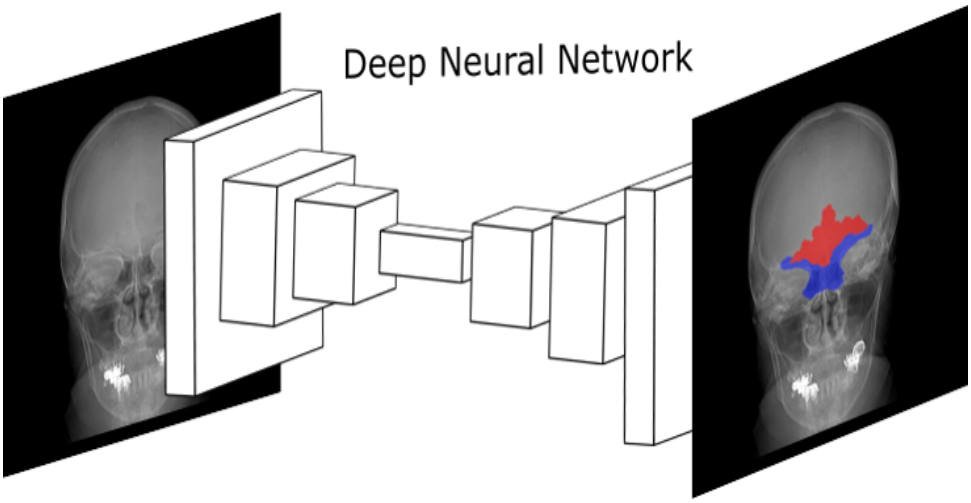

Figure 27: A general scheme of IS framework for automating the stage 1.

IV.2 Stage 2. AM-PM overlay

The goal is to reproduce the acquisition parameters of the AM radiographs and, in consequence, to obtain the best 2D projection of the 3D PM image with respect to the 2D AM image. This superimposition process can be automated using a 3D-2D IR approach [MTLP12, OT14] based on an optimization process, that searches for the best match between the silhouette of a skeletal structure in an AM radiograph and a 2D projection of the 3D PM skeletal structure (either obtained via the segmentation of a PM CT or digitized with a 3D scanner), see Fig. 28. The optimizer cannot assume any parameter value related to perspective or initial pose, since the AM radiographs were acquired in an unknown conditions, where pose and radiograph device are unknown. These requirements make classic 3D-2D IR techniques not suitable for CR (see Section VI.1). Thus, more sophisticated techniques should be considered in order to solve it satisfactorily, such as advanced numerical search methods [Mod04] and evolutionary algorithms [DCS11].

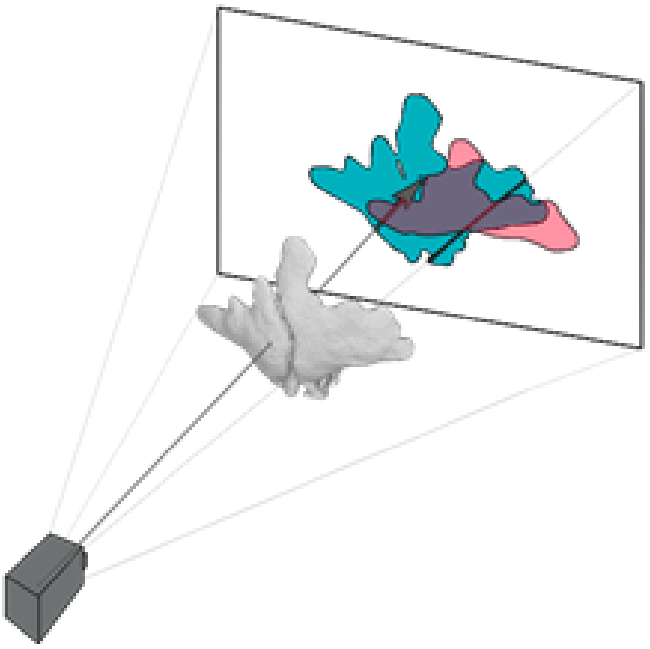

Figure 28: A general scheme of IR framework for automating the stage 2.

IV.3 Stage 3. Decision making

The final goal of the framework is to help forensic practitioners to take decisions based on one or several superimpositions of one or multiple skeletal structures (see Fig. 29). The system will integrate these partial assessments with other aspects affecting the final identification decision as:

- The quality of the bones (preservation) and images (resolution, artefacts, etc.) examined.
- The special characteristics of the bone under study, the uncertainties related to the whole process (e.g. segmentation errors).

- The aggregation of multiple evidences from the same bone (comparison with more than one AM image, assessment of more than one anatomical criterion).

All the previous issues are considered in the decision process followed by forensic experts. Modelling this human reasoning task involves two different problems: knowledge representation under uncertainty and combination of multiple evidences.

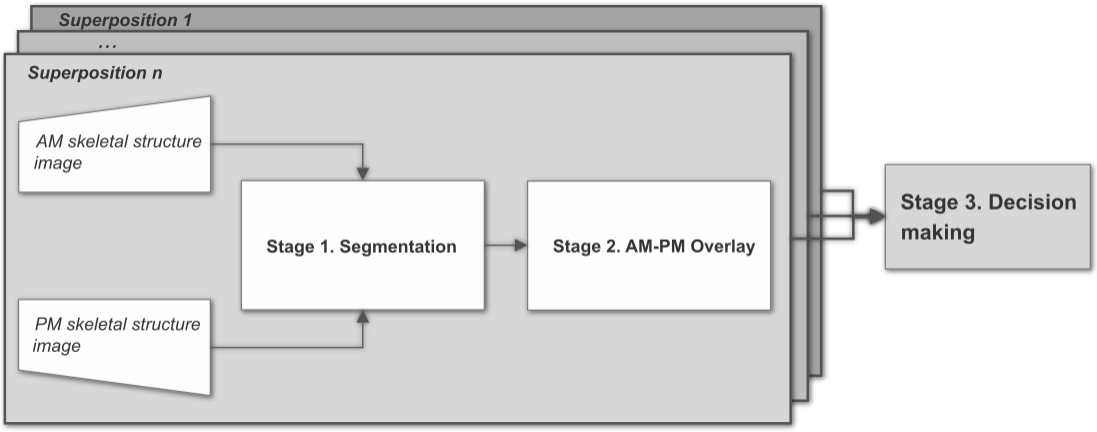

Figure 29: Three stages proposal for automating comparative radiography with one or several superimpositions of one or multiple skeletal structures.

This decision-making stage can be automated using a hierarchical decision support system [CAICW18]. The proposed hierarchical information fusion system comprises the following levels of abstraction (although some of them can be skipped depending on the available data), where the information is fused using aggregation functions [BPC+07] (see Fig. 30):

- **Level 4 (Criteria):** This level analyses a superimposition under a certain criterion using multiple metrics.

- **Level 3 (Superimposition):** It aggregates all the criteria with which a superimposition can be analyzed (performed at level 4), such as morphological differences between AM and PM information. This level also aggregates all the information related to a superimposition, such as the quality of images employed and the visibility of the skeletal structure.

- **Level 2 (Skeletal structure):** It aggregates all the information of all the superimpositions of a skeletal structure (performed at level 3) as well as the quality of the skeletal structures involved in the superimposition (e.g. state of conservation), the discriminatory potential of each skeletal structure, and the presence of special/non-frequent characteristics.

- **Level 1 (Subject):** It aggregates all the information available of the same subject, considering multiple skeletal structures and multiple superimpositions of each of them.

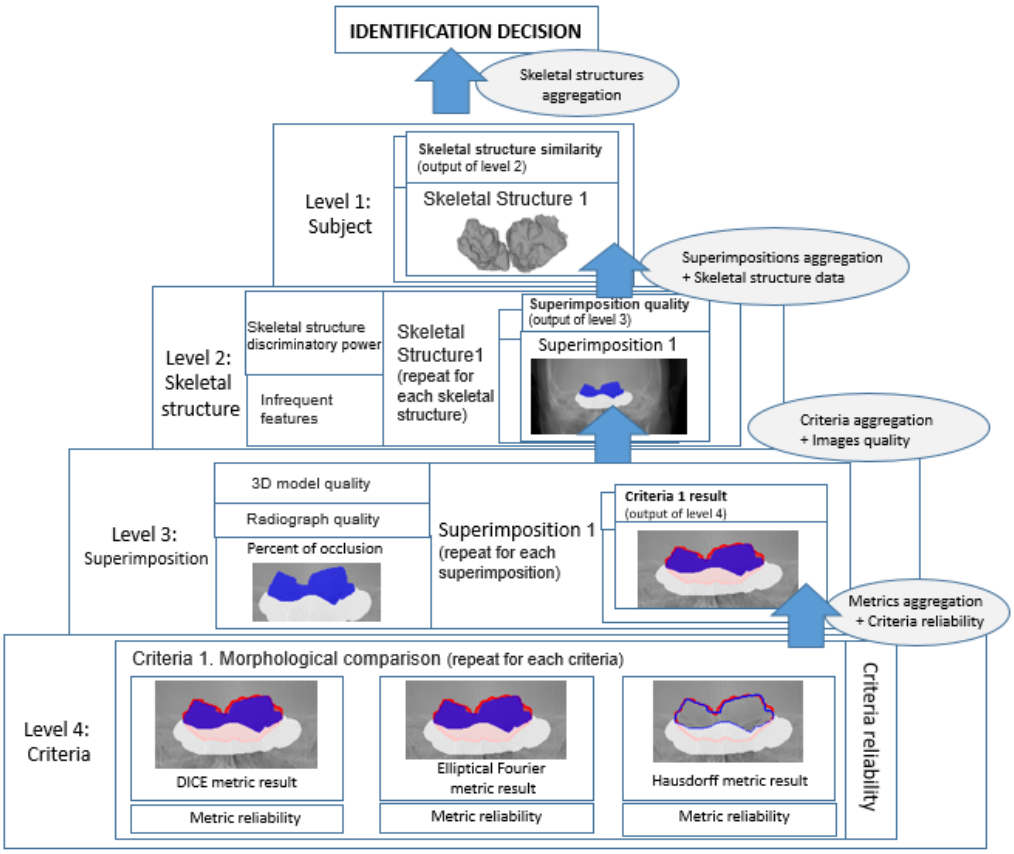

Figure 30: Overview of the 4-level hierarchical decision support system in the 3D-2D scenario. The input of the system is one or more superimpositions of one or more skeletal structures, and the output is the degree of confidence that the AM and PM data belong or do not belong to the same subject.

IV.4 Comparative radiography framework developments addressed in this PhD dissertation

The contributions of this PhD dissertation to the development of this framework, apart from its design, is **the development of all processes involved in achieving the best superimposition of the AM and PM images**. Particularly, we develop and validate an IS framework for **segmenting skeletal structures in radiographs** (stage 1), see Chapter V, as well as an IR framework for **generating PM radiographs simulating the acquisition set-up of the AM-ones**, see Chapters VI and VII. Providing an automatic solution for these two stages is crucial for a wider acceptance of the CR techniques by the scientific community [ARG+10], since generating PM radiographs is the main drawback of manual approaches and the reason why some experts recommend to only use CR-based identification technique as a last resort. Furthermore, a simplified decision support system (based only on one skeletal structure, one superimposition, one criterion, and one metric) is also developed with the purpose of validating the previous developments.

Chapter V

Deep learning for semantic segmentation of skeletal structures

'Words can be like X-rays if you use them properly – they'll go through anything. You read and you're pierced" — Aldous Huxley

V.1 Introduction

In this chapter, we tackle the automatic segmentation of two anatomical structures using convolutional neural networks (ConvNets): i) the clavicle in chest radiographs, including the segmentation of hearts and lungs at the same time as will be explained later; ii) and the frontal sinuses in skull radiographs. To this end, we have proposed a **new ConvNet architecture**, called X-Net. X-Net incorporates structural changes in the state-of-the-art ConvNet architecture for segmenting radiographs, INET [NLM+18], as well as integrates several techniques barely used by radiograph segmentation algorithms, such as instance normalization [UVL16] and atrous convolution (a.k.a. dilated convolution) [CPK+16]. These modifications allow us to improve the segmentation accuracy of the state of the art. Further structural modifications (resulting in an extension termed X-Net+) also allow us to work with images up to 1024×1024 in only one GPU (an example of a segmentation obtained by X-Net+, our best network, is shown in Fig. 31). We have also proposed a simplification of X-Net and X-Net+, called RX-Net (Reduced X-Net) and RX-Net+, respectively, that reduces even more both memory usage and training time, while keeping similar results. Lastly, we have investigated **single-class** (a ConvNet is trained to segment each organ separately) and **multi-class** (a ConvNet is trained to segment all organs simultaneously) segmentation approaches to elucidate which one is more suitable for the task at hand.

This chapter is structured as follows. Section V.2 reviews the current state of the art in radiograph segmentation. Section V.3 describes our proposals. Section V.4 presents experiments and results.

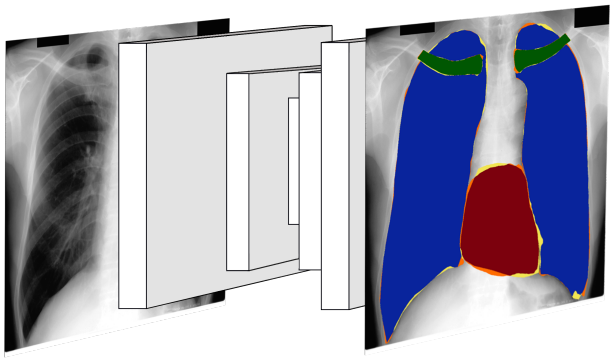

Figure 31: Example of segmentation obtained by X-Net+. The overlap between the ground truth and the segmentation obtained is displayed in green, blue, and red for clavicle, lungs and heart, respectively. The over-segmented area is displayed in orange, and the under-segmented one in yellow.

V.2 Related works

The automatic segmentation of X-ray images has been extensively studied since the '70s, at least regarding the segmentation of lungs, rib cage, heart, and clavicles [TSNF73, WS77]. Nevertheless, just a few works [FFM07, FFM08, TKWK08] have been published, in the late 2000s, for the segmentation of frontal sinuses in radiographs. Conventional methods rely on prior knowledge [ZPW+09] to delineate anatomical objects from X-ray images. Modern approaches utilize deep convolutional networks and have shown superior performance [RFB15]. More than 150 research works dealing with this problem were already published during the twentieth century [VGRV01], raising the number to 331 at present, according to Scopus[1]. Most works have focused on the segmentation of a single organ, mainly the lungs [MHS17, SGG+14, YLL+18] for their medical importance[2]; followed by the heart [BKP14], where most works plainly extrapolate the approaches used for lungs; and lastly the clavicles [HSdJ+12], the organ whose segmentation entails greater difficulty (reflected in a lower quality of the final segmentation [NLM+18]). Despite the great advances made in the automatic segmentation of these organs, limitations still persist, such as the need to use down-sampled radiographs or the irregularity and imprecision of the edges resulting from segmentation, which reduce their applicability in clinical settings.

V.2.1 Automatic segmentation of clavicles in chest radiographs

The Japanese Society of Radiological Technology (JSRT) [SKI+00], in cooperation with the Japanese Radiological Society (JRS), created the standard digital image database with and without chest lung nodules (JSRT dataset) in 1998. Since then, the JSRT dataset has been used by a significant number of researchers in the

[1]Search performed the 8th September 2018 using the keywords (TITLE-ABS-KEY (chest AND X-ray AND segmentation) OR TITLE-ABS-KEY (chest AND radiograph AND segmentation) AND NOT TITLE-ABS-KEY (computed AND tomography))

[2]According to the International Agency for Research on Cancer, lung cancer was the most common cause of cancer death in 2015 with 1.69 million deaths.

world for various research purposes such as image processing, image compression, and computer-aided diagnosis. In particular, the JSRT dataset represents the most popular dataset for chest radiograph segmentation, including high resolution images (2048×2048 size, 0.175mm pixel size) and high resolution segmentation masks (provided by [VGSL06]) of **clavicles**, as well as hearts and lungs. Those segmentation masks have a resolution of 1024×1024 (i.e. ground truth resolution). However, methods tested on JSRT are commonly evaluated using images of smaller resolution (256×256).

The state of the art in lungs segmentation with the JSRT dataset [CYRC18] is based on a hybrid approach with four stages devoted to: 1) preprocessing the X-ray images by augmenting the contrast between the lungs and their surrounding area; 2) extracting the foreground (which incorporates the upper torso region) by using an intelligent block-based binarization; 3) excluding lung regions from the foreground through a series of spatial-based processing operations; and 4) employing an adaptive graph cut technique to locally refine the preliminary lung boundaries.

On the other hand, the state of the art in hearts [BFK18] and clavicles [NLM+18] segmentation is based on deep learning approaches. In particular, chest radiograph segmentation methods based on ConvNets have outperformed prior state-of-the-art methods based on classical techniques. Firstly, an encoder-decoder network called U-Net [RFB15] was studied for multi-class segmentation of lungs, heart and clavicles achieving comparable, or higher, accuracy on most of the structures when compared with the state-of-the-art segmentation methods [Wan17]. They also studied the differences between multi-class and single-class training approaches, showing that for U-Net a multi-class approach helps the deep neural network to converge faster and deliver better segmentation results on the clavicles than the single-class. However, network outputs present holes inside the targeted structures as well as artifacts (i.e. small isolated segmented areas), which were solved with a post-processing step based on a level-set method. Afterwards, another work proposed a small modification of U-Net, called LF-SegNet [MHS18]. LF-SegNet modified the up-sampling strategy, incorporated normalization techniques such as batch normalization [IS15], and employed data augmentation, slightly improving the performance on lungs segmentation in both the JSRT dataset and the Montgomery dataset [JCA+14] (that includes 138 chest radiographs and ground truth only for the lungs). Very recently, several articles tackled the segmentation of chest radiographs employing fully ConvNets [HMS18, Wan17]. Besides, a generative adversarial network approach called dual-path adversarial learning (DAL) based on a hybridization of a fully convolutional network and U-Net [BFK18] was proposed for different kinds of medical IS problems. DAL was tested for the segmentation of lungs and hearts, trained with images of 512 × 512 and evaluated on images on the ground truth resolution 1024 × 1024 resulting in the state of the art for heart segmentation.

A work closely related to ours was published by Novikov et al. [NLM+18] in 2018, where a modification of U-Net, called InvertedNet (INET), segmented the three organs and achieved state-of-the-art results for clavicle segmentation. INET outperformed prior state-of-the-art methods by reducing the number of filters per convolutional layer, therefore decreasing the possibility of over-fitting. Furthermore, motivated by the Gaussian noise inherited from the X-ray images acquisition process, INET added Gaussian dropout layers [SHK+14] (see Subsection III.2.3 for further details about Gaussian dropout) and utilized a weighted loss function based on

the Dice Similarity Coefficient (DICE) [Sør48]. This new form of dropout amounts to adding a Gaussian distributed random variable with zero mean and standard deviation equal to the activation of the unit. INET only considers the multi-class approach and does not compare with training the network for only one class. Finally, it can only be used with down-sampled images (with a resolution of 256×256 pixels) and the main option pointed by the authors to operate with higher resolution images was the use of a multi-GPU scenario, unlike in this work where we opted for the modification of the network architecture.

V.2.2 Automatic segmentation of frontal sinuses in skull radiographs

There are no public repositories of X-ray images including the ground truth segmentation of frontal sinuses. As a consequence, little research [FFM07, FFM08, TKWK08] has been performed for automating the segmentation of frontal sinuses in radiographs. The first two works [FFM07, FFM08] are semi-automatic methods requiring the selection of several initial seed pixels within the frontal sinuses by the forensic expert. From these seed pixels, the whole sinuses are segmented using a graph-based method called differential image foresting transform (DIFT) [FB04]. The DIFT method calculates the minimum path forests in the graph derived from the image, restricting the search to paths originated from the seeds. The method was tested in a dataset composed of 90 skull radiographs, but the results were reported in terms of the identification capability of the segmented frontal sinuses, i.e. equal error rate, instead of in terms of the segmentation accuracy. Meanwhile, the latter work [TKWK08] is based on an ad-hoc rule-based segmentation method. This method is composed of three consecutive stages: detection of the region of interest, detection of the bottom border of the frontal sinus, and detection of the top border of the frontal sinus. This method automates these stages using prior knowledge of the frontal sinuses anatomy, connectivity-preserving thresholding, and on watersheds. The method was tested in a dataset composed of 50 skull radiographs but the results were only qualitatively analyzed.

V.2.3 Room for improvement

To the best of our knowledge, there are no other works in the literature that apply the atrous convolution [CPK+16] to the segmentation of radiographs, while it is one of the key elements of the state-of-the-art network for IS in general [CPSA17]. Atrous convolution is a convolutional operation that introduces a spacing between the values in a kernel (the number of spaces between values is called dilated rate). This allows us to adjust filter's field-of-view and capture multi-scale context information without reducing the spatial dimensions of the features maps (i.e. a 3×3 kernel with a dilation rate of 2 will have the same field-of-view as a 5×5 kernel, while only using 9 parameters). However, this is computationally expensive and takes a lot of memory, as a consequence its use is normally preceded by a few pooling layers to make the feature maps computationally approachable as in DeepLab [CPSA17]. Besides, we are not aware of other works studying the compression/simplification of deep networks, devoted to medical image segmentation, for their deployment in single-GPU devices or to allow them to work with larger images. Many re-

searchers have pointed out that ConvNets suffer from heavy over-parameterization and can be efficiently reconstructed with only a small subset of its original parameters [DSD+13]. Therefore, several works have been published studying the simplification/compression of ConvNets reducing the required resources without significant loss in the original accuracy [HLM+16, KPY+15]. Furthermore, we also perform a comparison between multi-class and single-class approaches using a k-fold cross-validation protocol, which is a much more rigorous evaluation strategy than the one commonly employed in the deep learning literature (where generally a simple hold-out is used).

V.3 Proposals

V.3.1 Architectures

The deep architectures proposed in this chapter are inspired by INET [NLM+18] (described in Section V.2 and depicted in Fig. 32) which, in turn, is a modification of U-Net. INET is devoted to the segmentation of lungs, hearts and clavicles in chest radiographs, and represents the state of the art in clavicle segmentation.

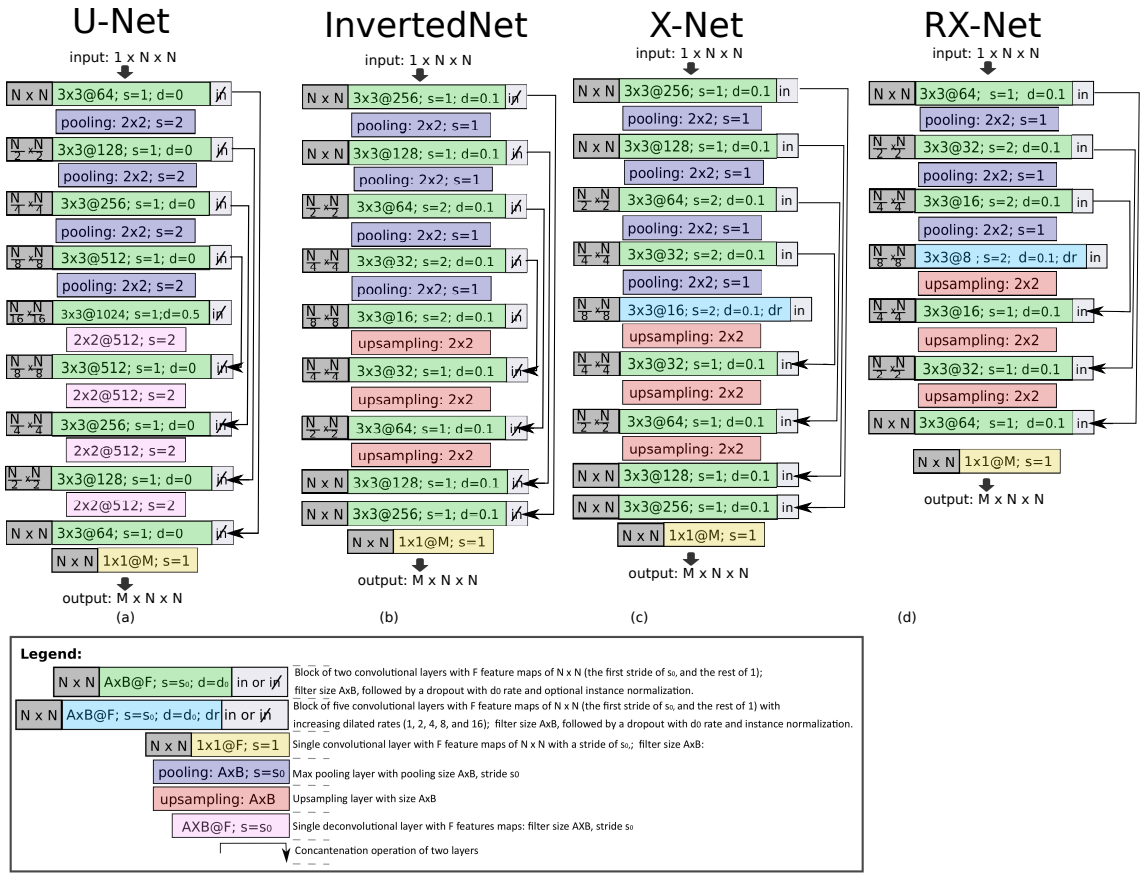

Figure 32: Schematic view comparing two preceding deep networks (a and b), and two of the deep networks proposed in this chapter (c and d). All architectures employ Gaussian dropout, except U-Net that uses conventional dropout. 'in' stands for 'instance normalization'.

Even if INET currently offers the best performance in chest radiographs multi-class segmentation, it does not include some very relevant advances made in the IS

research field (atrous convolution) and deep learning in general (instance normalization). Consequently, the first methodological contribution of this chapter is the introduction of a new architecture, called X-Net, that aims to increase the accuracy of INET through the inclusion of these advances in order to develop the automatic segmentation of some skeletal structures useful to perform CR-based identification. X-Net takes its name both from being designed to segment X-ray images and from the shape of the network (where, as usual, there is an encoding stage, which provides a reduced dimensionality representation of the input, and a decoding stage, that allows to recover an output of equal size to the input). First, X-Net takes advantage of instance normalization [UVL16] at the end of each convolutional layer to add a normalization factor, accelerate the training, improve generalization, and reduce the dependency on the weights initialization. Second, X-Net changes the two central convolutional layers of INET by five atrous convolutional layers with increasing dilated rates of 1, 2, 4, 8 and 16 (see Fig. 32). These specific dilated rates are the most commonly employed in recent literature [LHL+18, LCC+18, GLL+18]. Atrous convolutions are convolutions with upsampled filters that allow us to enlarge the field-of-view and, therefore, to take into account more contextual information. The combined use of atrous convolution and instance normalization leads to a greater gain in performance than when they are employed separately (detailed in Section V.4.4).

The second proposal introduced in this chapter, termed Reduced X-Net (RX-Net), consists of a simplification of X-Net with the aim of obtaining similar accuracy but with a significantly lower memory usage and training time. Importantly, the main source of memory usage is not the trainable parameters, but instead the feature maps. Reducing the number of features maps of resolution $N \times N$ will result in a large memory reduction. RX-Net represents one alternative that makes possible experimenting with images of higher resolution than the one generally used, as we will discuss in the next paragraph. Therefore, the simplification involves the elimination of the first and last convolutional blocks of X-Net (notice that the first and last layers have the largest activation maps), and the reduction to half the number of convolutional filters of each convolutional layer (see Fig. 32), since these changes lead to the larger reduction in the ConvNet memory usage with the smaller drop in accuracy.

One important goal in chest radiographs image segmentation is to design a network able to work without any down-sampling or at least to reduce it to the minimum possible. The objective is to avoid upsampling the results (or diminish its impact) since upsampling causes a loss of detail in the final segmentation. In this sense, our next proposals are able to manage images of up to 1024×1024 in a single GPU (see subsection V.4.3 for the technical details) without any down-sampling. Thus, these proposals are able to work with the ground truth resolution [VGSL06] of JSRT [SKI+00]. INET and X-Net would require too much memory or a multi-GPU scenario, being only able to work with images up to 256×256 (that is the resolution in which INET results, and most of the results in literature, are reported [NLM+18]). Our RX-Net can handle images up to 1024×1024, four times the usual resolution, but changing the input resolution results in a change in the relation between filter's field-of-view and feature maps, which would lead to a different behaviour than the one showed by RX-Net with images of 256×256. To avoid this drawback, a new architecture called RX-Net+ is proposed. This network is an incremental step from

RX-Net maintaining all their layers (except the last convolutional block) with the same resolution (i.e. the value $N/4$ of RX-Net+ is equal to 256).

It allows us to employ the weights resulting from training RX-Net with 256×256 in RX-Net+ for images of 1024×1024, which significantly reduces the total training time. Additionally, RX-Net+ adds two pooling layers at the beginning (they could be replaced by convolutional layers but, if the number of feature maps introduced as input to the pre-trained RX-Net block is different to one, it would make impossible to re-use the weights from RX-Net in RX-Net+) and two final convolutional blocks connected to the inputs of the first two pooling layers. Furthermore, performing the pooling within the ConvNet allows to pass high-resolution information to the final layers of the ConvNet (see Fig. 32). Notice that comparing RX-Net and RX-Net+ for images of 1024×1024 will allow us to study the importance of the relation between the filter's field-of-view and the feature maps. This approach could also be applied to X-Net giving rise to X-Net+. X-Net+ combines all the advantages of X-Net as well as allows to work with images of 1024×1024 in just one GPU. Training time and accuracy for both X-Net+ and RX-Net+ are improved thanks to training X-Net and RX-Net, respectively, then re-using the central block of common weights, and finally employing a simple fine-tuning. We will refer to our proposed deep networks as X-Net architectures, since they all are based on X-Net.

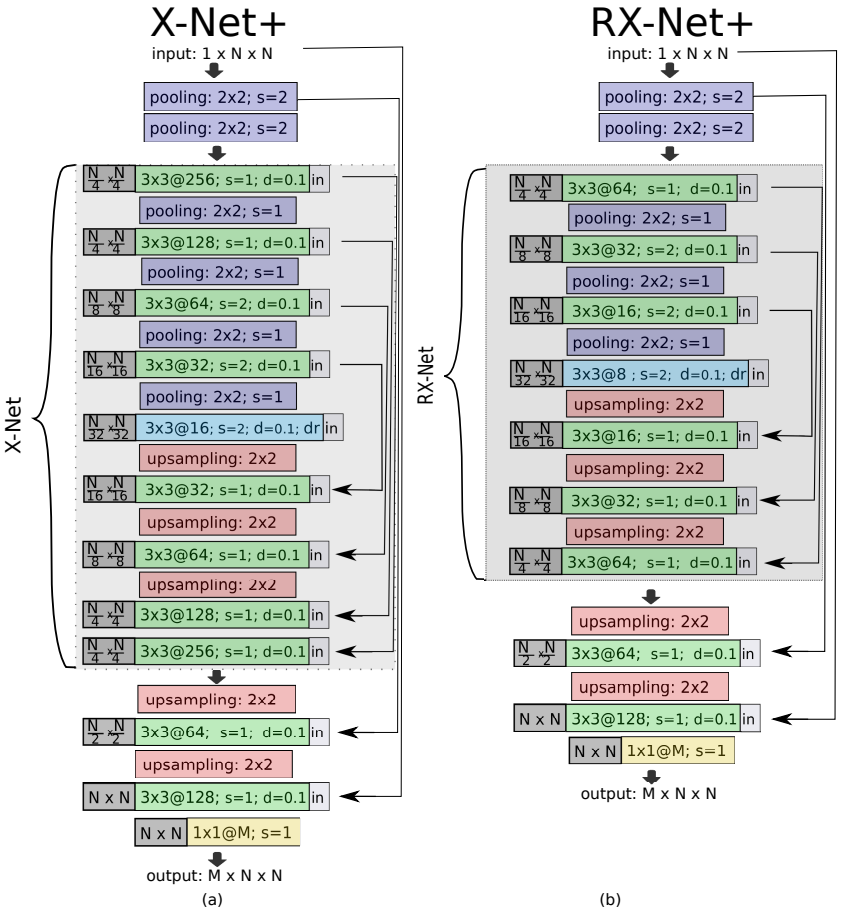

Figure 33: Schematic view of two of the deep networks proposed. These networks are extension of X-Net and RX-Net, respectively, allowing us to handle ground truth resolution images (i.e. 1024×1024) in just one GPU. The legend of this figure can be seen in Fig. 32.

V.3.2 Training strategies

Two strategies are compared to train all networks: a single-class approach (to train a network to segment only one organ, i.e. three different networks are required to segment the three organs) and a multi-class approach (to train a network to jointly segment the three organs). The loss function in the single-class approach is directly the usual DICE [Sør48] (see Section V.4.2) and for the multi-class segmentation we employ a balanced version of this measure, defined as the product of the DICE values obtained for each single organ, i.e. DICE $= \prod_{i=1}^{n} DICE_i$, where n is the number of classes to segment, and $DICE_i$ is the DICE value for the segmentation of class i (corresponding to each organ to segment). This loss function allows us to deal with the imbalanced nature of the chest radiographs segmentation problem (for instance, in the JSRT dataset ground truth [VGSL06], the 73.53%, 21.85%, and 4.62% of image pixels on average belong to lungs, hearts and clavicles, respectively) looking for solutions that properly segment the three classes. This loss function is stricter than others employed in the state of the art (for instance, the weighted mean in [NLM+18]) because we intend to encourage solutions that segment properly the three organs. INET has been trained using the weighted mean as loss function as in the original work [NLM+18].

V.3.3 Post-processing

The predictions provided for each pixel by the neural architectures range from 0 to 1. To turn this soft classification into a binary mask, it is necessary to threshold the output at a given value. In this work, we use the same threshold value (0.25) as in [NLM+18] where this value was fixed empirically based on a preliminary experimentation. This output does not ensure the presence of a single connected object for the heart and two for clavicles and lungs. Therefore, a last and very simple post-processing step is considered: the largest connected object is selected for the hearts, and the two largest ones for the clavicles and lungs (notice that this post-processing step is a simplified version of the one proposed by [VGSL06] since it does not fill the holes within the object). The algorithm employed for this task was the Block-Based Decision Table algorithm [GBC10] with 8-way connectivity. This very simple post-processing step has a minor but positive impact, as can be seen in Table 8, and it does not differ from other simple post-processing strategies employed [VGSL06, SVZ13, LST+16, QSMG17].

V.4 Experiments

The empirical evaluation of this work includes three experiments, the first two for segmenting chest radiographs and the last one for segmenting skull radiographs:

- The first experiment is devoted to the study of performance, precision, robustness, and the trade-off between accuracy and memory/time consumption of the X-Net architectures and INET with both single-class and multi-class training approaches for segmenting clavicles, as well as hearts and lungs. This study is performed using a 3-fold cross validation protocol as in [NLM+18].

- The goal of the second experiment is the comparison of the X-Net architectures performance (except the worst performing architectures from the latter experiment, that are excluded from the comparison) with the state-of-the-art results using a 10-fold cross validation to avoid any bias caused by the stochastic components of training a ConvNet for segmenting clavicles, hearts and lungs.

- Lastly, the third experiment is devoted to study the accuracy, robustness and memory/time consumption of the X-Net architectures for segmenting frontal sinuses and occlusion regions in skull radiographs. This experiment is performed using a 3-fold cross validation protocol, and again single-class and multi-class training strategies are compared.

It is important to highlight the computational cost of performing the experimentation following a rigorous experimental design in deep learning (cross validation). Overall, around 1608 hours (67 days) were required to perform Experiment I, around 1840 hours (77 days) were necessary to run Experiment II, and around 960 hours (40 days) were necessary to run Experiment III. Detailed information about training times are included in Sections V.4.5, V.4.6, and V.4.7.

V.4.1 Data

The dataset employed in the experiments I and II is the JSRT dataset [SKI+00]. It is the most widely used dataset in chest radiographs segmentation. This dataset is composed of 247 chest radiographs of 2048×2048 pixels with a grayscale depth of 12 bits. These images contain manual/ground-truth segmentations of the lungs, clavicles, and heart[VGSL06] with a resolution of 1024×1024 pixels, where $\sim$73%, $\sim$5%, and $\sim$22% of pixels belong to lungs, clavicles, and hearts, respectively.

Meanwhile, the dataset employed in experiment III is comprised by 234 skull radiographs of 512×512 pixels with a grayscale depth of 12 bits, provided by the *Hospital de Castilla-La Mancha*, Spain. The frontal sinuses and the occlusion region were manually segmented by two trained forensic anthropology MSc students (Andrea Cerezo Vallecillo and José Manuel Pérez Jiménez) from the Physical Anthropology lab (PAL) of the University of Granada.

V.4.2 Performance metrics

Three metrics are employed to quantitatively evaluate the quality of the segmentation results obtained: HD [BTE98], JI [Jac12], and DICE [Sør48]. The HD represents a measure of the spatial distance between two sets of points: it is the largest of all distances from any point in the resulting segmentation to the closest point in the ground truth and a value of 0 indicates perfect agreement. Meanwhile, the DICE and the JI measure set agreement: a value of 0 indicates no overlap with the ground truth while a value of 1 indicates perfect agreement. Notice that, DICE and JI are equivalent metrics, and one can be derived from the other. Thus, for the comparison between X-Net architectures only the HD and JI will be reported. However, in order to facilitate the comparison with competitor methods (Experiment II), the DICE is also included in the tables.

Our final goal is to be able to segment radiographs in the original resolution of their ground truth segmentation (i.e. 1024×1024), and not in a down-sampled resolution, because up-sampling to the ground truth resolution will worsen the accuracy of the final segmentation. As a consequence, all results are reported in the ground truth resolution, either if they correspond to the ConvNet output (e.g. X-Net+ and RX-Net+) or to the up-sampled version of it (e.g. INET, X-Net and RX-Net).

V.4.3 Experimental set-up

The first experiment involves the application of **6 deep network configurations** (INET and X-Net can only run using the 256×256 resolution, RX-Net with 256×256 and 1024×1024 resolutions, and lastly X-Net+ and RX-Net only with the 1024×1024 resolution), **and two training strategies** (single-class and multi-class approaches). INET, X-Net and RX-Net are trained from scratch for 4000 epochs (since that was the number of epochs required by INET to converge according to [NLM$^+$18]). X-Net+ and RX-Net+ are trained for 100 epochs using as initialization the weights of X-Net and RX-Net in the shared layers, respectively. These are tested using a **3-fold cross validation** approach (as in INET [NLM$^+$18]), where one fold is devoted to testing (33% of all available data), and each one of the remaining two folds is divided into training and validation (90% and 10%, respectively). Furthermore, the results are evaluated with and without post-processing to measure the contribution of this refinement step. To sum up, 12 deep networks are evaluated (see Table 3), rising up to 24 architectures if we include results with and without post-processing (see Table 4). The notation employed to refer to each model uses the following labeling protocol: *<Network Name>_<Single(s) or Multi(m) organ problem>_<Input/Output resolution>*. As an example, the architecture INET_m_256 corresponds to INET trained to solve the multi-class problem on 256×256 images.

The second experiment involves the comparison of INET and the best proposals from the latter experiment in terms of accuracy (X-Net+ for single-class and multi-class) and accuracy-memory/time balance (RX-Net+) with a **10-fold cross validation**, where on each fold 80% of data are used for training, 10% for validation, and 10% for test. This allows us to study more rigorously the proposals reducing possible bias, as remarked in [LMAPH18], caused by the stochastic effect inherent to the training process or the effect that different training and test sets have on the final performance. The results obtained by X-Net architectures are compared among them and with the state-of-the-art methods (see Tables 6, 7, and 8).

Both experiments (Sections V.4.5 and V.4.6) include the results provided by our implementation of INET. This allows us to replicate the exact same experimental conditions in all methods and, therefore, to perform a fair comparison. The only exception is Table 8, dedicated to the comparison with the state of the art, where we show the original results reported on each paper. The difference between the results provided by our implementation of INET and the results reported in [NLM$^+$18] is minor, as can be seen by comparing the results in the original paper with Tables 3 and 6, and can be due to several reasons: from differences in the partitions employed in the 3-fold and 10-fold, respectively; differences in the batch size employed (as theirs is not reported in the chapter); or just the inherent stochastic behavior of training a network from scratch.

The third experiment involves the application of the two best network con-

figurations, X-Net+ and RX-Net+, for **frontal sinuses segmentation** using the original resolution, 512×512. These ConvNets are trained using two strategies, single-class and multi-class approaches, and the same protocol employed in the previous experiments. They are tested using a 3-fold cross validation approach, where one fold is devoted to testing (33% of all available data), and each one of the remaining two folds is divided into training and validation (90% and 10%, respectively). Furthermore, the results are evaluated with and without post-processing to measure the contribution of this refinement step. To sum up, eight networks are evaluated (see Tables 9 and 10), rising up to sixteen architectures if we include results with and without post-processing.

No data augmentation is performed in Experiments I and II. Meanwhile, data augmentation operations (flip, rotation and zoom) are employed in Experiments III, due to the smaller size of the dataset. Particularly, for each training image, five new images are generated using random rotations, between -15°and 15°. For each of the resulting images and the original one, five new images are generated using zoom or unzoom with zero padding, between 85% and 115%. Finally, all images are flipped. As a result, the size of the training dataset is multiplied by 50. The images are zero centered, as in [NLM+18], using the mean and standard deviation of the training set. The batch size was set to 1. The optimizer is Adam (with a learning rate of 1e-5, beta1 of 0.9, and beta2 of 0.999). The outputs of lower resolution than the ground truth (i.e. 1024×1024) are scaled using a bicubic interpolation, since it showed better results than the other alternatives tested, although the gap between the best and worst interpolation was lower than 0.001 according to the JI.

All experiments have been performed on an Nvidia Titan X with 12 GBs of memory using Keras 2.1.6 with TensorFlow 1.4.1 as backend.

V.4.4 Preliminary Experiment: Evaluating the influence of architectural changes on INET and post-processing for segmenting chest radiographs

The purpose of the first preliminary experiment is to measure the influence of the different architectural changes introduced on INET to obtain X-Net. The results of this ablation study are shown in Table 1. The best results are obtained by instance normalization together with atrous convolution, being both sources of improvement. However, we can claim that instance normalization has a greater contribution to this improvement. Instance normalization introduces some noise into the network, helping to improve its generalization ability. We hypothesize that, since we have at our disposal a small dataset, this noise inducing process contributes to enforce regularization and, therefore, to improve the results obtained.

A second preliminary experiment was performed to measure the impact of the post-processing step (see Section V.3.3). The post-processing step (see Table 2) has shown to improve the results according to both JI and HD, providing statistically significant differences according to Wilcoxon's rank sum test [Geh65] ($9.8 \cdot 10^{-90}$ for JI; and 0 for HD). On average, the JI improves from 0.895 to 0.899, and the HD from 86.699 to 35.069. This simple post-processing step is important for metrics that focus on the quality of the final contours (like HD), since removing the artifacts allows for a better comparison of the error in the boundaries of the segmented organs. We want to highlight that, even if the post-processing has a positive impact on the

Table 1: Summary of the preliminary experiments according to the average JI and HD of the three organs to study the influence of architectural changes on INET without post-processing. 'in' and 'ac' stand for 'instance normalization' and 'atrous convolution', respectively.

Network	**JI**			
	mean	**sd**	**min**	**max**
INET	0.885	0.012	0.685	0.944
INET + ac	0.892	0.013	0.721	0.953
INET + in	0.910	0.007	**0.832**	0.963
X-Net (INET + in + ac)	**0.925**	0.007	0.797	**0.967**
Network	**HD**			
	mean	**sd**	**min**	**max**
INET	132.37	72.38	100.33	255.27
INET + ac	128.27	68.10	99.03	239.50
INET + in	124.65	84.30	**97.54**	226.27
X-Net (INET + in + ac)	**121.68**	62.32	98.27	**188.73**

final result, almost all X-Net architectures without post-processing yield a better performance than our implementation of INET with post-processing.

Table 2: Summary of the preliminary experiments according to the average JI and HD of the three organs to study the influence of the post-processing.

Network	Without post-processing		With post-processing	
	JI mean	HD mean	JI mean	HD mean
INET	0.876	132.522	0.883	43.143
X-Net	**0.905**	91.653	**0.908**	**31.006**
RX-Net	0.899	**60.599**	0.899	34.400

V.4.5 Experiment I: Comparison of X-Net architectures and INET with single-class and multi-class strategies for segmenting chest radiographs

The results obtained for the single-class and multi-class strategies are shown in Table 3, employing JI and HD as evaluation metrics. The first conclusion worth mentioning is that single-class training strategies generally outperform multi-class strategies for chest radiographs segmentation. There are statistically significant differences in favor of the former with p-values, according to the Wilcoxon's rank sum test [Geh65] of 0.02 for the JI, and $6.5 \cdot 10^{-20}$ for HD. In particular, single-class approaches obtain the best segmentation results for clavicles and lungs, while the best results on hearts are obtained by a multi-class approach. Thus, despite multi-task learning has shown useful in other problems [ZLLT14, WVBWK15], its use must be studied for each particular problem. Finally, the comparison of the results of RX-Net and RX-Net+ for images of 1024×1024, i.e. RX-Net_m_1024 and RX-Net+_m_1024, provides support about the fact that the relation between the filter's field-of-view and the feature maps affects significantly to the performance. Since this simple post-processing has shown to be beneficial, all results of X-Net architectures and our implementation of INET include it (see Tables 3, 4, 6, and 7).

We rank the performance of the X-Net architectures, as well as our implementation of INET, in Table 4 according to JI and HD. Methods with a difference in performance smaller than 0.0025 and 5 for JI and HD, respectively, are considered

Table 3: Summary of results evaluated using JI and HD per architecture and organ. All X-Net and INET variants are included.

Network	Organ\Metric	JI					HD				
		mean	std	median	min	max	mean	std	median	min	max
INET_m_256	Clavicles	0.833	0.015	0.843	0.639	0.905	22.390	10.721	20.180	6.031	72.764
	Heart	0.869	0.024	0.894	0.511	0.955	50.245	32.378	40.464	13.867	195.971
	Lungs	0.951	0.006	0.957	0.842	0.972	56.795	40.207	42.804	13.176	229.373
INET_s_256	Clavicles	0.862	0.017	0.876	0.635	0.931	20.375	12.672	17.825	5.846	88.730
	Heart	0.866	0.032	0.896	0.350	0.961	51.971	33.839	40.764	13.403	185.146
	Lungs	0.949	0.007	0.957	0.821	0.974	49.071	37.883	35.783	11.362	207.247
X-Net_m_256	Clavicles	0.876	0.013	0.887	**0.701**	0.934	18.518	9.538	16.383	5.025	64.325
	Heart	0.890	0.014	0.905	**0.751**	0.963	38.317	18.297	34.434	11.369	**93.118**
	Lungs	0.961	0.004	0.965	0.903	0.976	36.183	25.384	27.083	10.599	**140.922**
X-Net_s_256	Clavicles	0.855	0.013	0.867	0.682	0.919	19.151	11.812	16.434	6.640	89.206
	Heart	0.889	0.017	0.905	0.654	0.969	37.501	18.953	34.505	10.295	100.593
	Lungs	0.959	0.004	0.961	0.892	0.976	36.909	29.552	**25.281**	10.343	176.254
X-Net+_m_1024	Clavicles	0.883	0.015	0.894	0.686	0.949	18.468	10.761	15.818	4.824	67.755
	Heart	**0.892**	0.014	0.903	0.735	0.965	37.732	19.318	33.318	10.889	118.317
	Lungs	**0.963**	0.004	**0.967**	**0.908**	**0.980**	38.352	27.830	27.611	10.181	144.794
X-Net+_s_1024	Clavicles	**0.885**	0.016	**0.896**	0.630	**0.953**	18.022	11.241	15.941	4.824	87.660
	Heart	0.890	0.016	**0.907**	0.706	**0.970**	**37.207**	20.542	**32.514**	**8.872**	107.929
	Lungs	**0.963**	0.004	**0.967**	0.896	**0.980**	**36.100**	29.485	26.605	**7.912**	184.837
RX-Net_m_256	Clavicles	0.860	0.017	0.874	0.661	0.929	20.047	10.479	17.202	7.066	70.187
	Heart	0.889	0.015	0.905	0.738	0.961	38.007	18.074	34.350	13.150	99.823
	Lungs	0.955	0.005	0.961	0.889	0.976	45.146	30.254	36.291	11.338	167.700
RX-Net_s_256	Clavicles	0.869	0.017	0.881	0.606	0.934	18.049	10.101	16.058	**4.667**	70.237
	Heart	0.883	0.016	0.898	0.704	0.963	40.068	19.971	36.872	11.312	105.527
	Lungs	0.959	0.004	0.963	0.899	0.976	38.694	29.625	27.335	11.105	168.505
RX-Net_m_1024	Clavicles	0.855	0.023	0.880	0.548	0.942	22.872	11.765	19.616	6.535	66.767
	Heart	0.874	0.020	0.894	0.657	0.967	46.055	26.915	39.795	13.631	155.417
	Lungs	0.951	0.007	0.959	0.823	0.976	51.204	35.061	41.329	14.524	190.915
RX-Net_s_1024	Clavicles	0.866	0.019	0.880	0.642	0.949	21.762	13.535	18.989	5.878	95.639
	Heart	0.855	0.032	0.880	0.367	0.961	49.850	31.894	41.383	12.801	182.362
	Lungs	0.953	0.006	0.961	0.869	0.976	49.069	35.060	37.430	11.656	190.086
RX-Net+_m_1024	Clavicles	0.867	0.018	0.880	0.612	0.940	19.472	9.751	16.962	7.333	60.711
	Heart	0.889	0.015	0.903	0.726	0.961	37.330	18.653	33.224	11.646	105.080
	Lungs	0.955	0.005	0.961	0.881	0.978	44.751	30.678	34.837	11.600	167.060
RX-Net+_s_1024	Clavicles	0.880	0.016	0.892	0.646	0.946	**17.728**	9.301	**15.695**	5.277	**53.971**
	Heart	0.883	0.017	0.896	0.686	0.965	40.472	21.172	35.936	11.200	119.962
	Lungs	0.961	0.004	**0.967**	0.894	0.978	38.596	29.747	26.927	10.051	169.883

Table 4: Average ranking position of X-Net architectures and INET per organ and metric (JI, HD, and their average) using 3-fold cross validation [NLM+18]. Two networks are considered equal if the difference in performance between them is lower than 0.0025 for JI and 5 pixels for HD.

Network\Metric	Clavicles			Lungs			Hearts			3 organs		
	JI	HD	Aver.	JI	HD	Aver.	JI	HD	Aver.	JI	HD	Aver.
X-Net+_s_1024	**2.2**	**5.3**	**3.8**	**4.8**	**2.8**	**3.8**	**2.8**	**4.2**	**3.5**	**3.3**	**4.1**	**3.7**
X-Net+_m_1024	**2.2**	**5.3**	**3.8**	**4.8**	3.8	4.3	**2.8**	**4.2**	**3.5**	**3.3**	4.4	3.9
X-Net_m_256	3.3	**5.3**	4.3	**4.8**	**2.8**	**3.8**	**2.8**	**4.2**	**3.5**	3.7	**4.1**	3.9
RX-Net+_s_1024	2.8	**5.3**	4.1	**4.8**	5.2	5.0	6.7	5.5	6.1	4.8	5.3	5.1
X-Net_s_256	9.7	**5.3**	7.5	**4.8**	3.8	4.3	4.7	**4.2**	4.4	6.4	4.4	5.4
RX-Net_s_256	6.3	**5.3**	5.8	**4.8**	5.2	5.0	7.0	5.5	6.3	6.1	5.3	5.7
RX-Net+_m_1024	6.8	**5.3**	6.1	6.7	7.3	7.0	5.2	**4.2**	4.7	6.2	5.6	5.9
RX-Net_m_256	8.7	**5.3**	7.0	6.2	8.7	7.4	4.0	**4.2**	4.1	6.3	6.1	6.2
RX-Net_s_1024	6.8	7.7	7.3	6.2	9.5	7.8	11.7	10.3	11.0	8.2	9.2	8.7
INET_s_256	7.5	7.0	7.3	9.5	9.0	9.3	10.3	11.0	10.7	9.1	9.0	9.1
RX-Net_m_1024	9.7	11.2	10.4	11.3	10.2	10.8	9.3	9.7	9.5	10.1	10.3	10.2
INET_m_256	12.0	9.5	10.8	9.2	9.7	9.4	10.7	11.0	10.8	10.6	10.1	10.3

equivalent. This ranking does not show the values of JI and HD, but the average position of each network for a given metric and organ. Thus, the values of the ranking goes from 1 to the number of networks, and smaller values are associated with a better performance. All our proposals outperform INET (even the reduced ones which require lower resources than INET), with INET being the worst performing approach in the comparison. It is important to remember that INET is the current state-of-the-art approach in multi-class chest radiographs segmentation. Another important conclusion is that, generally, ground truth resolution approaches (1024×1024) outperform downsampled approaches. In particular, the best method in all rankings is X-Net+ in ground truth resolution using a single-class training approach (see Fig. 34 for some segmentation examples), with X-Net+_m_1024 being the second best performing approach.

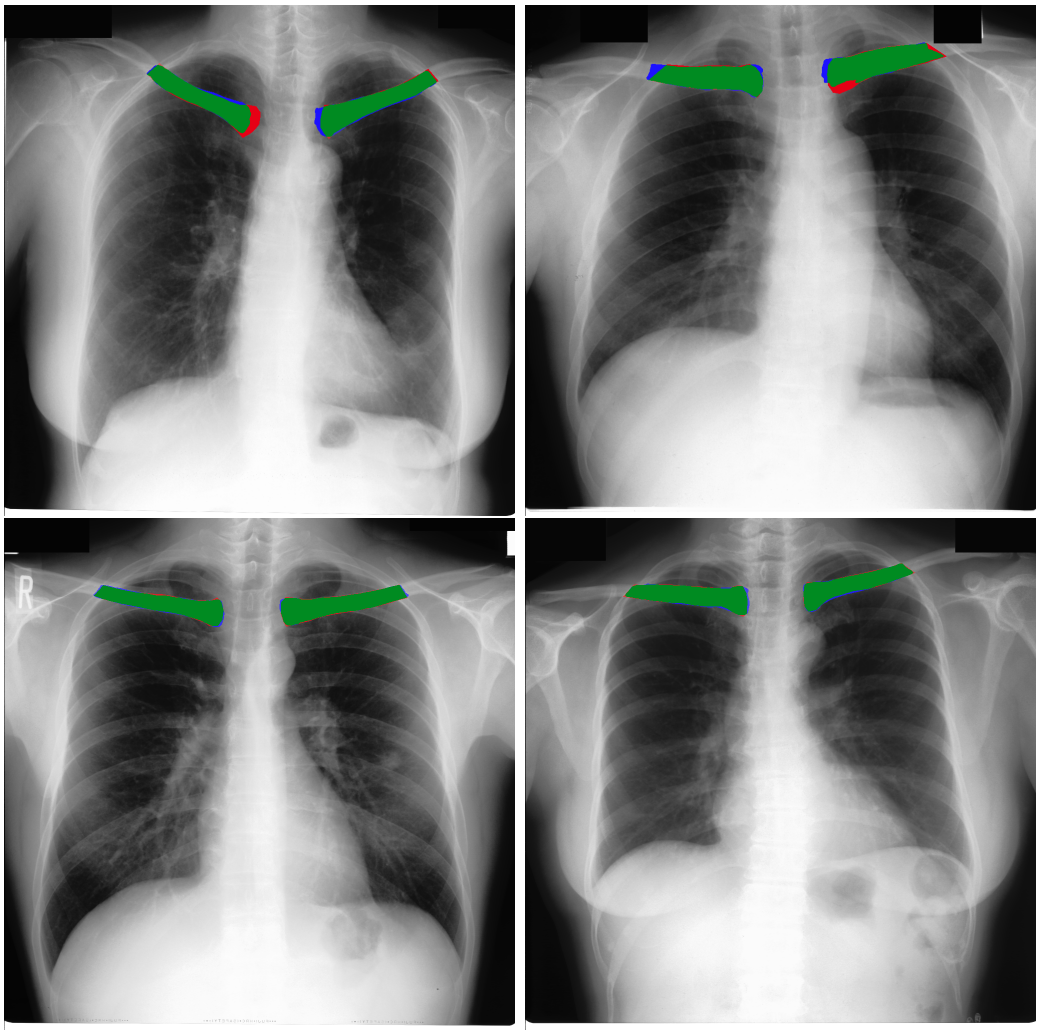

Figure 34: Examples of segmented clavicles obtained by X-Net+. In green, area correctly segmented. In red, areas that were, erroneously, not segmented. In blue, over-segmented areas. (Top) Examples of segmentations with errors of around 0.84 and 0.91 for JI and DICE, respectively. (Bottom) Examples of segmentations with errors around 0.92 and 0.96 for JI and DICE, respectively.

The time required to train INET was about 26 hours per run (i.e. 26 hours for multi-class approach and 78 hours for single-class since three networks are trained), and 9 GBs of memory are necessary (for both the single-class and multi-class con-

Table 5: Summary of the average time and memory requirements of X-Net architectures and INET with both the mono-class and multi-class approaches.

Network	**Muti-class time (h)**	**Single-class time (h)**	**GPU memory (GB)**
INET (256 × 256)	26h	26h×3=78h	9GB
INET (1024 × 1024)	Cannot be trained due to its GPU memory requirements.		
X-Net (256 × 256)	36h	36h×3=108h	9GB
X-Net (1024 × 1024)	Cannot be trained due to its GPU memory requirements.		
X-Net+ (1024 × 1024)	36h+3h =39h	(36h+3h)×3=117h	12GB
RX-Net (256 × 256)	12h	12h×3=36h	3.5 GB
RX-Net (1024 × 1024)	55h	55h×3=165h	11GB
RX-Net+ (1024 × 1024)	12h+2h =14h	(12h+2h)×3=42h	10GB

figuration). X-Net requires 36 hours and 9 GBs to train, while the finetuning of X-Net+, from the X-Net weights, takes only 3 hours (for a total of 39 hours), requiring almost 12GBs of GPU memory. RX-Net requires only 12 hours and 3.5 GBs with images of 256 × 256, and 55 hours and 11 GBs with images of 1024 × 1024. Meanwhile, RX-Net+ with images of 1024×1024 takes only 2 hours to be finetuned from the weight of RX-Net (256) (i.e. a total of 14 hours) and 10 GBs. Thus, RX-Net outperforms INET in accuracy but also reduces the required memory and the training time (see Table 5 for a summary of the time and memory requirement of all the architectures). Overall, around 1608 computational hours (67 days) were required to perform the 3-fold cross validation.

Given that X-Net, the proposal that is closest to INET, is better than INET (see rankings of Table 4, where X-Net_s and X-Net_m are systematically ranked above their INET counterparts), we can conclude that the modifications introduced in X-Net are responsible for such improvement. Thus, the use of atrous convolution and instance normalization to improve the results is highly recommended.

V.4.6 Experiment II: Comparison with state-of-the-art approaches for segmenting chest radiographs

The results obtained by the best X-Net architectures employing 10-fold cross validation are shown in Table 6. All those results include post-processing. The results of INET [NLM+18] correspond to our implementation, in order to perform a comparison as rigorous as possible with the same 10-folds. The comparison of Tables 3 and 6 shows that the results obtained have not changed significantly from the 3-fold to the 10-fold cross validation protocol. X-Net architectures are robust to different initialization and training-test subsets. This is supported by the unchanged positions of the different proposals in the ranking showed in Table 7. Lastly, the Nemenyi test [HW99] was performed to look for statistically significant differences between the best ranked proposal, X-Net+_s_1024, and all the other networks. The test showed that there is no statistical significant difference with X-Net+_m_1024 with p-values larger than 0.1 for both and HD. Therefore, both X-Net+_s_1024 and X-Net+_m_1024 must be considered the best performing approaches. More specifically, when employing JI and DICE as evaluation metrics, X-Net+_s_1024 is better for clavicles, and X-Net+_m_1024 for lungs and hearts. However, X-Net+_s_1024 becomes also the best method in clavicles when HD is considered. The Nemenyi test finds statistically significant differences with all the other networks with a p-value always lower than $1 \cdot 10^{-06}$ for both JI and HD. In particular, for INET_m_256,

Table 6: Summary of JI and HD results per architecture and organ employing a 10-fold cross validation protocol to the best performing architectures in Experiment I (see Section V.4.5).

Network	Organ\Metric	JI					HD				
		mean	std	median	min	max	mean	std	median	min	max
INET_m_256	Clavicles	0.835	0.018	0.850	0.663	0.905	21.937	10.946	19.732	8.218	55.271
	Heart	0.850	0.033	0.883	0.546	0.944	64.605	41.183	50.472	20.807	178.663
	Lung	0.953	0.005	0.959	0.885	0.972	51.679	36.886	41.390	13.676	163.012
X-Net_m_256	Clavicles	0.848	0.017	0.862	0.667	0.910	20.507	12.043	17.734	7.519	60.869
	Heart	0.881	0.019	**0.901**	0.663	0.951	**42.179**	23.456	36.361	14.808	119.029
	Lung	0.951	0.005	0.957	0.890	0.972	47.416	35.574	35.761	13.287	170.676
X-Net_s_256	Clavicles	0.859	0.013	0.869	**0.748**	0.912	18.381	9.789	16.133	7.826	**51.733**
	Heart	0.878	0.017	0.892	**0.718**	0.955	45.603	24.703	40.355	14.842	110.086
	Lung	0.951	0.006	0.957	0.880	0.972	46.333	37.479	32.960	**11.806**	177.618
X-Net+_m_1024	Clavicles	0.874	0.018	0.889	0.678	0.940	20.361	11.891	17.336	7.587	59.583
	Heart	**0.883**	0.018	0.898	0.698	**0.959**	42.638	27.027	**35.422**	13.773	132.061
	Lung	**0.957**	0.005	**0.961**	**0.898**	**0.976**	46.477	34.094	35.544	13.321	**161.642**
X-Net+_s_1024	Clavicles	**0.883**	0.015	**0.896**	0.745	**0.944**	**18.357**	11.567	**15.795**	**6.595**	60.463
	Heart	0.878	0.018	0.896	0.714	0.957	43.671	24.269	37.844	**13.342**	**107.652**
	Lung	0.955	0.006	**0.961**	0.881	0.976	**46.248**	37.529	**32.567**	12.189	174.856
RX-Net_m_256	Clavicles	0.838	0.018	0.852	0.645	0.905	21.515	12.103	18.920	8.449	61.395
	Heart	0.876	0.019	0.896	0.682	0.951	44.508	23.747	38.192	16.442	117.139
	Lung	0.947	0.006	0.951	0.866	0.969	53.350	35.161	44.273	17.195	162.182
RX-Net_s_256	Clavicles	0.845	0.017	0.860	0.685	0.908	19.882	11.984	17.109	7.620	61.342
	Heart	0.864	0.020	0.878	0.684	0.946	50.112	30.619	42.176	16.232	147.602
	Lung	0.947	0.007	0.955	0.860	0.970	51.975	42.228	38.710	13.080	198.489
RX-Net+_m_1024	Clavicles	0.864	0.019	0.881	0.653	0.934	20.718	11.953	17.965	7.962	62.627
	Heart	0.880	0.017	0.896	0.717	0.953	43.526	23.499	37.026	15.560	111.432
	Lung	0.951	0.006	0.957	0.873	0.972	51.912	37.211	41.463	16.028	172.276
RX-Net+_s_1024	Clavicles	0.871	0.019	0.890	0.676	0.942	19.404	12.283	16.332	7.292	63.905
	Heart	0.866	0.020	0.880	0.689	0.949	50.412	31.395	43.437	16.247	144.748
	Lung	0.925	0.012	0.940	0.779	0.969	79.469	45.211	69.918	21.169	208.965

Table 7: Average ranking position of the best X-Net architectures and INET per organ and metric (JI, HD, and their average). Two networks are considered equal if the difference in performance between them is lower than 0.0025 for JI and 5 pixels for HD. A 10-fold cross validation protocol is applied to the best performing architectures in Experiment I (see Section V.4.5).

Network\Metric	Clavicles			Lungs			Hearts			3 organs		
	JI	HD	Aver.	JI	HD	Aver.	JI	HD	Aver.	JI	HD	Aver.
X-Net+_s_1024	**1.8**	**3.2**	**2.5**	4.4	3.7	4.0	3.9	**3.5**	**3.7**	3.3	**3.5**	**3.4**
X-Net+_m_1024	2.5	5.4	3.9	**3.8**	**2.8**	**3.3**	**3.3**	4.1	**3.7**	**3.2**	4.1	3.6
X-Net_s_256	4.8	**3.2**	4.0	4.1	3.9	4.0	4.0	4.9	4.4	4.3	4.0	4.1
X-Net_m_256	6.5	5.4	5.9	4.1	3.4	3.8	4.0	3.7	3.8	4.8	4.2	4.5
RX-Net+_m_1024	4.1	6.4	5.2	4.5	5.6	5.0	4.5	4.4	4.4	4.3	5.4	4.9
RX-Net_s_256	6.7	5.2	5.9	5.2	5.5	5.4	6.5	6.2	6.4	6.1	5.6	5.9
RX-Net_m_256	7.8	6.0	6.9	6.3	6.2	6.2	4.9	4.3	4.6	6.3	5.5	5.9
RX-Net+_s_1024	2.8	3.4	3.1	8.7	8.5	8.6	6.4	6.0	6.2	6.0	6.0	6.0
INET_m_256	8.3	7.0	7.7	4.1	5.5	4.8	7.7	8.1	7.9	6.7	6.9	6.8

it obtains a p-value of $5.2 \cdot 10^{-15}$ for JI and $1.2 \cdot 10^{-12}$ for HD.

The time required to train a fold of all architectures with the two training approaches is lower since only 3000 epochs are performed instead of the 4000 from the previous experiment, and also the number of X-Net architectures compared is lower. Nevertheless, the computational time needed to tackle this experimentation is significantly higher because a 10-fold cross validation were performed, and thus *circa* of 1840 computational hours (i.e. 77 days) were required.

Table 8 shows the comparison of our best X-Net-based architectures using a 10-fold cross validation protocol, with and without post-processing, and including state-

of-the-art approaches. JI and DICE represent the results reported in the trained resolution (indicated by the number in parentheses at the end of the name of the method). JI Full and DICE Full report the results in the original resolution of the segmentation mask (i.e. 1024×1024). DICE and DICE Full (as well as JI and JI Full) have the same value for methods trained with the original image resolution. Notice that all approaches report better results in the down-sampled resolution than in the ground truth resolution. The reason is that the resulting segmentation is evaluated more roughly, and thus we lose details and nuances. Since the segmentation results are always more precise in the ground truth resolution, we employ it as reference to highlight in bold the performance of the different algorithms under comparison. Results without the post-processing step are also reported to allow a fair comparison with "pure" deep learning methods. As a conclusion, X-Net+ provides better results in the ground truth resolution than all the other methods in the state-of-the-art (9 competitor approaches) for clavicles and lungs. It also yields comparable results with the state-of-the-art method [BFK18] for heart segmentation (with a difference in performance smaller than 0.01 (JI) and 0.005 (DICE)). X-Net+ also outperforms the human observer in lungs and hearts (see Table 8). Importantly, X-Net+ without post-processing yields comparable results to X-Net+ with post-processing.

Table 8: Comparison of our best X-Net-based architectures, with and without post-processing, with state-of-the-art approaches. The best results in the ground truth resolution, 1024×1024, are displayed in bold per organ and metric. Cells containing a "—" represent either that the proposed method does not tackle the segmentation of the organ, or that the results at the original or down-sampled resolution are not reported. Values calculated from other metric, where only one of them was reported, are marked with a "*".

	Clavicles				**Lungs**				**Hearts**			
Method	**JI**	**JI Full**	**DICE**	**DICE Full**	**JI**	**JI Full**	**DICE**	**DICE Full**	**JI**	**JI Full**	**DICE**	**DICE Full**
Human observer [VGSL06]	—	0.896	—	0.945*	—	0.946	—	0.972*	—	0.878	—	0.935*
TVC2018 (512) [BFK18]	—	—	—	—	—	0.951	—	0.975	—	**0.893**	—	**0.943**
WPC2018, a.k.a. LF-SegNet (224) [MHS18]	—	—	—	—	0.951	—	0.975*	—	—	—	—	—
WPC2018-2, a.k.a. FCN (224) [HMS18]	—	—	—	—	0.959	—	0.979*	—	—	—	—	—
JBHI2018 (256) [YLL+18]	—	—	—	—	0.952	—	0.975	—	—	—	—	—
MP2017 (256) [XSM+17]	—	—	—	—	0.955	—	0.977	—	—	—	—	—
N2018 (256) [CYRC18]	—	—	—	—	0.963	0.948	0.983	0.974	—	—	—	—
MIA2012 (256) [HSdJ+12]	0.860	—	0.925*	—	—	—	—	—	—	—	—	—
SCIA2017 (256) [Wan17]	0.863	—	0.926*	—	0.959	—	0.979*	—	0.899	—	0.947*	—
TMI2018, a.k.a. INET (256) [NLM+18]	0.868	—	0.929	—	0.950	—	0.974	—	0.882	—	0.937	—
X-Net+_m_1024 without post-proc.	0.871		0.931		0.954		0.976		0.879		0.935	
X-Net+_m_1024	0.874		0.933		**0.956**		**0.978**		0.884		0.938	
X-Net+_s_1024 without post-proc.	0.880		0.936		0.954		0.976		0.863		0.927	
X-Net+_s_1024	**0.883**		**0.938**		0.955		0.977		0.879		0.935	
RX-Net+_m_1024 without post-proc.	0.859		0.924		0.948		0.973		0.876		0.934	
RX-Net+_m_1024	0.864		0.927		0.951		0.975		0.880		0.936	
RX-Net+_s_1024 without post-proc.	0.867		0.929		0.924		0.960		0.849		0.919	
RX-Net+_s_1024	0.870		0.931		0.925		0.961		0.865		0.928	

V.4.7 Experiment III: Tackling the segmentation of frontal sinuses in skull radiographs

The results obtained for the X-Net architectures are shown in Table 9 employing DICE, JI and HD as evaluation metrics. The reported results according to DICE and JI metrics are not comparable to the reported results in the clavicle segmentation problem. Apart from the differences related to the employment of a different skeletal structure, the underlying reason is that there is not an anatomical limit between the frontal sinuses and the occlusion region. In fact, the occlusion region contains the continuation of the frontal sinuses by definition (see Chapter IV). As a consequence, these limits vary significantly between the predicted and GT segmentations, since no method can possibly detect the GT limit. Nevertheless, the HD metric is robust to this problem, allowing to study both the quality of the upper area of the frontal sinuses segmentations, and the presence of gaps between the frontal sinuses and the occlusion region. The best performing method is again X-Net+ (see Fig. 35 for some segmentation examples). However, in contrast to chest radiographs, multi-class training strategies generally (X-Net+_m_1024) outperform single-class strategies (X-Net+_s_1024) for head radiographs segmentation. The rationale behind this fact is that X-Net+_m_1024 barely produces gaps between the frontal sinuses and the occlusion region, while X-Net+_s_1024 cannot deal satisfactorily with them. Thus, it is necessary to study multi-task learning for each particular problem, even between similar segmentation problems. Furthermore, all single-class methods rank better than their multi-class counterparts (see Table 7). Lastly, the rest of conclusions remain unchanged: (1) the simple post-processing step is beneficial; and (2) the greater the image resolution, the better the obtained results.

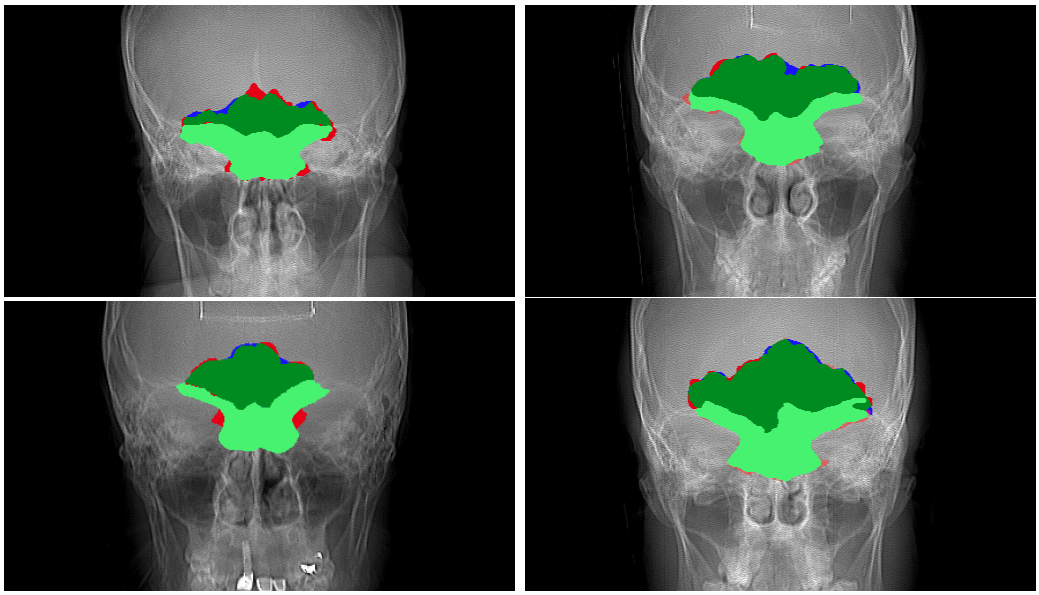

Figure 35: Examples of segmented frontal sinuses obtained by X-Net+. In green, area correctly segmented (frontal sinuses in dark green, and occlusion region in light green). In red, areas that were, erroneously, not segmented. In blue, over-segmented areas. The errors related to limit between the frontal sinuses and the occlusion regions are not reported in these images since these are misleading as previously stated. (Top) Examples of segmentations with errors of around 0.66 and 0.79 for JI and DICE, respectively. (Bottom) Examples of segmentations with errors around 0.69 and 0.84 for JI and DICE, respectively.

Table 9: Comparison of the X-Net-based architectures, with and without post-processing, for segmenting frontal sinuses and occlusion regions in head radiographs.

Network	Object\Metric	Without post-processing					
		JI		DICE		HD	
		Mean	Std	Mean	Std	Mean	Std
RX-Net_m_128	Frontal Sinuses	0.564	0.082	0.721	0.152	67.383	387.932
RX-Net_m_128	Occlusion Regions	0.643	0.051	0.783	0.098	20.801	15.693
RX-Net_s_128	Frontal Sinuses	0.570	0.074	0.726	0.137	28.136	28.033
RX-Net_s_128	Occlusion Regions	0.649	0.042	0.787	0.081	21.540	15.964
RX-Net+_m_512	Frontal Sinuses	0.648	0.083	0.787	0.154	107.455	535.785
RX-Net+_m_512	Occlusion Regions	0.690	0.057	0.817	0.107	22.787	27.841
RX-Net+_s_512	Frontal Sinuses	0.656	0.074	0.792	0.139	28.014	34.417
RX-Net+_s_512	Occlusion Regions	0.692	0.048	0.818	0.092	21.376	16.962
X-Net_m_128	Frontal Sinuses	0.582	0.073	0.735	0.135	30.410	34.792
X-Net_m_128	Occlusion Regions	0.639	0.050	0.780	0.094	27.948	39.624
X-Net_s_128	Frontal Sinuses	0.587	0.069	0.740	0.130	28.700	29.917
X-Net_s_128	Occlusion Regions	0.646	0.048	0.785	0.092	34.494	46.622
X-Net+_m_512	Frontal Sinuses	**0.667**	0.069	**0.801**	0.129	29.728	38.711
X-Net+_m_512	Occlusion Regions	**0.699**	0.047	**0.823**	0.090	**19.286**	12.149
X-Net+_s_512	Frontal Sinuses	0.653	0.071	0.790	0.133	**24.510**	18.577
X-Net+_s_512	Occlusion Regions	0.692	0.052	0.818	0.099	24.231	27.970
Network	Object\Metric	With post-processing					
		JI		DICE		HD	
		Mean	Std	Mean	Std	Mean	Std
RX-Net_m_128	Frontal Sinuses	0.564	0.083	0.722	0.152	65.299	382.655
RX-Net_m_128	Occlusion Regions	0.643	0.052	0.783	0.099	19.564	9.163
RX-Net_s_128	Frontal Sinuses	0.572	0.073	0.728	0.137	22.582	12.126
RX-Net_s_128	Occlusion Regions	0.649	0.043	0.787	0.082	20.113	10.052
RX-Net+_m_512	Frontal Sinuses	0.649	0.085	0.787	0.156	105.490	535.104
RX-Net+_m_512	Occlusion Regions	0.689	0.058	0.816	0.110	18.203	9.600
RX-Net+_s_512	Frontal Sinuses	0.657	0.075	0.793	0.139	21.077	13.364
RX-Net+_s_512	Occlusion Regions	0.691	0.049	0.817	0.094	19.275	9.746
X-Net_m_128	Frontal Sinuses	0.585	0.073	0.738	0.136	22.341	11.075
X-Net_m_128	Occlusion Regions	0.642	0.049	0.782	0.094	19.487	9.215
X-Net_s_128	Frontal Sinuses	0.589	0.073	0.741	0.136	22.944	14.609
X-Net_s_128	Occlusion Regions	0.649	0.047	0.787	0.089	20.564	9.486
X-Net+_m_512	Frontal Sinuses	**0.670**	0.069	**0.802**	0.129	**20.799**	13.142
X-Net+_m_512	Occlusion Regions	**0.699**	0.048	**0.823**	0.092	**18.016**	9.436
X-Net+_s_512	Frontal Sinuses	0.655	0.071	0.792	0.133	21.296	12.507
X-Net+_s_512	Occlusion Regions	0.690	0.057	0.817	0.107	19.625	13.311

Table 10: Average ranking position of the best X-Net architectures per segmentation object and metric (JI, HD, and their average). Two networks are considered equal if the difference in performance between them is lower than 0.0025 for JI and 5 pixels for HD.

Network\Metric	Frontal Sinuses			Occlusion Regions			All		
	JI/DICE	HD	Aver.	JI/DICE	HD	Aver.	JI/DICE	HD	Aver.
X-Net+_m_512_post	**2.0**	**3.5**	**2.8**	**3.5**	**3.5**	**3.5**	**2.8**	**3.5**	**3.1**
RX-Net+_s_512_post	4.8	**3.5**	4.2	5.2	5.0	5.1	5.0	4.3	4.6
X-Net+_s_512_post	4.8	5.3	5.1	4.3	5.3	4.8	4.6	5.3	5.0
RX-Net+_m_512_post	5.5	7.3	6.4	4.7	**3.5**	4.1	5.1	5.4	5.3
X-Net+_m_512	**2.0**	13.7	7.8	**3.5**	5.7	4.6	**2.8**	9.7	6.2
RX-Net+_m_512	6.0	7.7	6.8	4.7	12.3	8.5	5.3	10.0	7.7
X-Net+_s_512	4.8	9.3	7.1	5.7	10.8	8.3	5.3	10.1	7.7
RX-Net+_s_512	6.0	11.0	8.5	4.5	9.8	7.2	5.3	10.4	7.8
X-Net_m_128_post	11.2	**3.5**	7.3	14.2	6.5	10.3	12.7	5.0	8.8
X-Net_s_128_post	11.3	6.0	8.7	11.5	9.3	10.4	11.4	7.7	9.5
RX-Net_s_128_post	14.0	6.5	10.3	11.5	6.7	9.1	12.8	6.6	9.7
RX-Net_m_128_post	14.0	9.7	11.8	12.8	6.5	9.7	13.4	8.1	10.8
RX-Net_m_128	14.0	12.0	13.0	12.8	9.3	11.1	13.4	10.7	12.0
RX-Net_s_128	14.0	12.0	13.0	11.5	11.0	11.3	12.8	11.5	12.1
X-Net_s_128	10.3	12.3	11.3	11.5	16.0	13.8	10.9	14.2	12.5
X-Net_m_128	11.2	12.7	11.9	14.2	14.7	14.4	12.7	13.7	13.2

Chapter VI

Evolutionary image registration for 3D-2D skeletal structure's silhouette overlay

'Nothing was your own except the few cubic centimetres inside your skull." — George Orwell

VI.1 Introduction

The superimposition process for CR is complex. This complexity has its origin on multiple factors such as the unknown set-up of the AM radiograph or the fact that image intensities are not reliable or even not captured. Most 3D-2D IR approaches are designed for a controllable set-up, which is a common situation in many medical domains [RRM+05, SGW+12, JBVH+06]. Therefore, they can assume a calibrated case, with only 6 degrees of freedom (DoF), where the parameters related to the perspective distortions are known, and with a initial pose close to the GT pose (i.e. an error of around 20 mm in translation and 20°in rotation in [JBVH+06], a maximum target registration error [vdKPT+05] of 16 mm in [RRM+05], etc). However, these assumptions are not suitable for CR since the AM radiograph is generally taken under unknown conditions (neither the pose nor the radiograph device are known in advance). Therefore, the search for the optimal solution in the CR scenario is more challenging. Of course, there are a few exceptions such as Feldman et al. [FAB95], which proposed a 3D-2D IR method based on the silhouette that does not rely on assumptions about the initial pose by using free-form curves and surfaces. However, this is only applicable in the calibrated case (6 DoF). To the best of our knowledge, there is no other 3D-2D IR approaches based on the silhouette considering 7 DoF without these constraints.

The objective of this work is twofold. First, to propose and validate a novel computer-aided paradigm, based on a 3D-2D IR feature-based approach (second stage), for the superimposition of the silhouettes of a 3D PM model of any bone or cavity and an AM radiograph. This is validated with synthetic images of two bones (clavicles and patellae) and one cavity (frontal sinuses). Second, to study how optimization performance and both variability and differences in the segmentation

performed by human operators, affect the identification using synthetic and real images of frontal sinuses (a first and partial approach to tackle the third stage).

This chapter is structured as follows. Section VI.2 describes our proposal to tackle the 3D bone scan-2D radiograph superimposition problem. Section VI.3 presents the experiments and their results.

VI.2 Image registration for comparative radiography

There is not a universal standard for any IR method because several considerations of the particular application must be taken into account. Nevertheless, IR methods usually require the components presented in Section III.3 (see Fig. 36 for the 3D-2D IR proposal scheme for CR): (1) the model (PM 3D surface model of the bone) and the scene image (AM radiograph); (2) the projective transformation responsible of generating a 2D image from a 3D object; (3) the expert knowledge/context information of the problem that delimit the target transformation (radiographs acquisition protocols); (4) a similarity metric, which measures the resemblance of a 2D projection with the original 2D image (overlapping); and (5) an optimizer, which looks for the best parameters for the transformation to minimize the error of the similarity metric (to be developed). The composition of each element for our framework is introduced in the following subsections.

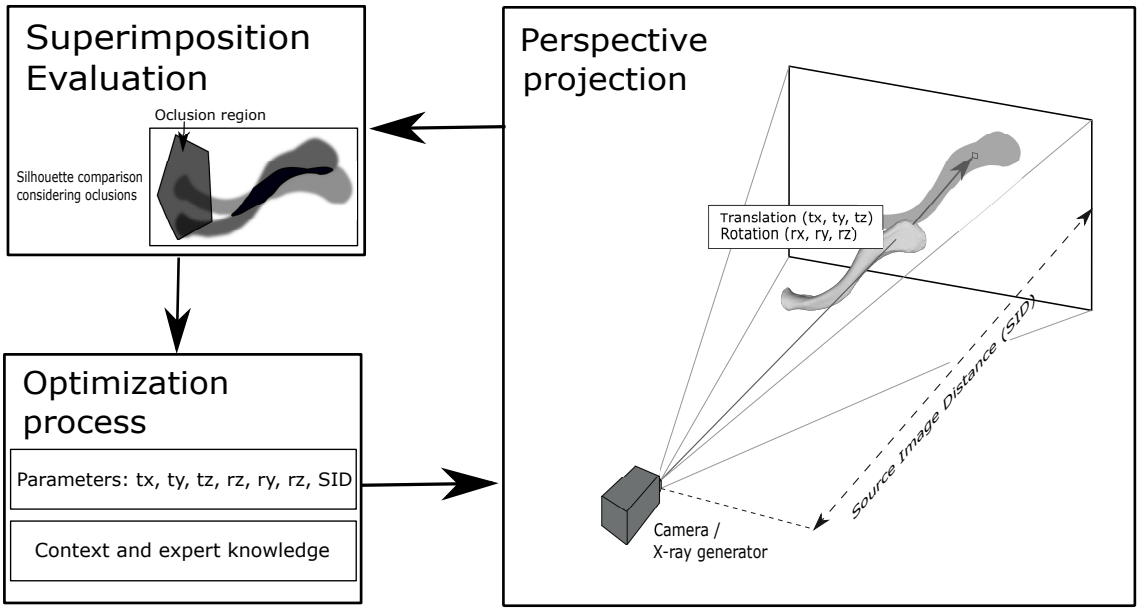

Figure 36: Scheme of the proposal of 3D-2D IR for CR.

VI.2.1 AM and PM images

The raw images to be registered are as follows:

- The 2D image: an AM radiograph acquired in an image receptor, which is usually a flat surface (i.e. flat panels or photographic films) whose size frequently ranges from 240 mm × 350 mm to 430 mm × 430 mm (although there are

other sizes such as those used for tooth radiographs) [BL13] and whose resolution is around 3-10 pixels per mm depending on the technology. However, the resolution could be lower in older radiographs.

- The 3D image: a PM 3D surface model acquired either by scanning a "clean" bone with a laser surface scanner or segmenting the bone in a CT of the deceased (in both cases, the scale of the 3D image is in mm).

In this proposal, the IR process is guided by the silhouette of the bone or cavity requiring the segmentation of the raw images. In the 2D image, the skeletal structure and the occlusion region have to be segmented (see Fig. 26). In the 3D image, there are two different scenarios depending on the raw image: CT or 3D surface scan. In the first scenario, a bone is segmented by thresholding the CT (i.e. a volumetric image) according to the corresponding Hounsfield units. When dealing with cavities like frontal sinuses, a further hindrance has to be addressed. By their nature, cavities can be connected among them and even with the external air. To overcome this problem, the cavity is first isolated by using one or several planes. These planes are horizontal or vertical and must go through a bone landmark (i.e. in frontal sinuses, it is a horizontal plane that goes through a clearly identifiable landmark called the *crista galli*). Finally, the internal air of frontal sinus is selected thresholding with the particular Hounsfield units. In 3D surface scans, no preprocessing is needed although internal cavities such as sinuses cannot be acquired. Lastly, in both scenarios, the center of mass of the 3D surface is moved to the coordinate center.

In general, these segmentation processes always add an unavoidable degree of error in both the 3D and 2D images. Nevertheless, the proposed method may tolerate small segmentation errors, what will be subject of study of this chapter.

VI.2.2 Projective transformation

A projective transformation describes a mapping from 3D to 2D coordinates. Projective transformations are classified according to the type of camera that they model into [HZ03]: perspective projection that models a pinhole camera; and orthographic projection that models an orthographic camera.

The projective transformation behind a radiograph can be modeled (for the most part) with a simple pinhole camera [Mer15]. A simple pinhole camera is a simplification that only considers 7 DoF (6 extrinsic parameters: 3 translations and 3 rotations of the camera; and 1 intrinsic parameter: focal distance) out of the 11 DoF of a the complete model (not considering changes in the rest of intrinsic parameter of a pinhole camera: principal point, assumed in the centre of the image; aspect ratio of the pixels, assumed square; and skewness). Notice that, in a radiograph the perspective distortion is related to the source to image receptor distance (SID) [Mer15] (see Fig. 36) instead of the focal distance. Most works consider a calibrated scenario (only 6 DoF) and the SID is assumed as known which is not the case for the CR problem. [RRM$^+$05, SGW$^+$12, JBVH$^+$06]. Thus, 7 DoF are considered for the CR problem with the perspective transformation.

Meanwhile, an orthographic camera is a particular case of a pinhole camera located at the "infinity" and thus it does not model perspective distortions (see Section III.3.2). Almost all clinical X-ray images have perspective distortions (asides those with a large SID as cephalometric images with a SID of 4 meters). However

even though it does not model a radiograph, it is worth of study to test whereas it is sufficient for identification purposes or not. In addition, it is more mathematically tractable (only 6 DoF: 2 translations, 3 rotations, and 1 scale) and the constraint of the translations and the scale does not require expert knowledge.

The 3D surface model does not have any intensity information (as stated in Section 2) and thus a Digitally Reconstructed Radiograph (DRR) [RRM+05] can not be obtained. Fortunately, a 2D projection can still be provided with a ray-tracing approach producing a binary image with the silhouette of black color and white background. In our case, the implementation of the CGAL library [The17] has been used. This is the most time consuming part of the IR process and thus its optimization is crucial. Therefore, the ray-tracing is only calculated in the surrounding of the silhouette of the segmented bone in the AM radiograph (2.5 times its bounding box), which is the only region of interest for the metric. This approach requires significantly less calculation and time (on average, it takes 0.020 seconds for a projection of 1290×1050 pixels in a standard computer) than a DRR approach (e.g. 0.025 seconds for a DRR of 512×512 pixels using GPUs [SGW+12]), although they are not comparable.

VI.2.3 Parameters and their Constraints using Expert Knowledge

In summary, the parameters of the perspective transformation (7 DoF) are the translation (t_x, t_y, and t_z) in world coordinates, the rotation (r_x, r_y and, r_z), and the SID. Meanwhile, the parameters of the orthographic transformation (6 DoF) are translation (t_x and t_y), that represents the position of the center of the silhouette in the 2D image with respect to the center of the image, rotation (r_x, r_y, and r_z), and scale (s), that represents the percentage of pixels occupied by the bounding box of the silhouette in the image.

The ranges of the parameter are only delimited by the acquisition protocol [BL13] as stated in the introduction. Radiographs are taken with the body in a known position (posteroanterior, anteroposterior, or lateral) and thus the rotation is known with a certain margin of error in both transformations (e.g. $\pm 10°$ or $\pm 20°$ in Euler angles). In the perspective transformation, the acquisition protocols serve us to also delimit the translation and SID and to set the dimensions of the image receptor and its resolution (pixels per mm). The translation on x-axis and y-axis are limited by the width and height of the image receptor respectively. The SID is also limited by the protocol with a margin of error and at the same time the SID also limits the translation on the z-axis since the body is placed as close as possible to the image receptor. Meanwhile, in the orthographic transformation, the translation is not limited by the acquisition protocol but instead by the limits of the image in normalized coordinates of the image for both axis (from -1 to 1). The scale is limited by the percentage of pixels expected to be occupied by the silhouette in the radiograph (from 5% to the 80%). Lastly, the dimension and resolution of the image receptor are not needed.

VI.2.4 Similarity metric

To measure the similarities of two silhouettes/regions, several similarity metrics have been proposed due to its importance in the field of computer vision [VH01] and in its application to medical imaging [MTLP12]. The most utilized to measures the overlap of silhouettes is the DICE metric [Sør48].

However, metrics based on the overlap or on the distance between silhouettes are not robust against occlusion and do not consider partial matching. In intensity-based IR for medical imaging, a Region of Interest (ROI) [PWL+98] is usually considered to address the occlusion problem and increase the accuracy in the visible area. It defines a region with a mask that restricts the evaluation to that region. In our problem, it makes more sense to consider it in the opposite way. The metric is computed in the area of the segmented bone in the AM radiograph but excluding the region where the expert has doubts in the segmentation stage (the occlusion region).

Therefore, a Masked DICE metric that combines the DICE metric with a ROI approach is proposed (see Fig. 37). It computes the overlap of the two silhouettes in the whole image except in the mask region (the occlusion region) (see Eq. VI.1). Notice that, in cases without occlusion the Masked DICE value is equal to the DICE value. Lastly, the metric value is set to 1.5 when the projection of the 3D bone is outside the field of view.

$$\text{Masked DICE} = \frac{2 \cdot |(I_A \setminus M) \cap (I_B \setminus M)|}{|I_A \setminus M| + |I_B \setminus M|} \tag{VI.1}$$

where I_A is the set of pixels of object A (segmented bone) silhouette, I_B is the set of pixels of object B (PM project bone) silhouette, and M is the occlusion region.

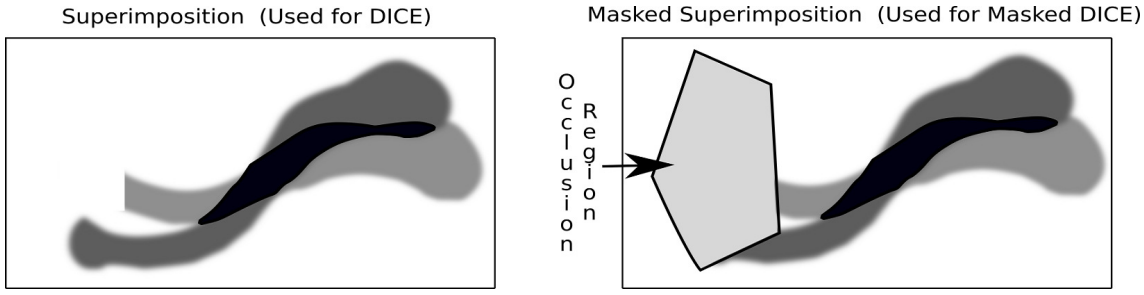

Figure 37: (Left) Superimposition of the silhouette of the AM bone (partially occluded) and the projection of the PM 3D bone; (Right) Masked superimposition by the occlusion region of the silhouette of the AM bone and the projection of the PM 3D bone.

Furthermore, preliminary tests were performed also using the Hausdorff distance but it was discarded because of the poor convergence of the optimizer even in cases without occlusion and being more time consuming.

VI.2.5 Optimizer

A preliminary analysis of the search spaces has been performed using the fitness distance correlation [TVCC05] (see Eq. VI.2 for the distance function) with synthetic data of a clavicle, a patella and a frontal sinus (see Section 4.1.1.).

$$\text{Dist} = \frac{\sum_{i=1}^{n} \left| \frac{t_i - min_i}{max_i - min_i} - \frac{GT_i - min_i}{max_i - min_i} \right|}{n} \tag{VI.2}$$

where n is the number of parameters, t_i is the parameter i-th of a transformation t, GT_i is the parameter i-th of the ground truth transformation GT, min_i is the minimum possible value of the parameter i-th, and max_i is the maximum possible value of the parameter i-th.

This has uncovered the highly multimodality nature (i.e. it has multiple local minima) of the search space with a sample of 100.000 random transformations for each bone/cavity even in the simplest scenario (no occlusion) as shown in Fig. 38. The fitness distance correlation according to the Pearson's correlation coefficient [Pea95] is 0.85 for the orthographic camera model (strongly correlated) and 0.47 for the perspective one (weakly correlated). Thus, the search space of the perspective camera model is more complex than that of the orthographic one. This correlation will decrease as the complexity of the problem increases.

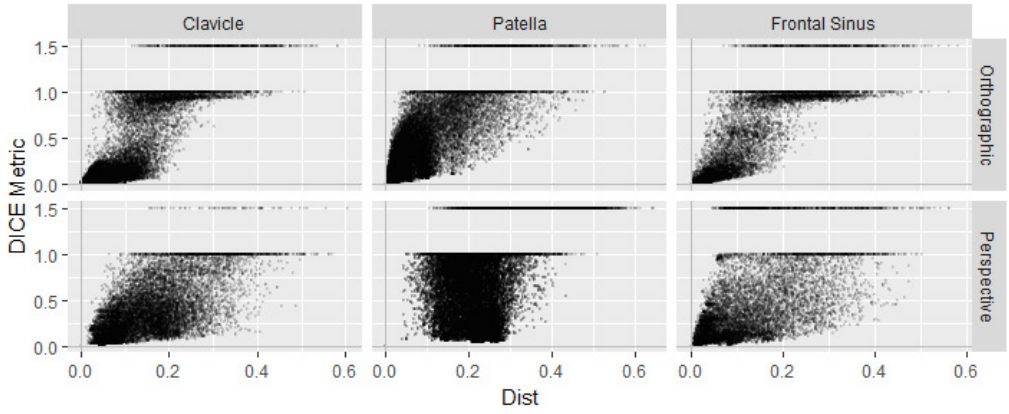

Figure 38: Scatter plots of DICE metric of a transformation versus its distance to the ground truth transformation according to bone/cavity, and camera model.

Thus, to tackle such a complex optimization scenario, two approaches are proposed. The first approach is based on a numerical optimization method called BOBYQA [Pow09], that has already been used for IR. The second approach is based on a metaheuristic called differential evolution (DE) [SP97], that has shown a great performance on the global optimization problems of the Evolutionary Computation Congress Competition CEC-2013 [QL13] and it is easy to use because it has few parameters to set (see Section III.4.1.1). The components of both IR methods for CR are introduced in the following subsections.

VI.2.5.1 EG-BOBYQA

Several numerical optimization methods based on both linear search (Nelder-Mead, BFGS, LBFGS) and trust region (Levenberg-Marquardt, BOBYQA) were tested to solve our IR problem using the DLIB library [Kin12]. Preliminary experiments were carried out and, among the considered methods, the best performing one was BOBYQA. In fact, it is considered the state-of-the-art trust region numerical optimization method [Pow09, MTWL16] and has been already used for IR, including 3D-2D IR scenarios. In particular, it was used for 6 DoF pose estimation from

X-ray [MTWL16]. In that contribution, BOBYQA was compared with Hill Climbing and Nelder-Mead achieving a similar accuracy but with a significant advantage in convergence speed, which is a relevant factor in our problem due to the high computational cost of generating 2D projections.

Even so, the accuracy obtained by BOBYQA in our 3D-2D IR problem was insufficient because its performance greatly depends on the initialization. It resulted to be especially sensitive to the rotation parameter and as stated in the introduction a close initialization is unrealistic in the CR problem as in many other IR approaches. To avoid this problem, we proposed an EG-BOBYQA method (estimation grid-BOBYQA). It includes an initialization grid with the rotation parameters. Furthermore, the translation and scale are estimated for each configuration of the grid with the orthographic transformation (see Eqs. VI.3 and VI.4 respectively), and the estimated solution is improved using BOBYQA.

$$\text{Translation} = c_{AM} + (c_{PM} - b_{PM}) \tag{VI.3}$$

where c_{AM} is the center of mass of the bone in the AM image, c_{PM} is the center of mass of a projection of the PM 3D model with a certain rotation in the center of coordinates and known scale, and b_{PM} is the center of bounding box in the projection. Notice that $c_{PM} - b_{PM}$ vary with each rotation value.

$$\text{Scale} = \frac{P_{AM}}{\frac{P_{PM}}{B_{PM}}} \tag{VI.4}$$

where P_{AM} is the percentage of pixels occupied by the bone within the AM image, P_{PM} is the percentage of pixels occupied by the bone within a projection of the PM 3D model with a certain rotation in the center of coordinates, and B_{PM} is the percentage of pixels occupied by the bounding box of the PM 3D model within the same projection (it is set up to the 20% of the image, although other percentages are valid). Notice that $\frac{P_{PM}}{B_{PM}}$ vary with each rotation value.

However, a grid with a step of one degree over a certain rotation range in the three axes (where the rotation parameters are delimited) is computationally unapproachable when the rotation range is superior to 20°. Therefore, it is tackled with several grids with a decreasing step and rotation range around the best solution of the previous grid. Several configurations were tested and the best tradeoff (in terms of time and accuracy) configuration was a step of a quarter of the rotation range for each grid and a decrease of the rotation range of the following grid to [best rotation - $\frac{step}{2}$, best rotation + $\frac{step}{2}$] for the three axes until a step smaller than 1 degree is reached.

The translation and scale are estimated with enough accuracy when a near rotation (around 1 degree) is tested and there are no significant occlusions. However, the method is not accurate in the presence of occlusions. Therefore, BOBYQA is used to refine the estimation. However, it significantly increases the execution time and the required number of evaluations and thus this refinement is limited to 50 evaluations (again a tradeoff value between time and accuracy).

Finally, the best solution of the grid is optimized again with BOBYQA without limit of evaluations. See Fig. 39 for a graphical explanation of the proposed EG-BOBYQA algorithm for 3D-2D CR.

The main drawback of this approach is that it is only applicable with the orthographic transformation because in the perspective transformation the scale and the

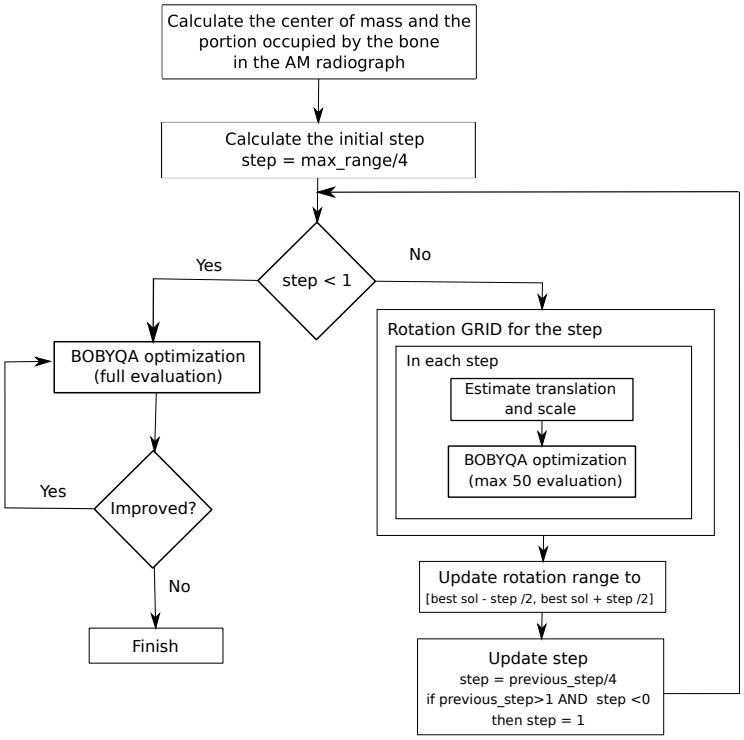

Figure 39: Schema of the EG-BOBYQA optimization proposal.

translation cannot be directly estimated because they depend on several parameters (i.e. the scale depends on the translation on the z-axis and the SID).

VI.2.5.2 Differential Evolution

As reviewed in Section III.4, metaheuristics [CDS06], and in particular DE [SDGTC12], have shown a great performance on 3D-3D IR problems. DE is a variant of an evolution strategy [Bey13] proposed by Storn and Price [SP97]. It is an stochastic search technique for solving optimization problems over continuous spaces. It has been successfully applied to optimization problems including non-linear, non-differentiable, non-convex and multi-modal functions [Cha08]. It has also been widely applied in many real-world problems because of its robustness, fast convergence, and its reduced number of parameters to set [DS11]. In particular, it was considered for IR in [DFDCMT08] with a very good performance. See Section III.4.1.1 for further details.

VI.3 Experiments

The experimental study is divided into three parts. The first experiment is devoted to the study of performance, precision and robustness of the different methods proposed with simulated CR problems of different bones/cavities (frontal sinuses, clavicles and patellae) where the AM and PM data belong the same person (only positive cases). Meanwhile, the goal of the second experiment is to study how optimization performance affect the identification capability of the best proposal in a n-to-n scenario with simulated CR problems of frontal sinuses. Last, the third experiment studies how segmentation errors, inter-expert variability and optimiza-

tion performance affect the identification capability of the best proposal in a n-to-n scenario with real CR problems of frontal sinuses. This experiment is performed with frontal sinuses only because we do not have real AM radiographs of clavicles and patellae available.

For all the experiments, a stop criteria is set when the optimizer reaches an error lower than 0.01%, which means that the 99.99% of the pixels outside the occlusion region are superimposed correctly with respect to the GT.

Furthermore, the parameters of the optimizers were also set for all the experiments in preliminary studies. In particular, the best configuration for BOBYQA's initial trust region was set to 0.05 (see Section 3.5.1), while the best configuration achieved in the case of the DE considered the following parameter values: 100 individuals, 50000 evaluations, a crossover probability P_c of 0.5, and F set to 0.5.

All the experiments have been performed on a computing server with 12 nodes that have an Intel Core i7 4930k 3.4 GHz, running Ubuntu 16.04.

VI.3.1 Experiment 1: Validation of the image registration approach

VI.3.1.1 Data set generation

Simulated CR problems were used to validate the capability of our method to accurately perform 3D-2D IR. A simulated CR problem is composed of a 3D surface model and a 2D perspective projection of the 3D model with a random transformation (within a given parameter range) to be superimposed. It allows the quality of a superimposition to be objectively measured, even if there are occlusions present, because its GT projection without occlusions is known.

To do so, 10 clavicles and 10 patellae from the bone collection of the Physical Anthropology lab at the University of Granada were scanned with a laser range scanner (Artec $Spider^{TM}$ 3D scanner). Furthermore, 10 3D surface models of frontal sinuses were obtained by manually segmenting 10 CTs (provided by the Hospital de Castilla la Mancha, Spain) using 3D slicer 4.5.0-1 (see Section VI.2.1). The frontal sinuses and clavicles 3D surface models were placed in frontal positions and the patellae 3D surface models in lateral positions because they are respectively the most common acquisition positions in a radiography [BL13]. For each of the 30 3D surface models, 5 perspective projections were generated within the ranges showed in Table 11 with a resolution of 3 pixels per mm. These ranges were set following the international acquisition protocols [BL13] and the constraints stated at Section 3.3. The image receptor dimension and SID for the frontal sinuses and patellae are 240 mm $\times$ 300 mm and 1000 mm respectively, and for clavicles 430 mm $\times$ 350 mm and 1800 mm respectively, resulting in images of 750 $\times$ 900 pixels for frontal sinuses and patellae, and 1290 $\times$ 1050 for clavicles. From each of the 150 simulated CR problems, two additional simulated problems were generated with an increasing degree of occlusion of the target bone of 15% and 30% in order to model bone silhouette occlusion in real radiographs. These occlusions are located at the bottom for the frontal sinuses and the patella and in the mid-region near the sternum for clavicles because that is where the occlusion usually takes place on real radiographs. A total of 450 positive simulated CR problems with their corresponding GT were generated.

Table 11: Parameter range of each bone/cavity for the perspective transformation constrained by international acquisition protocols [BL13] and expert knowledge (see Section 3.3).

Parameter	Bone/Cavity		
	Frontal Sinuses	Patellae	Clavicles
t_x (mm)	[-125, 125]	[-125, 125]	[-210, 210]
t_y (mm)	[-150, 150]		[-175, 175]
t_z (mm)	[900 - 200, 900 + 200]		[900 - 200, 1700 + 200]
r_x, r_y, and r_z (degrees)	[-10°, 10°], [-20°, 20°]		
SID (mm)	[1000 - 100, 1000 + 100]		[1800 - 100, 1800 + 100]

VI.3.1.2 Experimental set-up

This experimentation involves the application of two different optimizers (EG-BOBYQA and DE), two kinds of projective transformations (perspective and orthographic), and two rotation ranges ($\pm$10°and $\pm$20°) for each of the 450 CR cases (generated as described in Section 4.1.1). Notice that, the only parameter ranges altered are those related to the rotation of the bone/cavity and the rest remain unchanged (see Table 11 and Section 3.3 for the ranges of the perspective and orthographic transformations, respectively). However, the EG-BOBYQA optimizer cannot be used with the perspective transformation because the translation and the scale cannot be directly estimated. In summary, a total of 2700 experiments were carried corresponding to:

1. Orthographic transformation: 2 optimizers (EG-BOBYQA and DE), 2 rotation values ($\pm$10°and $\pm$20°), and 450 simulated CR problems (notice that, the 450 cases were generated with a perspective transformation), i.e. an overall of 1800 experiments.

2. Perspective transformation: 1 optimizer (DE), 2 rotation values ($\pm$10°and $\pm$20°), and 450 simulated CR problems, i.e. an overall of 900 experiments.

Since the DE approach is based on a stochastic process, 16 independent runs were performed for each problem instance to compare the robustness of the methods and to avoid any possible bias when the DE optimizer is utilized. The initialization of each run is random in the whole parameter range for all the degrees of freedom of its corresponding projective transformation. A closer initialization would be unrealistic in a real identification scenarios where the AM radiograph would have been taken in unknown conditions.

VI.3.1.3 Ground truth metrics

Two GT metrics are considered to objectively measure the quality of the results. The first one is GT DICE. This metric measures the percentage of not superimposed pixels of the 2D projection obtained by the optimizer and the GT projection (which is equivalent to the AM 2D projection used by the optimizer, but without occlusions). Notice that, in cases without occlusion Masked DICE is equal to the original GT DICE (see eq. VI.1).

To avoid any possible bias caused by the high correlation between the two DICE metrics, a second independent metric is also utilized, the mean reprojection distance error (mRPD) [vdKPT$^+$05], that allows to perform standardized evaluation in 3D-2D IR. It measures the average distance from each 3D point of the 3D surface model

to a reprojected line (which is a line composed of all the 3D points whose projection under a certain transformation results in the same 2D point). A reprojected line is calculated using the inverse of the transformation obtained by the optimizer and the 2D projection of the 3D point projected with the GT transformation (see eq. VI.5).

$$\text{mRPD} = \frac{1}{m} \cdot \sum_{i=1}^{m} |||p_i, P_{reg}^{-1}(P_{GT}(p_i))||| \tag{VI.5}$$

where p_i is the i-th 3D point of the 3D surface model, m is the number of points of the 3D surface model, P_{GT} is the GT projective transformation, $P_{GT}(a)$ is the 2D point resulting from the projection of the 3D point a with the GT transformation, P_{reg} is the projection transformation obtained by the optimizer, and $P_{reg}^{-1}(b)$ is basically a straight line composed of all the 3D points that multiplied by P_{reg} result in the 2D point b.

VI.3.1.4 Results

The results obtained are shown in Table 12 according to the Masked DICE metric, the GT DICE metric, and the mRPD metric. The GT metric is strongly correlated with the Masked DICE metric (the metric that guided the optimizer) with a correlation of 0.845 according to the Pearson's correlation coefficient [Pea95]. Meanwhile, the mRPD metric is also correlated with both metrics for the perspective transformation (0.780 for Masked DICE and 0.801 for GT DICE).

Table 12: Summary of the Masked DICE metric results, the GT metric results, and the mRPD metric results according to bone/cavity type, camera model, and optimizer.

Bone	Optimizer	Camera Model	Masked DICE		GT DICE		mRPD (mm)	
			mean	sd	mean	sd	mean	sd
Clavicle	DE	Ortho.	0.015	0.011	0.037	0.018	12.128	14.442
		Persp.	**0.001**	0.003	**0.002**	0.005	**0.055**	0.088
	EG-BOBYQA	Ortho.	0.044	0.025	0.083	0.065	12.714	14.373
Patella	DE	Ortho.	0.014	0.016	0.029	0.034	11.981	16.270
		Persp.	**0.005**	0.008	**0.015**	0.025	**0.761**	1.544
	EG-BOBYQA	Ortho.	0.035	0.020	0.103	0.081	12.148	16.892
Frontal Sinus	DE	Ortho.	0.014	0.055	0.020	0.058	8.471	3.571
		Persp.	**0.001**	0.003	**0.002**	0.006	**0.028**	0.067
	EG-BOBYQA	Ortho.	0.029	0.034	0.051	0.060	8.285	3.399

As expected, it can be clearly seen how the results obtained by both optimization approaches with the perspective transformation perform better (with a mean error lower than 0.1 mm for clavicles and frontal sinuses, and lower than 1 mm for patellae according to the mRPD metric) than with the orthographic transformation (with a mean error always higher than 8 mm according to the mRPD metric). That fact is confirmed by the Wilcoxon's test [Geh65] and the sign test obtaining p-values of $2.2 \cdot 10^{-16}$ and $8.6 \cdot 10^{-203}$ respectively. This difference can be explained since only the perspective transformation can reproduce the perspective distortions present in the AM simulated radiographs. For the orthographic projection, the DE optimizer performs better than EG-BOBYQA in average and standard deviation. Furthermore, DE's mean is significantly lower, which is confirmed by the Wilcoxon's test obtaining a p-value of $2.2 \cdot 10^{-16}$. DE also outperforms EG-BOBYQA in most scenarios, which is confirmed by the sign test [DM46] obtaining a p-value of $9.2 \cdot$

10^{-110}. The robustness of both optimizers can be improved because the results have a dispersion that leads some runs to have quite large errors (up to 40 % for the GT metric and 60 mm according to mRPD metric).

The minimum error obtained by the DE optimizer (i.e. the result of the best run) for the perspective transformation is lower than a 0.5% of badly superimposed pixels according to GT DICE metric or 0.1 mm according to mRPD metric for clavicles, patellae and frontal sinuses (see Fig. 40). Meanwhile, the min error for the orthographic projection is higher (around 1% for GT DICE and 2.5 mm for mRPD with patellae and frontal sinuses, and 3% for GT DICE and 1 mm for mRPD with patellae and frontal sinuses). Furthermore, occlusions have a visible effect on the accuracy according to the GT DICE Metric but not according to the mRPD. Meanwhile, the influence of the rotation affect mainly to robustness but not significantly according to both metrics.

A first-sight conclusion is the strong influence of bone/cavity on the performance (see Fig. 40). Better results are always obtained for frontal sinuses in terms of accuracy and robustness than for clavicles and patellae, probably due to the singularity of the visible region of the frontal sinuses. For instance, frontal sinuses have been used in several works for identification [QFS+96] and are different even in homozygous twins whereas the clavicles and the patellae have been mainly used for short listing [NSGF16, SWCT11]. Last, the worst results are obtained for patellae in terms of robustness for the lower singularity of the visible region.

The main weakness of both optimizers is the computational time required to achieve the results. As shown in Table 13, the time to obtain a superimposition is large (30 minutes in average), making it hard to run the algorithm again when a bad superimposition is achieved. EG-BOBYQA requires more time than DE for the orthographic model (73 and 25 minutes in average, respectively), which makes the DE optimizer better also in terms of computation time. This is due to the high number of evaluations performed by the optimizers in order to overcome the problem of not relying on an initialization (notice that, the time required for clavicles is higher due to the larger size of their radiographs).

Table 13: Summary of the required computation time (in seconds) according to bone/cavity type, camera model, and optimizer.

Bone	Optimizer	Camera.Model	Time (seconds)	
			mean	sd
Clavicle	DE	Orthographic	3775	954
		Perspective	**2370**	981
	EG-BOBYQA	Orthographic	6947	2381
Patella	DE	Orthographic	1121	400
		Perspective	**841**	493
	EG-BOBYQA	Orthographic	2052	714
Frontal Sinus	DE	Orthographic	2128	760
		Perspective	**966**	611
	EG-BOBYQA	Orthographic	3421	1342

With the perspective transformation, DE usually reaches the stop criteria before the 300-th generation which explains its lower run time. However, the orthographic transformation does not reach the stop criteria, despite it usually converges at the 200-th generation.

In summary, the DE optimizer is able to obtain good superimpositions for both camera models, but DE has only shown a robust behavior for frontal sinuses due to the singularity of their silhouettes.

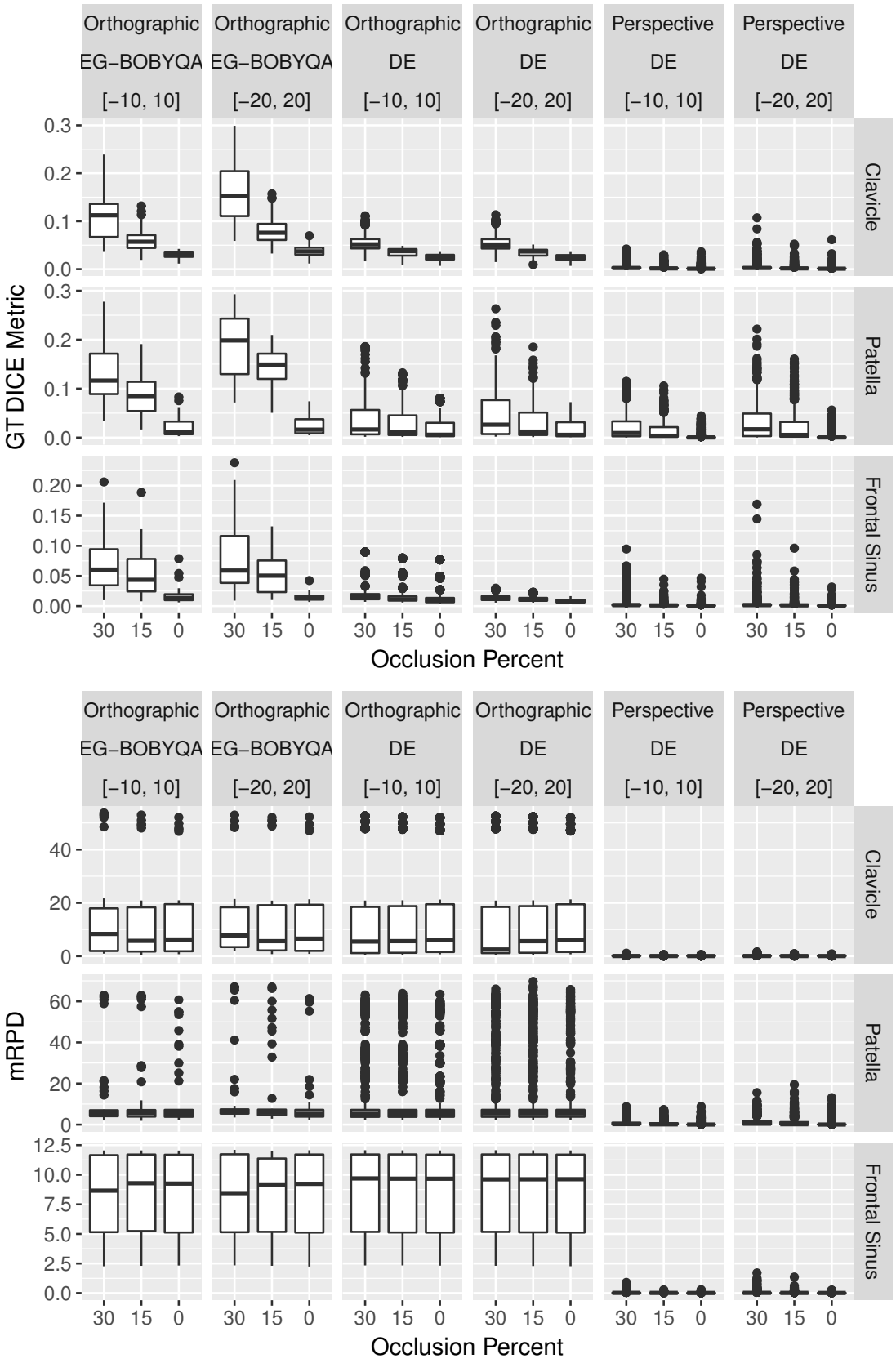

Figure 40: Boxplots of the minimum errors according to bone/cavity, camera mode, and optimizer for the GT DICE metric (Top) and mRPD metric (Bottom).

VI.3.2 Experiment 2: Validation of the automatic CR methodology

The aim of this experiment is to study how the precision of DE affects the identification capability with simulated CR problems of frontal sinuses.

VI.3.2.1 Data set generation

For the experiment, 2 perspective projections were generated for each of the 10 frontal sinuses within the ranges of Table 11. For each of theses 20 frontal sinuses projections, two additional simulated problems were generated with an increasing degree of occlusion of the target bone of 15% and 30%, resulting in 60 simulated AM projections and 10 PM 3D surface models. These are combined into 600 simulated CR problems (60 positive cases and 540 negative cases).

VI.3.2.2 Experimental set-up

This experimentation involves the application of the best optimizer according to Section 4.1 (DE), the two rotation ranges ($\pm 10^\circ$ and $\pm 20^\circ$, and again the rest of parameter ranges remain unchanged), and the two projective transformations (perspective, and orthographic), resulting in 2400 experiments.

Since the DE approach is based on a stochastic process, several runs must be performed in order to validate the robustness of the method. However, this study has already been performed for frontal sinuses with the DE optimizer (see Section 4.1.) showing the robustness of the optimizer with a positive case. For this reason (and due to the great amount of computational time required to perform again 16 independent runs), only 2 independent runs are performed.

VI.3.2.3 Results

Promising results were obtained. Positive and negative cases have shown a great difference in terms of fitness according to the Masked DICE Metric (see Fig. 41). The GT DICE and mRPD metrics are not suitable for this study because they cannot be known in real CR cases. Furthermore, the positive cases always rank the first in the ranking resulting of ordering each AM projection against all the PM 3D models according to the Masked DICE metric for the perspective camera model. Meanwhile for the orthographic camera model, the positive cases rank the first in the 99% of the cases but this percentage goes up to the 100% if the best run of each experiment is considered.

VI.3.3 Experiment 3: Validation of the CR methodology on real cases

The third experiment is devoted to study how the accuracy of our best proposals affects the identification on real CR problems of frontal sinuses. Furthermore, this experiment studies the validity of the orthographic and perspective camera models for identification, while analyzing the effect of inter-observer segmentation errors.

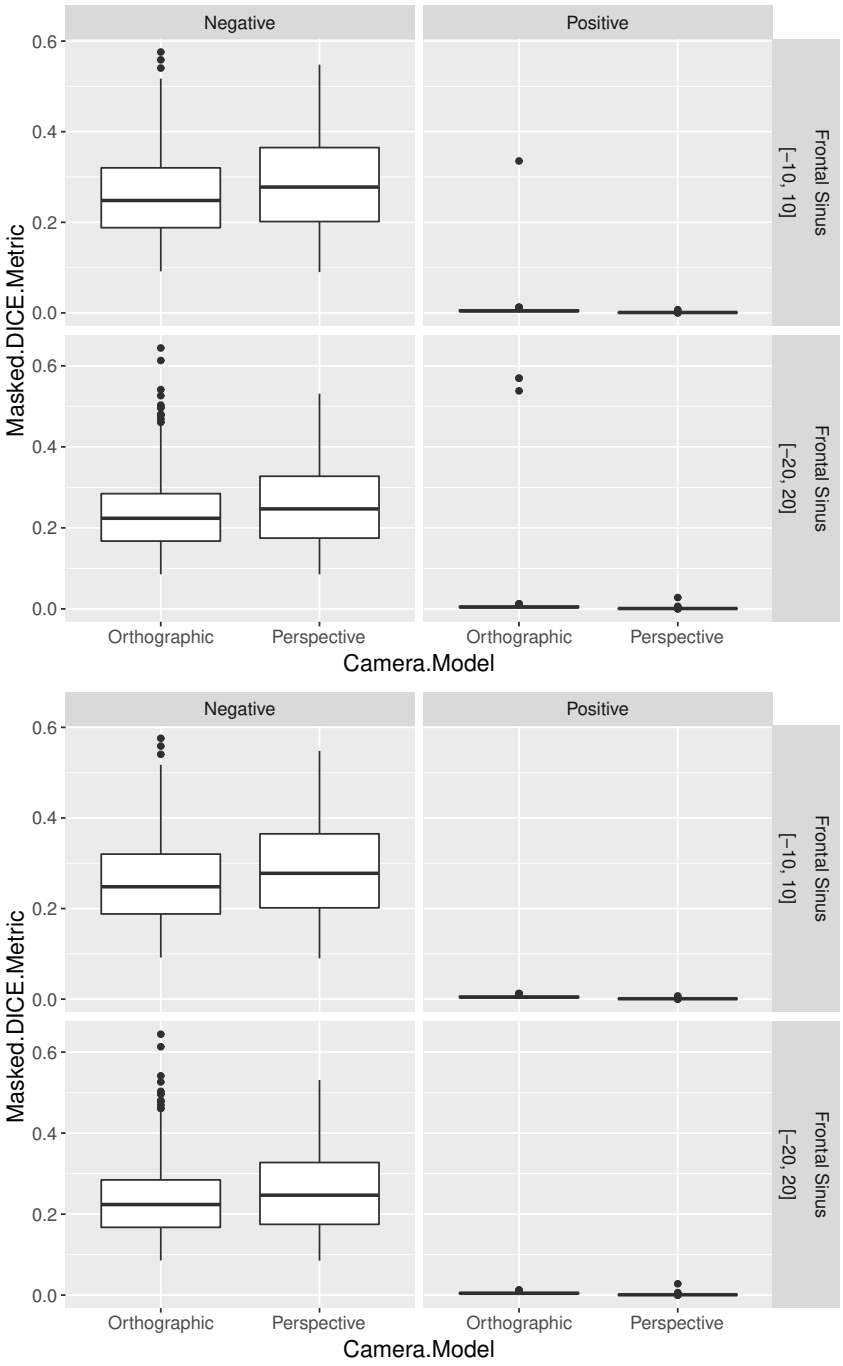

Figure 41: Boxplots of the mean (Top) and minimum (Bottom) error for the Masked DICE metric of positive or negative cases according to the rotation range and the camera model.

VI.3.3.1 Data set

The data set for this last experiment is composed of 10 pairs of a PM 3D surface model and a AM radiograph of a frontal sinus. Each of the AM radiographs was segmented by 3 different persons (as seen in Section 3.1.), and also a consensus segmentation was calculated (which contains the pixels segmented in at least 2 out of 3 manual segmentations). This results in 40 real segmented AM radiographs and 10 PM 3D surface models which are combined into 400 real CR problems (40 positive cases and 360 negative cases).

VI.3.3.2 Experimental set-up

This experimentation involves the application of the best optimizer (DE), the two rotation ranges ($\pm 10°$, $\pm 20°$), and the two projective transformations (perspective, and orthographic), resulting in 1600 experiments.

Since the DE approach is based on a stochastic process, 16 independent runs with random initializations were performed for each experiment, due to the DE optimizer has not been tested with real radiographs manually segmented yet.

VI.3.3.3 Results

The results are reported using Cumulative Match Characteristic (CMC) curves [LJ05] to study the identification capabilities of the proposal as done in [CAICW18]. A CMC curve measures the probability that the correct match for a identification case is present in a candidate list of the r best matches, where r denotes the position in the rank. For example, rank 5 identification accuracy denotes the probability that the correct match is one of the subjects in a list of the top 5 matches. To focus on the potential identification capability of the method and not in the robustness only the best run of each experiment is considered. In this case, the rank 1 includes the 10% of the sample (i.e. 1 out of the 10 comparisons performed for each AM case with a given configuration). The rank 2 includes the 20% of the sample, and so on until reaching the rank 10 that includes the 100% of the sample.

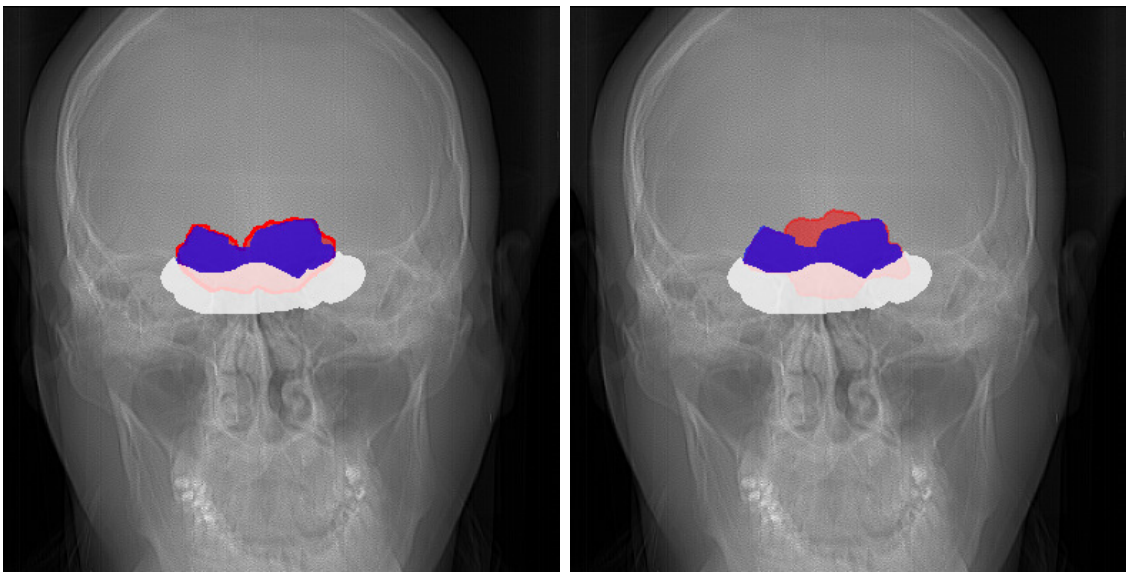

Figure 42: Example of a positive case with a frontal sinus ranked in the first position (left) and a negative case ranked in the second position (right). The AM segmentation is represented in blue, the projection of the PM 3D model in red, and the occlusion region in white.

The results obtained show that the 58% of the 160 experiments with positive cases rank the first in the ranking resulting of ordering each AM projection against all the PM 3D models according to the Masked DICE metric. The percentage goes up to 85% if a rank 3 identification is considered. Furthermore, the CMC curve shows that the perspective camera model has a better performance than the orthographic one (see Fig. 43a). For instance the perspective camera model obtains a probability of 97% with a rank 4 identification while the orthographic model only

obtains a 87%. Furthermore, the rotation range also has a significant effect in the performance (see Fig. 43b). The best rotation range was the ±10° as in the first experiment. Lastly, the performance is greatly affected by the segmentation errors (see Fig. 43c).

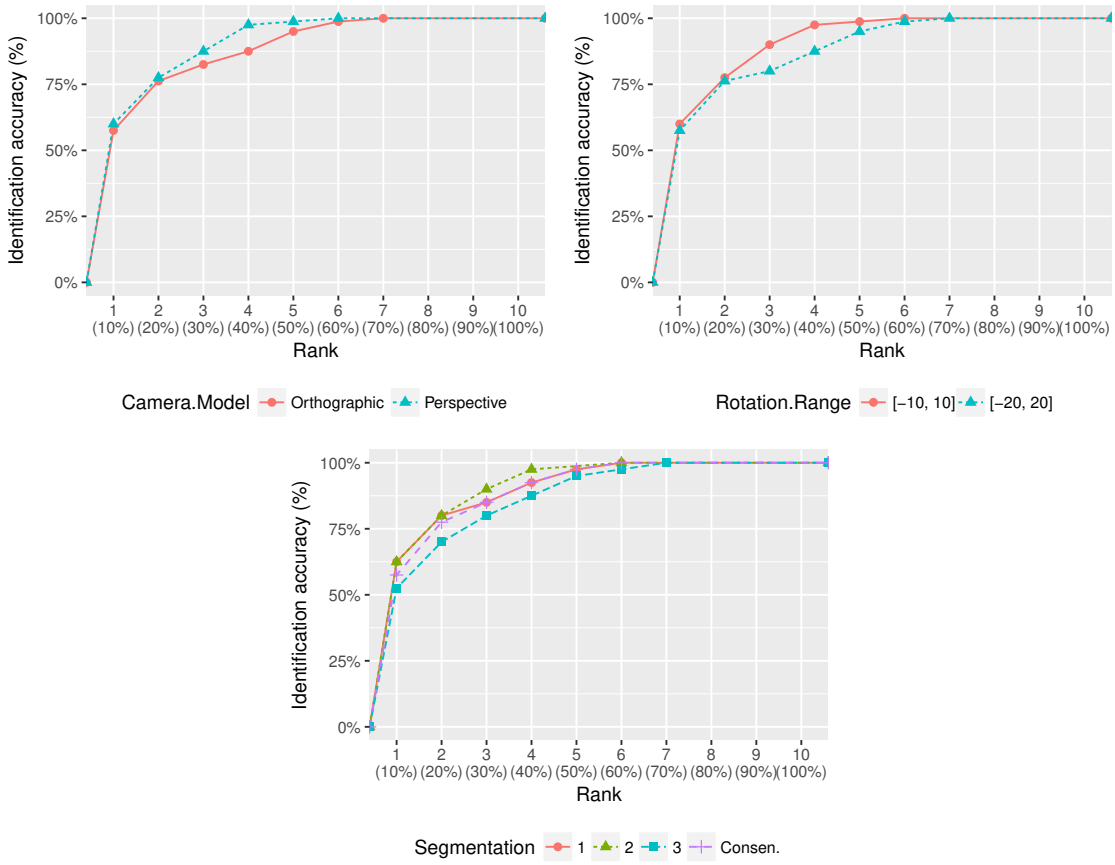

Figure 43: (Top left) CMC curve according to the camera model. (Top right) CMC curve according to the range of rotation. (Bottom) CMC curve according to the segmentation.The number in parenthesis represents the percentage of the sample included in each rank.

Deepening into the differences according to the AM case, the results show that five of the ten frontal sinuses always rank the first regardless of the camera model, the rotation range, or the segmentation (see Fig. 44). In addition, another two frontal sinuses also rank the first but only with one of the segmentations (segmentation 2). The AM radiographs of these two frontal sinuses present several differences depending on the segmentation on the available radiographs, due to a lower visibility of the upper area (which is its most characteristic and identifying part). Lastly, the remaining 3 frontal sinuses never rank the first (ranking between the second and the sixth position depending on the scenario) and their segmentation were the hardest presenting an occlusion region even on the upper part (again due to the low visibility).

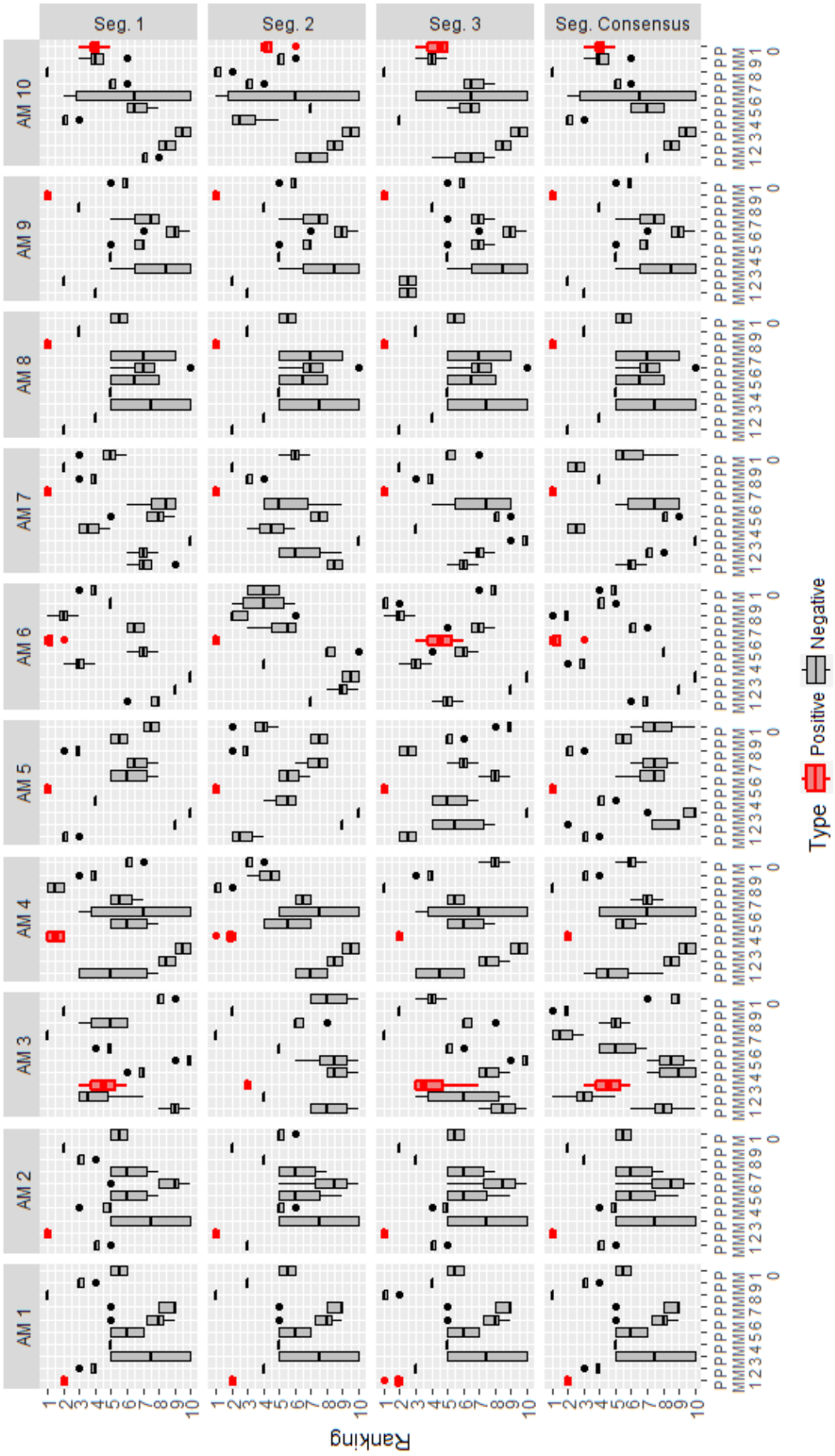

Figure 44: Boxplots of the rankings according to AM case, PM case and segmentation.

Chapter VII

Performance analysis of the real-coded evolutionary algorithm for comparative radiography

'There is something at work in my soul, which I do not understand." — Mary Shelley

VII.1 Introduction

Most limitations of current CR approaches are overcome by the evolutionary 3D-2D IR approach presented in the previous chapter. However, the proposed method still shows the following drawbacks: (1) the projective transformations cannot reproduce the perspective distortion of radiographs where the X-ray generator was not perpendicular to the image receptor (e.g. in the Water's projection of radiographs of frontal sinuses [TBV02]); (2) the robustness of the DE algorithm is low in some cases, especially with clavicles and patellae, where it led to bad superimpositions in some runs, due to the stochastic nature of DE and the highly multimodal search space tackled (see Section VI.2.5 for further details of the landscape analysis); and (3) the large amount of time required to obtain a superimposition with DE (1,800 seconds on average). This elevated time is motivated by the high computational cost required by each evaluation (on average, it takes 0.25 seconds to obtain a projection of 1290×1050 pixels in a standard computer), uncovering the computationally expensive optimization nature of the CR problem as well as the high number of evaluations needed by the optimizer to converge.

Apart from the high computation requirements of CR, the numerical optimization problem underlying the superimposition process is complex, even when the perspective distortions related to the first drawback are not modeled, resulting in a highly multimodal search space. The complexity of the problem comes from several sources, such as the segmentation errors and the inherent uncertainty in both the AM and PM image, the lack of assumptions regarding the initialization, the strong interrelation among the parameters (e.g. the apparent size of the projection is affected by all the parameters in a perspective projection), the dependency on the singularity of the bone, etc. Thus, the choice of the optimizer plays a crucial role in the superimposition process. While numerical methods have proved to be insufficient,

considering an evolutionary IR method based on DE has shown a great performance but still suffers from the second and third drawbacks described. Therefore, a comparative study of several high performance RCEAs [BFME97, YG10, MLH18, ZYQ19], with particular focus on those tested in complex real-world problems as well as in competitions from the IEEE Congress on Evolutionary Computation (CEC), is necessary in order to determine the influence of the RCEA considered in the 3D-2D IR framework.

The goal of this chapter is thus three-fold. Firstly, to propose and validate a new projective transformation that can reproduce the perspective distortion of any kind of radiograph. This new projective transformation can model regular radiographs as well as radiographs where the X-ray generator is not perpendicular to the image receptor, as in the Water's projection of radiographs of frontal sinuses [TBV02]. Furthermore, this new projective transformation can also model small alignment errors in regular radiographs. Secondly, to perform a comparative study of several state-of-the-art RCEAs looking for better accuracy, robustness, and convergence speed in the automatic CR process. These two goals are studied with synthetic images of three skeletal structures (clavicles [SWCT11], patellae [NSGF16], and frontal sinuses [QFS+96]), which have been commonly utilized in the CR literature as well as along this dissertation. Thirdly, to study the performance of the best RCEA in real images of frontal sinuses as well as its robustness to intra-expert and inter-expert segmentation variability. With this third goal, we will be able to evaluate the actual capability of our methodology to implement an automatic CR-based identification method to assist the forensic anthropologist.

This chapter is structured as follows. Section VII.2 describes the additions to the IR methodology, the new projective transformation, and the state-of-the-art RCEAs utilized. Section VII.3 presents experiments and results.

VII.2 Methodology

The methodology is similar to that proposed in the previous chapter following a IR approach with five components (the data, the projective transformation, the expert knowledge that delimit the transformation, the similarity metric, and the optimizer). These five components are further detailed in Chapter VI. The main contribution of this chapter is in the proposal of a new projective transformation and the analysis of the optimizers (the second and fifth components, respectively) that will be detailed in the following subsections (see Fig. 45).

VII.2.1 Projective transformation

The projective transformation [HZ03] behind a radiograph image is, in most of the cases, a simple perspective transformation obtained using a pinhole camera model [Mer15]. As already described, simple perspective transformation considers 6 extrinsic parameters (3 translation and 3 rotations) and 1 intrinsic parameter (focal distance; assuming that the rest of intrinsic parameters of a complete perspective transformation are known: the principal point is located in the center of the image, pixels' aspect ratio is square, and no skewness). Particularly, in a radiograph, the focal distance is represented by the SID [Mer15] (see Fig. 45).

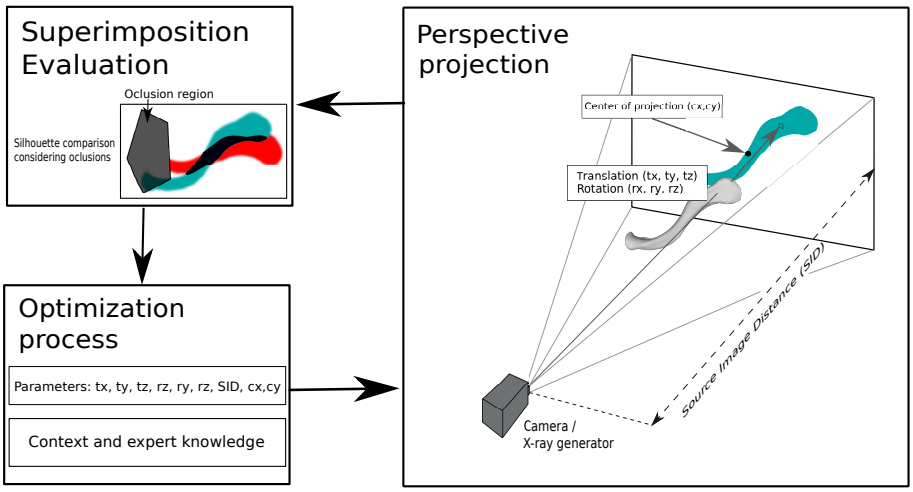

Figure 45: Scheme of the proposal of 3D-2D IR for CR. Three main interconnected blocks are represented: (Right) the projective transformation to obtain a projection of the 3D model with 9 parameters: translation (t_x, t_y, and t_z), rotation (r_x, r_y, and r_z), and perspective distortions (SID, c_x, and c_y); (Top left) The similarity metrics that compares the PM projection (colored in blue) and the AM segmentation (colored in red) considering an occlusion region (colored in gray); (Bottom left) the optimization process to estimate the 9 parameters of the registration transformation that are only weakly limited by the context and expert knowledge from the X-ray acquisition protocol.

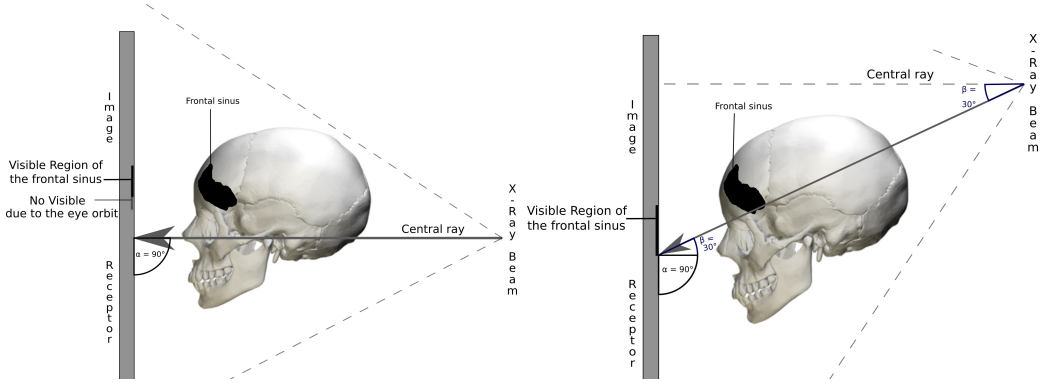

Figure 46: (Left) Diagram of a frontal sinus radiograph with a posteroanterior view, where the ray between X-ray generator and the center of the image receptor is perpendicular. (Right) Diagram of a frontal sinus radiograph with a Water's view, where the ray between X-ray generator and the center of the image receptor is not perpendicular.

However, radiographs acquired with procedures where the ray that joins the X-ray generator and the center of the image receptor is not perpendicular cannot be modeled with a simple perspective transformation. That is the case of frontal sinuses radiographs taken in one of the acquisition protocols of the Water's view (see Fig. 46 for a graphical example). In these radiographs, the acquisition protocols [BL13] establish that the X-ray bean is angled at β to the center of the receptor. It causes that the principal point of the image is not located at the center of the images (as can be seen in Fig. 46) and can be located even outside the image limits. Thus, to model these radiographs, a more complex perspective transformation that also models changes in the principal points is needed (resulting in 9 parameters to be optimized). The movement of the principal point in an axis can be calculated

according to the following equation:

$$c_i = SID \cdot \frac{sin(90 - \beta_i)}{sin(\beta_i)} \tag{VII.1}$$

where c_i is the principal point displacement in the axis i and β_i is the angle of the ray that joins the center of the image receptor and X-ray generator in the axis i.

Furthermore, even radiographs taken in conventional views as the posterioanterior can be affected by this distortion (although with a minor effect), due to the small alignment errors between the image receptor and X-ray generator and the modeling of changes in the principal point can also be beneficial for them.

To sum up, two projective transformations are considered in this contribution, aiming to improve the performance of the automatic CR-method: the simple perspective projection with 7 parameters (t_x, t_y, t_z, r_x, r_y, r_z, and SID) from the previous chapter and a new more complex perspective projection with 9 parameters (t_x, t_y, t_z, r_x, r_y, r_z, SID, β_x, and β_y). The two transformations will be referred from now on as P7 and P9, respectively. Their parameters' ranges are stated in Section VII.3.1.

VII.2.2 Real-coded evolutionary algorithms for the image registration optimizer

As stated in Section VI.2.5, RCEAs can tackle complex and multimodal optimization problems. In our case, this complexity has increased with the new projective transformation P9. This can be confirmed by studying the fitness's landscape of the CR problem by using the fitness-distance correlation [TVCC05] (see Eq. VI.2 for the distance function). As in previous chapters, the complexity of the CR problem is uncovered by studying its simplest scenario, i.e. synthetic data without occlusions or segmentation errors. To analyze the simplest optimization scenario, a sample of 200,000 random transformations near to the GT transformation have been generated and evaluated for each skeletal structure (clavicles, patellae and frontal sinuses) and perspective transformation (P7 and P9), as shown in Fig. 47. This depicts many poor superimpositions with a small distance to the GT transformation, as well as good superimpositions with a big distance to the GT transformation. It hints the multimodality of the search space. Furthermore, the fitness distance correlation according to the Pearson's correlation coefficient [Pea95] is 0.47 for P7 and 0.42 for P9, both weakly correlated, confirming that P9 is more complex and multimodal than P7.

To tackle the real-coded optimization problem of P7 and P9, six RCEAs are studied and fine-tuned using the Masked DICE metric as fitness function. The RCEAs to be studied are the following (see Section III.4.1 for further details): (1) DE, a classic RCEA and the one originally used in our methodology; (2) L-SHADE, one of the best self-adaptive variants of DE; (3) CMA-ES, a classic RCEA that has outperformed DE in many problems; (4) BIPOP-CMAES, one of the best modern variations of CMA-ES; (5) CRO-SL, a powerful RCEA that is the state-of-the-art method in 3D-3D IR problems but is complex to fine-tune; and (6) MVMO-SH, a novel RCEA that has obtained groundbreaking results in costly optimization problems [ERWS14].

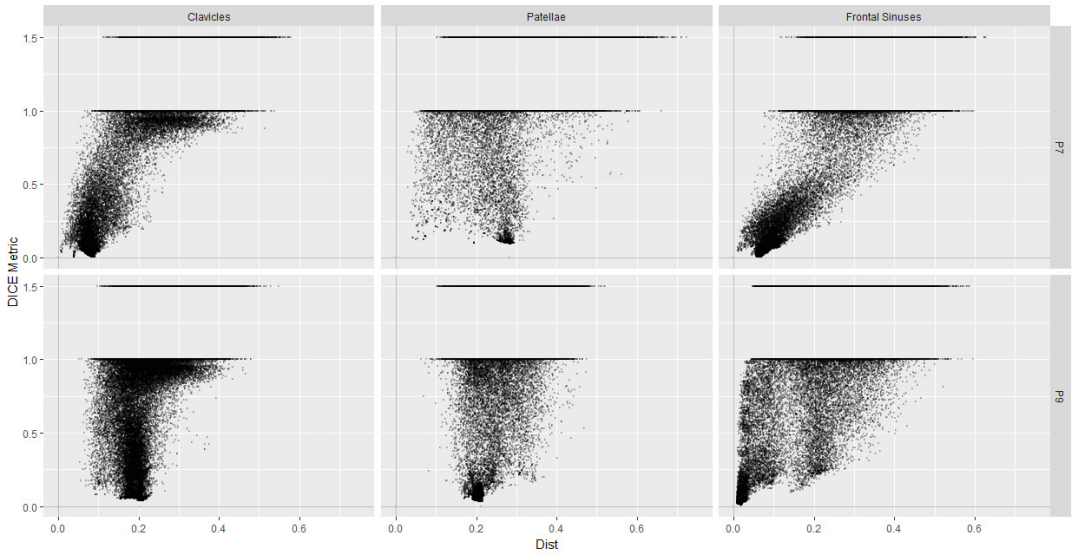

Figure 47: Scatter plots of DICE metric of a transformation versus its distance to the GT transformation according to bone/cavity, and perspective transformation.

VII.3 Experiments

The experimental study is divided into three parts. The first experiment is devoted to fine-tune the different RCEAs to find their best configuration in terms of accuracy and robustness. For this experiment, only simulated CR problems (positive cases, i.e. the AM and PM data belong to the same person) of frontal sinuses are considered, since these are of great forensic interest and result in the most complex optimization scenario (as it has to model both posterioanterior and Water's views). Furthermore, it is computationally unaffordable (because of its high computational cost) to perform this experimentation also with clavicles and patallae. Meanwhile, the second experiment is devoted to compare the best configuration of each RCEA with simulated CR problems of frontal sinuses, clavicles, and patellae with P7 and P9 in order to find the best RCEA in terms of accuracy and robustness. The third part is devoted to study the identification capability of the proposed IR framework using P9 and the best resulting RCEA, in turn MVMO-SH, in real images of frontal sinuses, as well as its robustness regarding intra-expert and inter-expert segmentation errors.

The same stop criteria is established for all the RCEAs to allow a fair comparison in terms of computational resources. The optimization process ends when at least one of the following three conditions holds: (1) the maximum number of evaluations is reached. This value is set to 50,000 evaluations (it includes the evaluations performed by the LS methods); (2) the optimization process has got stuck. It is considered that the optimization process has stagnated when it has performed 10,000 evaluations without improving the fitness of the best solution; and (3) the optimization process has archived a good solution/superimposition. A solution is considered of good quality when it shows an error lower than 0.001 in terms of fitness (i.e. the 99.9% of the pixels are correctly overlapped).

All the experiments (I, II and III) have been performed on the high performance computing server Alhambra from the University of Granada composed of 1808 cores Fujitsu PRIMERGY CX250/ RX350/RX500 nodes running Red Hat Enterprise 6.4, although on average only 50 cores were available for this experimentation. Further-

more, several preliminary experiments were performed in the supercomputing center of Galicia (CESGA). It is important to remark the large computational cost of the experimentation following a rigorous experimental design of a computationally expensive optimization problem as CR. Overall, around 1,950 computation hours (81 days) were required to perform Experiments I, II and III when the 50 cores were available uninterruptedly.

VII.3.1 Simulated dataset

The dataset employed in Experiments I and II is formed by 900 simulated CR problems (i.e. 300 for each skeletal structure to be studied), each of them composed of a 3D surface model and a random 2D perspective projection (either generated with P7 or P9) of the 3D model with occlusions. In all these simulated CR problems, the GT transformation and the GT projection without occlusions are known allowing to objectively measure the quality of the superimpositions archived.

The dataset has been generated using 30 3D surface models (10 of each skeletal structure studied in this work, i.e. 10 frontal sinuses, 10 clavicles, 10 patellae) obtained as in the previous chapter (see Section VI.3.1.1). Particularly, frontal sinueses' models were obtained by manually segmenting CTs (provided by the Hospital de Castilla la Mancha, Spain) using 3D Slicer 4.5.0-1 [PHK04]. Meanwhile, clavicles and patellae' models were obtained by scanning bones (from the bone collection of the Physical Anthropology Lab at the University of Granada) using a laser range scanner (Artec $Spider^{TM}$ 3D scanner). All these 3D models were placed in their respective most frequent positions in a radiograph [BL13] (a frontal position for frontal sinus and clavicle's models, and a lateral one for patella's models). For each 3D surface model, 10 perspective projections (5 with P7 and 5 with P9) were randomly generated within the ranges showed in Table 14 (these ranges have been set based on international acquisition protocols [BL13] and are detailed in Chapeter VI with the exception of the new parameters β_x and β_y. Notice that, these parameters are set to 0 with the P7 transformation). The parameters β_x and β_y have been added to model small alignment errors in the posterioranterior view for clavicles and patellae, as well as to model posterioranterior and Water's views for frontal sinuses (as stated in Section VII.2). With frontal sinuses, the parameter β_y has a larger range to allow the optimizer to adapt automatically to both posterioranterior and Water's views. In addition, the rotation range has been increased to [-40, 40] to study the robustness of the RCEA to a greater uncertainty on the initial pose of the 3D model. These projections are generated with a resolution of 2 pixels per mm, resulting in images of 480×600 pixels for frontal sinuses and patellae, and 860×700 pixels for clavicles. Lastly, in order to model the occlusions present in real radiographs, two additional projections were generated with occlusion on the skeletal structure of 20% and 40% for each of the previous projective projections. The occlusion ranges are greater than in the previous chapter to test the RCEAs in a more complex optimization scenario.

VII.3.2 Real dataset

The dataset employed in Experiment III was provided by the *Hospital de Castilla-La Mancha*, Spain, and is composed of 180 CTs and 180 radiographs where the

Table 14: Parameter range of each skeletal structure according to international acquisition protocols [BL13] and expert knowledge.

Parameter	Bone/Cavity		
	Frontal Sinuses	Patellae	Clavicles
Image receptor dimension (mm)	240 × 300		430 mm × 350
t_x (mm)	[-125, 125]	[-125, 125]	[-210, 210]
t_y (mm)	[-150, 150]		[-175, 175]
t_z (mm)	[900 - 200, 900 + 200]		[900 - 200, 1700 + 200]
r_x, r_y, and r_z (degrees)	[-40°, 40°]		
SID (mm)	[1000 - 100, 1000 + 100]		[1800 - 100, 1800 + 100]
β_x (degrees)	[-10°, 10°]		
β_y (degrees)	[-50°, 10°]	[-10°, 10°]	

frontal sinuses are visible. The data was segmented by two forensic anthropology MSc students from the Physical Anthropology lab (PAL) of the University of Granada. All CTs were segmented by the forensic student A (Andrea Cerezo Vallecillo), and all radiographs were segmented by forensic students B (José Manuel Pérez Jiménez). Furthermore, forensic student A and B performed five segmentations of 40 CTs and 40 radiographs, respectively, to study the intra-expert segmentation error. Finally, both forensic students, A and B, also segmented 50 radiographs and 50 CTs, respectively, to study the inter-expert segmentation error.

Table 15: Summary of all the parameters of the different RCEAs and their studied values in Experimentation I

Fixed parameters					
General par.	Number of evaluations: 50,000				
DE	$p = 100$	$F = 0.5$	$P_c = 0.5$		
L-SHADE	$r^{arc} = 2$[1]	None			
CMA-ES	None				
BIPOP-CMA-ES	None				
CRO-SL	$p_0 = 0.4$[2]	$n_{LS} = 50$[3]	$F = 0.5$		
	Substrates = (Harmony search, DE, Cauchy Mutation[4], SBX, and BLX-α)				
MVMO-SH	f_{start} =1	$d_r = 1$	$GP = 5$	$p_{LS} = 0.015$	
	Parent selection strategy = sequential selection of the 1st variable, and the rest randomly[5]				
Parameters to fine-tune					**N°conf.**
DE	None. DE's parameters were already fine-tuned in Chapeter VI)				1
L-SHADE	p_{init} = (15, 20, 25)	pb = (0.05, 0.1, 0.15)	H = (2, 5, 10)		27
CMA-ES	λ & μ = (100 & 15, 40 & 15, d[6] & d[7]).		σ = (0.01, 0.1, 0.3)		9
BIPOP-CMA-ES	λ & μ = (100 & 15, 40 & 15, d[6] & d[7]).		σ = (0.01, 0.1, 0.3)		9
CRO-SL	p = (25, 50, 100)	δ = (0.1, 0.25, 0.4)			9
MVMO-SH	p = (1, 25, d[8])	A_s = (5, 10, 25)	f_{end} = (1.5, 2.5)		12

[1] other values (1, 3) were also studied in a preliminary experimentation with worst performance results.
[2] other values (0.15, 0.65) were also studied in a preliminary experimentation with worst performance results.
[3] other values (0, 100) were also studied in a preliminary experimentation with worst performance results.
[4] the Gaussian Mut. was studied as alternative to the Cauchy Mut. in preliminary experiments with worst results.
[5] the rest of selection strategies were tested in preliminary experiments with worst performance results.
[6] d = default value calculated according to the following equation: $\lambda = 4 + \lfloor 3\ln(n) \rfloor$. Thus, it is equal to 9 and 10 for P7 and P9, respectively.
[7] d = default value calculated according to the following equation: $\mu = \lambda/2$. Thus, it is equal to 4 and 5 for P7 and P9, respectively.
[8] d = default value calculated according to the following equation: 15*number_variables. Thus, it is equal to 105 and 135 for P7 and P9, respectively.

VII.3.3 Performance metrics

Two GT metrics are employed to objectively measure the quality of the superimpositions archived by RCEAs: GT DICE [Sør48] and the mean reprojection distance error (mRPD) [vdKPT+05]. The GT DICE metric measures the overlap between the GT projection's silhouette (equal to the simulated AM projection but without any occlusion) and the 2D projection's silhouette archived by the RCEA. However, the GT DICE metric and the fitness function (i.e. Masked DICE, see Section VII.2) are highly correlated (e.g. they are equal in cases without occlusions) and thus, to avoid any possible bias, the mRPD metric is also employed. mRPD is an standardized metric for the evaluation of 3D-2D IR methods by computing the retroprojection error between the transformation obtained by the RCEA and the GT transformation (see Chapter VI for further details of the utilization of mRPD in the CR problem). Notice that these metrics can be employed only in simulated CR problems since in real CR problems the GT projection and the GT transformation are unknown.

VII.3.4 Experiment I: Fine-tuning of the evolutionary algorithms for the CR problem

VII.3.4.1 Experimental set-up

This experimentation involves the application of six different RCEAs (DE, L-SHADE, CMA-ES, BIPOP-CMA-ES, CRO-SL, and MVMO-SH) and two kinds of projective transformations (P7 and P9) for each of the 300 CR cases of frontal sinuses to achieve our goal of determining the influence of the evolutionary optimizer used by the automatic CR method. As mentioned above, this experiment is meant to fine-tune the six RCEAs. While there are unsupervised methods for parameter tuning [LIDLC+16], they tend to evaluate a very large number of parameter configurations, making them infeasible for an expensive optimization problem as CR (since each configuration should be tested over the 300 CR problems to compare them rigorously). Therefore, we have utilized a grid search where the parameter values are chosen based on the recommendations present on the RCEA's original paper and on expert knowledge about its behaviour. Taking these considerations into account, the parameter grid shown in Table 15 was designed. Lastly, in order to allow for a fair comparison, every RCEA will have the same computational resources with maximum number of 50,000 evaluations.

In summary, a total of 67 parameter configurations were considered, resulting in 20.100 executions. For each of these executions, 10 independent runs were performed to study the robustness of the RCEAs for solving the CR problem due to their stochastic component. Thus, 201,000 runs (i.e. superimpositions) were performed. Each superposition takes 1,000 seconds on average, resulting in an experimentation of around 55,833 computation hours (2,326 computation days) that performed on the 50 available cores of computing server Alhambra required "only" around 1,100 computation hours (45 computation days).

VII.3.4.2 Results

Fig. 48 shows the results obtained by the different RCEAs and their configurations according to the GT DICE metric. The performance varies significantly depending on the RCEA and projective transformation in terms of mean and standard deviation values. Better results are always obtained with P7 proving that P9 is significantly more complex as stated in Section 3, which is confirmed by the Wilcoxon's test [Geh65] obtaining a p-value lower than $1 \cdot 10^{-15}$ with both metrics. CMA-ES is an exception obtaining better results with P9 but its results are still significantly worse than those provided by the other RCEAs with both P7 and P9. Nevertheless, P9 holds a greater forensic interest since it allows to model radiographical scenarios that P7 cannot model.

Studying the influence of the different parameters, it can be observed large differences for each RCEA, specially with respect to their sensibility to the parameter choice. L-SHADE presents the more robust behavior since the results are similar for the different parameter values for each one of the problems. On the contrary, CMA-ES gives very different results in P7 depending on the parameter values used (in P9 there are very similar). More in detail, the most influential parameter in CMA-ES seems to be sigma, σ, obtaining better results with higher σ values. In BIPOP-CMA-ES this tendency is increased, corroborating that σ parameter is clearly more influential in both problems. By setting an appropriate σ value, BIPOP-CMA-ES obtains for both problems better results than the majority of the remaining RCEAs but DE and MVMO-SH. MVMO-SH is very sensitive to the number of particles, p. In P7, results are clearly different with p=1 and p=25, obtaining two performance levels based on that parameter value. For P9, results show three very different performance levels, for p=1, p=25, and p=d. In both problems, the results provided by MVMO-SH with p=1 are worse than the other RCEAs but with p=25 it outperforms the majority of algorithms, and with p=d, MVMO-SH achieves the best results overall.

In terms of accuracy and robustness of the best configuration of each RCEA, the worst performing RCEA (i.e. the sixth position) is CMA-ES (best configuration: λ = 100, μ = 25, and σ = 0.3). It is followed by L-SHADE (p_{init} = 25, pb = 0.15, H = 2, and r^{arc} = 2) and CRO-SL (p = 100, and δ = 0.25) in the fifth and fourth positions, respectively, closely tied. Neither CMA-ES and L-SHADE nor CRO-SL can obtain better results than DE, the original RCEA for CR, either with P7 and P9. BIPOP-CMA-ES (λ = 100, μ = 25, and σ = 0.3) and DE are also closely tied (taking the third and second positions). Finally, the best RCEA in terms of average and standard deviation values, and confirmed by the Wilcoxon's test with p-values lower than $1 \cdot 10^{-7}$ in the comparison with all the other RCEAs, is MVMO-SH (p = d, A_s = 4, and F_{end} = 2.5).

MVMO-SH has greatly improved the previous results both in terms of accuracy and robustness with P7 (see Chapter VI). MVMO-SH has also successfully solved a more complex version of the CR problem based on the projective transformation P9, that allows us to model both posterioanterior and Water's views, as well as, being robust to occlusions up to the 40% of their silhouettes and rotation ranges of up to 80°([-40°, 40°]) in the three axis.

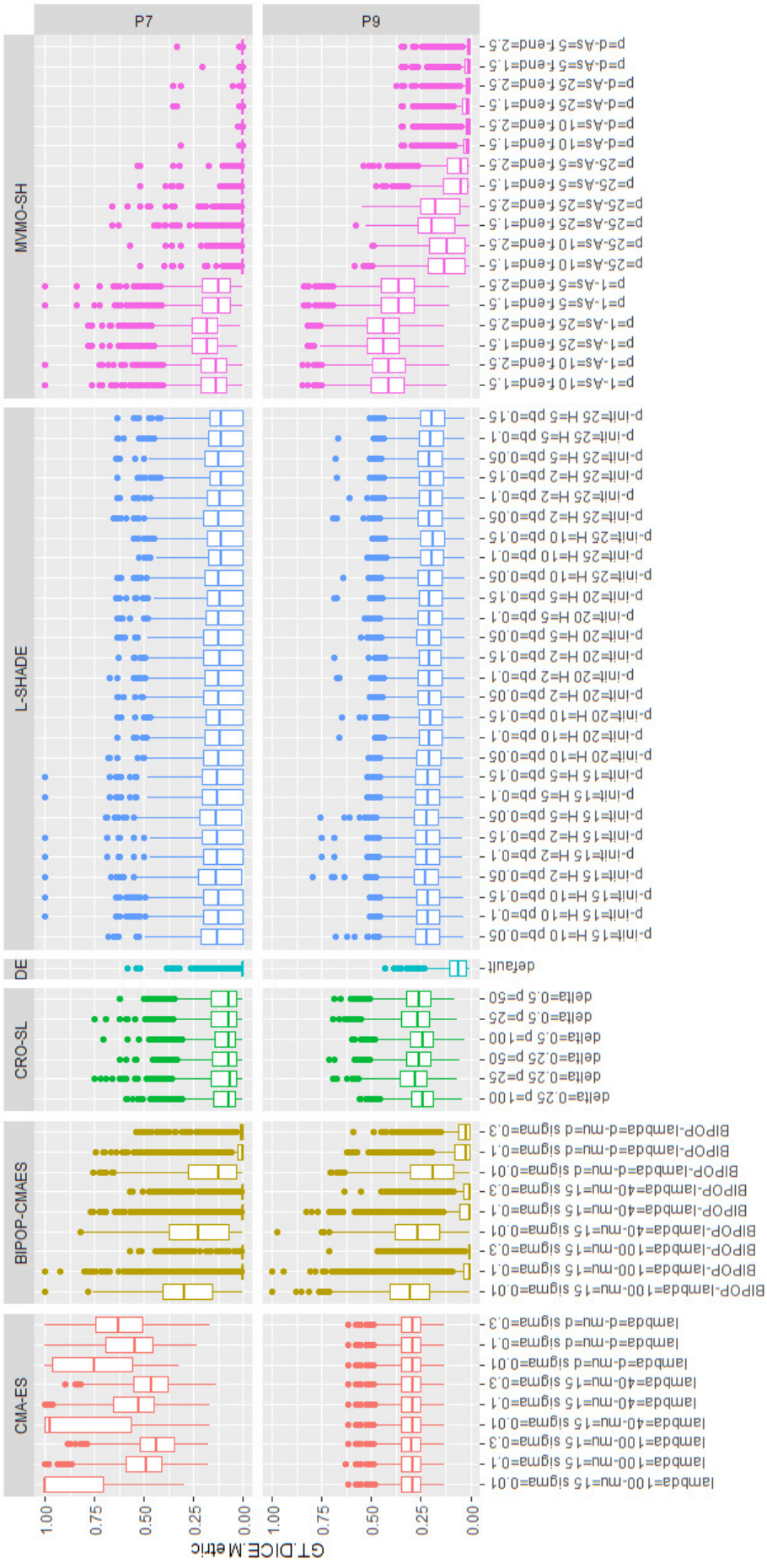

Figure 48: Boxplot of the minimum errors according to projective transformation, and optimizer for the GT DICE metric (Experiment I).

VII.3.5 Experiment II: Comparison of the RCEAs over all the CR problems

VII.3.5.1 Experimental set-up

This experimentation involves the application of the best configuration of the six different RCEAs (DE, L-SHADE, CMA-ES, BIPOP-CMA-ES, CRO-SL, and MVMO-SH) from Experiment I to all the 900 CR cases (300 frontal sinuses, 300 clavicles, and 300 patellae) using the two kinds of projective transformations (P7 and P9). The best configuration of the variable parameters of Table 15 for each algorithm are as follows:

- DE: $p = 100$, $F = 0.5$, and $P_c = 0.5$ (fine-tuned in Chapter VI).
- L-SHADE: $p_{init} = 25$, $pb = 0.15$, $H = 2$, and $r^{arc} = 2$.
- CMA-ES: $\lambda = 100$, $\mu = 25$, and $\sigma = 0.3$.
- BIPOP-CMA-ES: $\lambda = 100$, $\mu = 25$, and $\sigma = 0.3$.
- CRO-SL: $p = 100$, and $\delta = 0.25$.
- MVMO-SH: $p = \mathrm{d}$, $A_s = 4$, and $F_{end} = 2.5$.

In summary, the six RCEAs are applied to the 900 CR cases resulting in 3,000 executions. As in the first experiment, 10 independent runs are performed to avoid any possible bias caused by the stochastic component of the RCEA, resulting in 30,000 runs/superimpositions and around 200 computation hours (8 days) when performed using the 50 cores.

VII.3.5.2 Results

Table 16 shows the results obtained by the different RCEAs according to Masked DICE, GT DICE, and mRPD metrics. In view of those results, the impact of the considered skeletal structure on the RCEA's performance depicted in Chapter VI has been reduced but not eliminated. When a P7 transformation is considered, the best results are still obtained with frontal sinuses, followed by clavicles and patellae. This is probably due to the frontal sinus' silhouettes are more singular than those from clavicles and patellae. In fact, frontal sinuses are usually employed for identification [QFS+96], while clavicles and patellae are mainly employed for short listing [NSGF16, SWCT11]. However, when P9 is considered, clavicles achieve the best results since the optimization problem to solve with frontal sinuses is more complex (notice that, β_y has a range of 50°compared with the 20°of clavicles and patellae). Nevertheless, frontal sinuses are able to obtain significant results with a mean error of 0.02 (i.e. an error of only the 2% of the pixels of the silhouette) and 14 mm according to GT DICE and mRPD metrics, respectively. They also show a low standard deviation of 0.009 and 29 mm for GT DICE and mRPD metrics, respectively. As in P7, patellae had the last position due to their lower singularity.

In this experiment, MVMO is again the best RCEA for CR in terms of mean and standard deviation values, as confirmed by the Wilcoxon's test [Geh65] obtaining a

p-value equal or lower than $2 \cdot 10^{-16}$ in the comparison with the other RCEAs considering the two metrics and the three bones. The rest of the RCEAs are ranked as follows: the second best is DE; the third best is BIPOP-CMA-ES, that outperforms DE in some particular scenarios (e.g. with patellae and P9); followed by L-SHADE and CRO-SL with no significant differences between them (p-value of 0.166 in the Wilcoxon's test); and the worst results are obtained by CMA-ES.

Table 16: Summary of the results according to projective transformation, skeletal structure type, and RCEA optimizer (Experiment II).

Bone	**Opt.**	**Proj. Tran.**	**Masked DICE**		**GT DICE**		**mRPD**	
			Mean	**Sd**	**Mean**	**Sd**	**Mean**	**Sd**
Frontal Sinus	CMA-ES	P7	0.414	0.140	0.446	0.130	8.736	4.767
		P9	0.272	0.067	0.307	0.078	46.306	36.856
	BIPOP-CMA-ES	P7	0.011	0.054	0.015	0.061	0.595	1.998
		P9	0.016	0.044	0.029	0.070	15.453	28.762
	CRO-SL	P7	0.073	0.069	0.111	0.100	2.458	2.930
		P9	0.198	0.075	0.249	0.084	44.723	35.268
	DE	P7	0.008	0.034	0.015	0.048	0.307	1.396
		P9	0.048	0.040	0.076	0.066	29.024	31.458
	L-SHADE	P7	0.079	0.085	0.113	0.110	2.553	3.213
		P9	0.147	0.071	0.202	0.091	49.439	31.324
	MVMO-SH	P7	**0.001**	0.009	**0.002**	0.009	**0.047**	0.369
		P9	**0.011**	0.020	**0.021**	0.042	**14.778**	29.968
Clavicle	CMA-ES	P7	0.542	0.130	0.564	0.140	22.695	16.785
		P9	0.519	0.139	0.537	0.149	32.044	17.304
	BIPOP-CMA-ES	P7	0.089	0.186	0.109	0.220	7.063	17.111
		P9	0.132	0.220	0.155	0.246	30.573	29.887
	CRO-SL	P7	0.107	0.134	0.149	0.178	10.092	18.339
		P9	0.133	0.122	0.176	0.153	27.121	16.946
	DE	P7	0.005	0.021	0.010	0.036	0.461	3.116
		P9	0.028	0.053	0.046	0.077	23.024	15.253
	L-SHADE	P7	0.105	0.142	0.129	0.159	7.396	15.648
		P9	0.111	0.149	0.136	0.164	33.862	20.716
	MVMO-SH	P7	**0.001**	0.000	**0.002**	0.002	**0.065**	0.051
		P9	**0.004**	0.004	**0.009**	0.009	**19.383**	14.008
Patella	CMA-ES	P7	0.273	0.117	0.330	0.116	15.318	15.878
		P9	0.268	0.136	0.326	0.122	22.001	15.865
	BIPOP-CMA-ES	P7	0.016	0.024	0.045	0.063	9.486	19.605
		P9	0.019	0.028	0.053	0.072	22.163	25.054
	CRO-SL	P7	0.043	0.033	0.096	0.070	12.395	19.434
		P9	0.080	0.054	0.152	0.092	22.558	19.350
	DE	P7	0.014	0.022	0.045	0.057	7.057	16.320
		P9	0.025	0.026	0.073	0.073	21.411	20.970
	L-SHADE	P7	0.096	0.083	0.143	0.089	14.228	23.969
		P9	0.146	0.144	0.184	0.134	28.048	23.771
	MVMO-SH	P7	**0.003**	0.010	**0.009**	0.026	**2.650**	13.130
		P9	**0.006**	0.011	**0.026**	0.044	**17.151**	18.216

Table 17 shows the mean and standard deviation, according to the Masked DICE metric, of each RCEA and projective transformation after 5,000, 10,000 and 50,000

Table 17: Summary of the Masked DICE metric results according to bone/cavity type, projective transformation, and RCEA optimizer at 5,000, 10,000 and 50,000 evaluations (Experiment II).

Opt	N. Ev.	P7		P9	
		Mean	Sd	Mean	Sd
CMA-ES	5,000	0.429	0.168	0.418	0.163
	10,000	0.422	0.169	0.391	0.167
	50,000	0.410	0.169	0.351	0.169
BIPOP-CMA-ES	5,000	0.174	0.180	0.221	0.200
	10,000	0.079	0.159	0.109	0.186
	50,000	0.053	0.137	0.075	0.167
CRO-SL	5,000	**0.087**	0.098	0.168	0.120
	10,000	0.078	0.094	0.146	0.109
	50,000	0.072	0.092	0.134	0.102
DE	5,000	0.096	0.061	0.152	0.080
	10,000	**0.036**	0.036	**0.078**	0.052
	50,000	0.009	0.027	0.034	0.042
L-SHADE	5,000	0.094	0.109	**0.142**	0.130
	10,000	0.093	0.108	0.135	0.128
	50,000	0.093	0.108	0.135	0.128
MVMO-SH	5,000	0.241	0.138	0.338	0.171
	10,000	0.157	0.096	0.269	0.145
	50,000	**0.001**	0.007	**0.006**	0.013

evaluations. Meanwhile, Fig. 49 reports the average time required by the RCEAs to reach a stop condition and the average results obtained according to the GT DICE metric. In view of Table 17, the convergence speed of MVMO-SH is lower than that of the other RCEAs, needing almost all the 50,000 evaluations to obtain significant results in terms of accuracy and robustness. On the contrary, the other RCEAs have a similar performance with 10,000 and 50,000 evaluations, and the only one showing acceptable results with only 10,000 evaluations is DE. However, after the 50,000 evaluations limit, the best RCEA in terms of time is also MVMO-SH (as can also be seen in Fig. 49). In that figure, we can also observe that every algorithm except CMA-ES, and sometimes DE, does not stop due to the maximum number of evaluations condition but to the premature convergence (worse case) or good superimposition (best case) stop conditions. CRO-SL and L-SHADE stops more than 90% of times for premature convergence, while BIPOP-CMA-ES stops for good superimposition more than half of the times. The most frequent stopping condition reached by MVMO-SH is the good superimposition in 92% of all executions, while the converged condition arises in 7%, and the maximum number of evaluations condition only in 1% of runs (see Fig. 50). Thus, MVMO-SH has obtained an improvement in accuracy, robustness, convergence, and run time in the solution of the CR problem. In general, every RCEA (but CMA-ES, and DE for P9) is not limited by the maximum number of evaluations and thus no further improvements are to be expected with further run times.

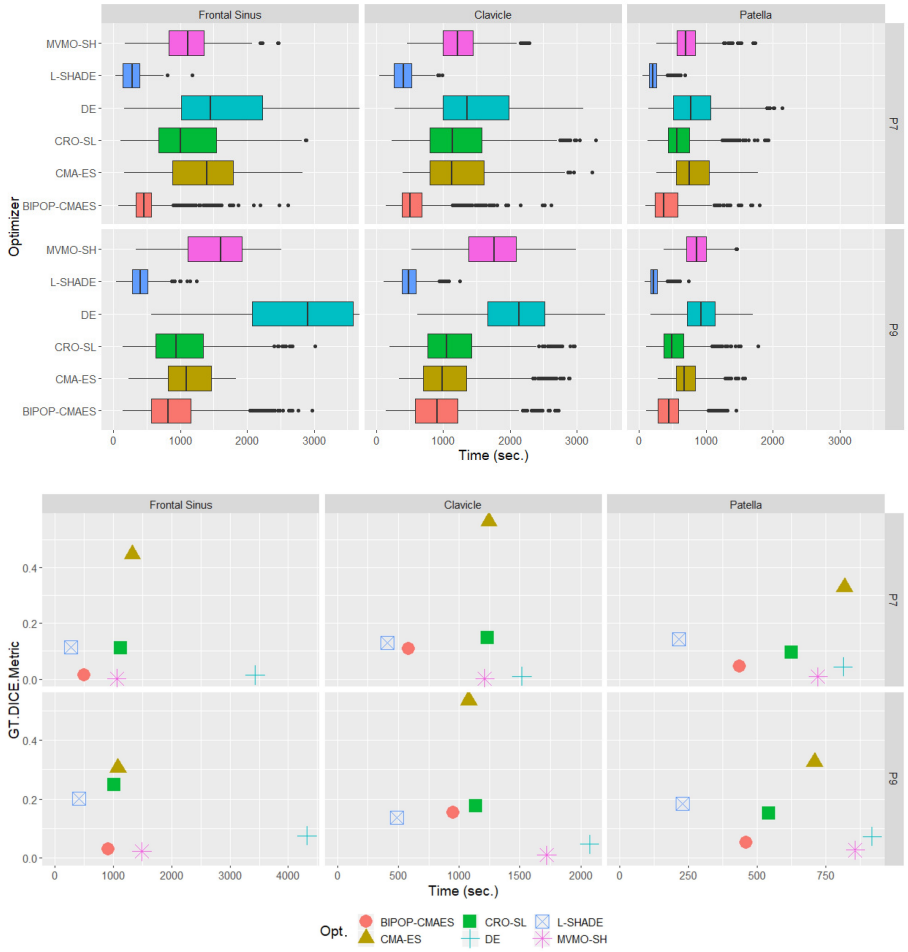

Figure 49: (Top) Boxplots of the time required to perform a superimposition according to projective transformation and RCEA optimizer. (Bottom) Relation between the average time (seconds) and the GT DICE metric according to projective transformation and RCEA optimizer (Exp. II).

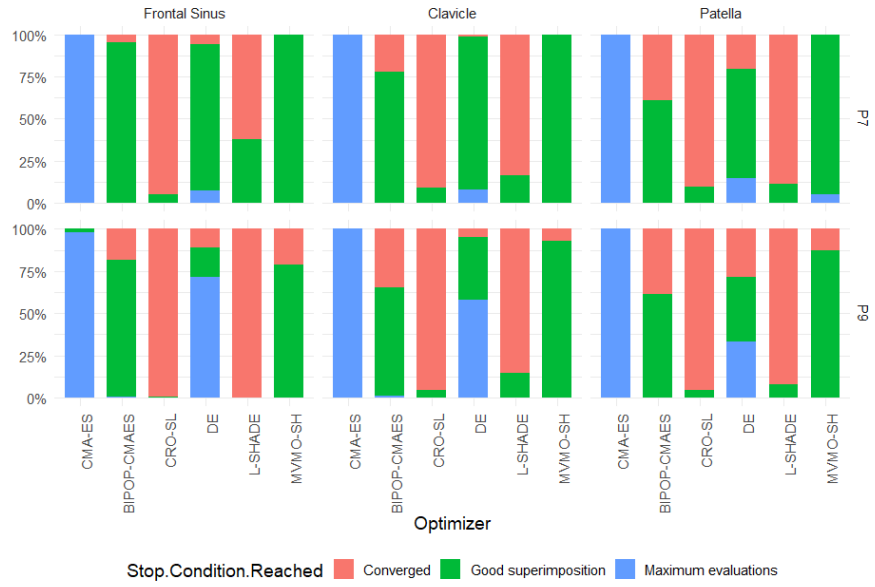

Figure 50: Boxplots of stop condition (defined in Section VII.3) reached by the optimization process according to skeletal structure, projective transformation and RCEA optimizer.

VII.3.6 Experiment III: Testing the identification capability of our 3D-2D IR-based CR framework with frontal sinuses

VII.3.6.1 Experimental set-up

This experimentation studies the identification capability of the proposed 3D-2D IR-based CR framework using frontal sinuses and the best RCEA configuration (MVMO-SH with $p = \mathrm{d}$, $A_s = 4$, $F_{end} = 2.5$, and P9). It is divided into 5 blocks or individual studies:

1. Reliability study: comparison of 180 radiographs against 180 CTs, in order to study the identification capability of frontal sinuses using the proposed IR framework.

2. Radiograph intra-expert study: comparison of 5 segmentations performed by the same forensic expert on 40 radiographs against 40 CTs.

3. Radiograph inter-expert study: comparison of 2 segmentations performed by different forensic experts on 50 radiographs against 40 CTs.

4. CT intra-expert study: comparison of 5 segmentations performed by the same forensic expert on 40 CTs against 40 radiographs.

5. CT inter-expert study: comparison of 2 segmentations performed by different forensic experts on 50 CTs against 40 radiographs.

In summary, 32,400 comparisons are performed in the reliability study (1), 8,000 in the radiograph intra-expert study (2), 5,000 in the radiograph inter-expert study (3), 8,000 in the CT intra-expert study (4), and 5,000 in the CT inter-expert study (5). A total of 58,400 CR comparisons, or CR cases. Since previous experiments have already shown the robustness of MVMO-SH, and due to the large computational cost of employing again 10 repetitions, only 2 independent runs are performed. Each of the 116,800 runs takes on average 1,000 seconds, resulting in 32,445 hours of computation (or 1,352 computation days) that, performed on the 50 available cores of computing server Alhambra, required "only" around 650 computation hours (27 computation days).

VII.3.6.2 Results

Promising results have been obtained (see Fig. 51). In the reliability study, positive and negative cases have shown important differences in terms of fitness according to the Masked DICE Metric (see Fig. 52). However, this metric alone is not sufficient to precisely distinguish between positive and negative cases.

Therefore, the results are reported using CMC curves to study the identification capabilities of the proposal as done in [CAICW18] and in Chapter VI. To focus on the identification reliability of the method only the best run (out of two) of each experiment is considered. The results of the reliability study are significant (see Fig. 53). The positive case ranks in the first position in 50% of the cross-comparisons (out of 180 candidates, 0.5% of the total sample). It ranks in the first 10 positions in 80% of the times (5.5% of the sample). Finally, to reach a confidence level of

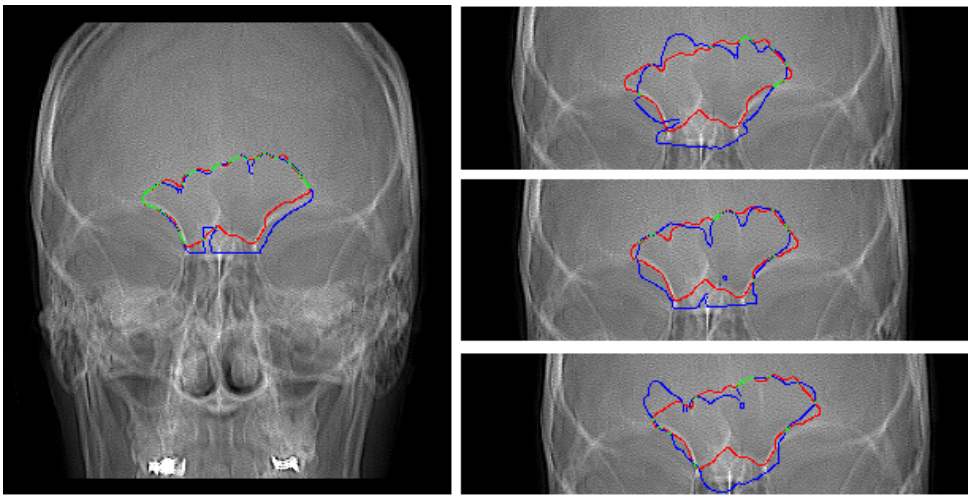

Figure 51: (Left) An example of a positive case, radiograph A compared against CT A; (Right) Example of negative cases, radiograph A compared against CTs A, B and C.

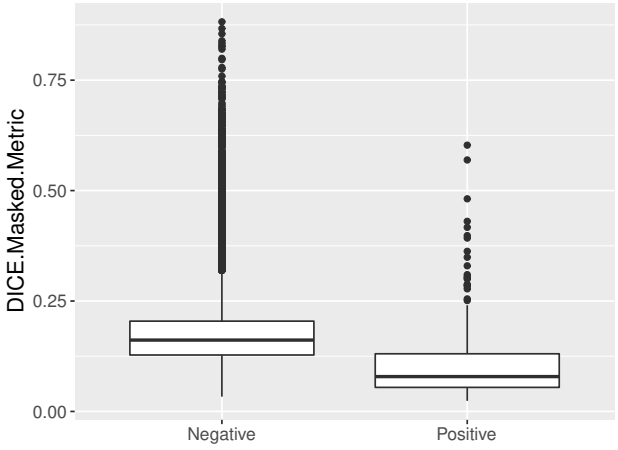

Figure 52: Boxplots of the minimum error of positive and negative cases according to the Masked DICE metric.

100% of success, we have to consider the first fifty positions (27% of the sample), i.e., in all the 32,400 cross-comparisons we will always find the positive case among the first fifty positions. One direct implication of this result is that the current framework with a very preliminary version of the decision making stage, based only on the value of the Masked DICE metric, is able to filter out 73% of the possible candidates with 0 error rate in a fully automatic manner.

Furthermore, the superimposition framework is robust to intra-expert (see Fig. 54) and inter-expert (see Fig. 55) segmentation errors in radiographs and in CTs, since results hardly vary between segmentations. In radiographs, the intra-expert error is small and barely affects the identification power of the proposed framework. Meanwhile, the inter-expert segmentation error has a greater effect, remarking the importance of automating the segmentation of radiographs. Nevertheless, the positive cases always rank within the first 30% of the cases with both segmentations. Lastly, in CTs, both the inter and intra segmentation errors are insignificant since the segmentation of skeletal structures in volumetric images is simpler and better defined than in radiographs. This is due to the fact that they do not suffer from fuzzy and overlapping boundaries as in radiographs.

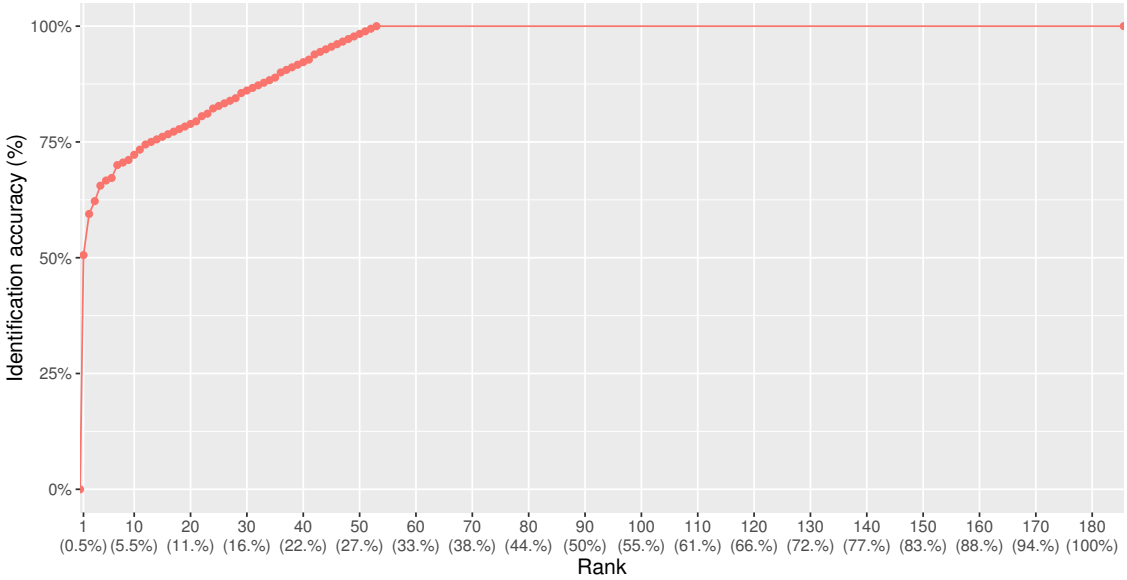

Figure 53: CMC curve of the comparison of 180 radiographs against 180 CTs.

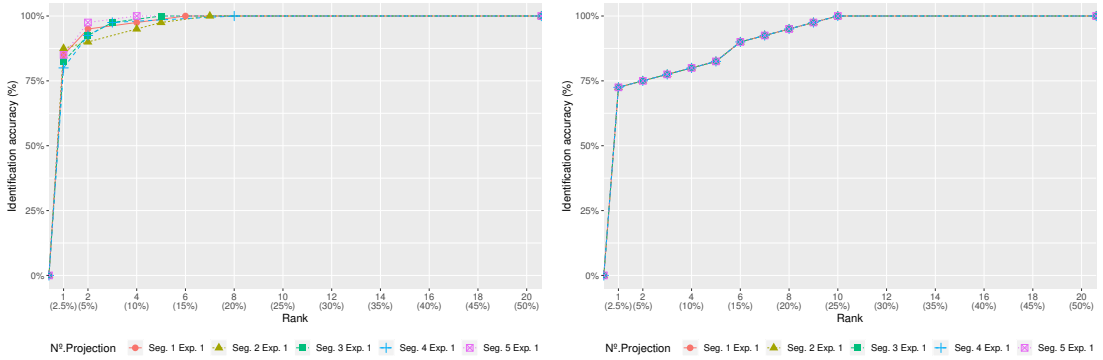

Figure 54: CMC curves of the intra-expert study (five segmentations perfomed by the same anthropologist; 40×40 comparisons per segmentation): (Left) Radiographs; (Right) CTs. In CTs, the intra-expert variation is non-existent and therefore only one curve is observed instead of five as they completely overlap.

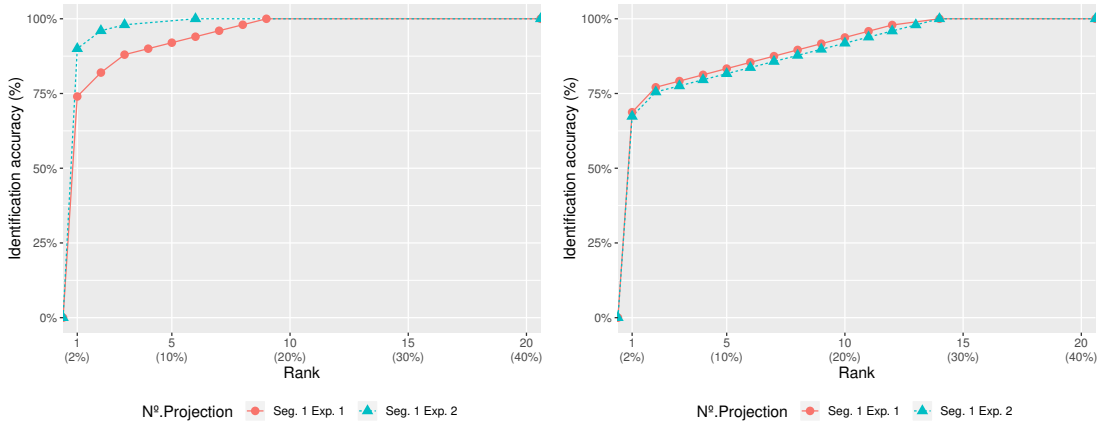

Figure 55: CMC curves of the inter-expert study (two different segmentations performed by two different anthropologists; 50×50 comparisons per segmentation): (Left) Radiographs; (Right) CTs.

Part III

Final remarks

Chapter VIII

Conclusions and future works

"And once the storm is over, you won't remember how you made it through, how you managed to survive. You won't even be sure, whether the storm is really over. But one thing is certain. When you come out of the storm, you won't be the same person who walked in. That's what this storm's all about." — Haruki Murakami

VIII.1 Conclusions

We have presented a novel framework for automating the comparison of 2D ante-mortem and 3D post-mortem materials, e.g. X-ray images and computed tomographies (or 3D surface scans), respectively. As a general description, the inputs of the proposed framework are the ante-mortem radiographs of all potential candidates and the 3D post-mortem images of the deceased (scanned with a laser range scanner or with a computed tomography scan), while the output is a short list of potential matches. To this end, we have divided the problem into three tasks or stages, segmentation, superimposition, and decision making, and we have automated each of them using artificial intelligence (in particular, soft computing) techniques.

In the first stage, skeletal structure segmentation, several new deep architectures are proposed to deal with this complex problem. First, X-Net focuses on improving the segmentation accuracy in images of 256×256 (the conventional resolution used in the literature). Second, RX-Net represents a simplification of X-Net focused on reducing the required computational resources (training memory and time) without significant loss in the original accuracy. Finally, X-Net+ and RX-Net+, are extensions of the former architectures that allow us to work with images up to 1024×1024, maintaining the original relation between filter's field-of-view and feature maps, and to transfer the learning from their precedent versions (X-Net and RX-Net, respectively, for images of 1024×1024).

Our best performing proposal for clavicle segmentation, X-Net+ for single-class segmentation obtains better results than the state-of-the-art methods with an average error of 0.884 (JI), 0.939 (DICE), and 18.022 (HD). The quantitative evaluation of our X-Net architectures by means of a rigorous experimental design protocol (10-fold cross validation, rankings and statistical tests) shows the empirical advantages of employing them in this task. Overall, the single-class training approach achieved better segmentation than the multi-class approach. This shows that multi-task learning is not always the best solution, despite its success in many other applica-

tions, and it must be analyzed for every particular problem separately. Furthermore, we have empirically shown, by comparing RX-Net and RX-Net+ for images of 1024×1024, that re-sizing a network to fit an input alters the relation between filter's field-of-view and the feature maps leading to a change in its behaviour. In our case, this change has significantly worsened the results of RX-Net obtaining the worst results among our proposals, meanwhile RX-Net+ has been ranked in the top 5. Meanwhile, X-Net+ outperforms all the other proposed architectures for the segmentation of frontal sinuses in head radiographs. Furthermore, multi-class learning approaches have shown to be better suited for the frontal sinuses and occlusion region segmentation problem, since both structures are strongly related. Multi-class X-Net+ has obtained segmentation errors of 0.668 (JI), 0.801 (DICE), and 20.799 (HD). Despite the low values of the JI and DICE metrics, the obtained segmentations are of high quality. The underlying reason is that there is not an anatomical limit between the frontal sinuses and the occlusion region. Thus, these limits vary significantly between the predicted segmentations and the GT segmentations. These variations bias the JI and DICE metrics. Nevertheless, the HD metric is more robust to the this problem regarding the fuzzy boundaries of the frontal sinuses, allowing to successfully measure the quality of the segmentation results.

The second stage is devoted to the calculation of the ante-mortem and post-mortem overlay. We have tackled the superimposition problem using an evolutionary 3D-2D IR approach based on the silhouette of the skeletal structure. We have studied three projective transformations (orthographic with 6 degrees of freedom, simple perspective with 7 degrees of freedom, and perspective with 9 degrees of freedom) and several numerical optimizers (e.g. BOBYQA), ad-hoc variants of numerical optimizer (EG-BOBYQA), and advanced RCEAs (DE, L-SHADE, CMA-ES, BIPOP-CMA-ES, CRO-SL, and MVMO-SH).

In summary, after a detailed analysis of the results obtained by the different RCEAs, we can conclude that the underlying optimization problem within comparative radiography is really complex for reasons such as the strong correlation among the parameters, their order of magnitude, the strong multimodality of the search space, and the high computational cost. We also confirmed there is a strong influence of the kind of RCEA employed. Advanced RCEAs such as CMA-ES, L-SHADE, and CRO-SL have not been able to obtain accurate results despite their good behavior in other real-world optimization problems. Nonetheless, promising results have been obtained with MVMO-SH overcoming BIPOP-CMA-ES and DE. The best configuration of MVMO-SH allowed us to obtain accurate superimpositions with an error lower than the 1% of the pixels for the simple perspective projection and lower than the 3% for the complete perspective projection in all the studied skeletal structures (frontal sinuses, clavicles, and patellae). Despite of its stochastic nature, it also showed a robust behaviour with a low standard deviation (frontal sinuses, 1% for the simple perspective projection and 4% for the complete perspective projection; clavicles, 1% with the simple and the complete perspective projection; and patellae, 3% for the simple perspective projection and 5% for complete perspective projection) according to the GT DICE metric. Furthermore, by using MVMO-SH, the strong dependency on the kind of bone or cavity is greatly reduced obtaining accurate results with every skeletal structure under study.

We have compared 180 skull radiographs against 180 skull CTs, where the frontal sinuses were segmented by forensic anthropology master students at the Physical

Anthropology lab (PAL) of the University of Granada. The positive case ranks in the first position (out of 180 candidates, 0.5% of the total sample) in 50% of the cross-comparisons. It ranks in the first 10 positions in 80% of the times (5.5% of the sample). Finally, to reach a confidence level of 100% of success, we have to consider the first 50 positions (27% of the sample). That is to say, in all the 32,400 cross-comparisons we will always find the positive case among the first 50 positions. One direct implication of the latter result is that the current framework with a very preliminary version of the decision making stage, based only on the value of the Masked DICE metric, is able to filter out 73% of the possible candidates with 0 error rate in a completely automatic way. Furthermore, the superimposition framework is robust to both intra-expert and intra-expert segmentation errors in radiographs and in CTs, since results hardly vary between segmentations (specially with CTs).

In conclusion, we have managed to automate the IS and IR tasks within the CR-identification process with promising results, as well as to design a preliminary version decision making stage, obtaining a significant performance in terms of accuracy and robustness even in the most complex version of the problems. The main drawback is the computation time required to obtain the superimpositions that, in spite of having been reduced to its minimum, is still high.

VIII.2 Future works

This is the first and probably most complex work of a future system intended to fully automate CR. However, there is still work to do before the framework is fully developed and validated:

- With respect to the **IS framework**, the first future work will be to design and validate an ad-hoc metric robust to the subjectivity of the limits between the frontal sinuses and the occlusion region. This new metric will prioritize the quality on the upper region of the frontal sinuses, due to its greater importance for identification. This new metric will allow to better train ConvNets for the given task and, consequently, to significantly improve the segmentation results in terms of utility for the CR task. Furthermore, we also aim to study the capability of our proposals to be applied to other problems, such as the segmentation of different sets of skeletal structures, different datasets, and different kinds of radiographs. Third, we aim to adapt our methods to the volumetric medical IS scenario following an approach similar to V-Net [MNA16]. Lastly, we would like to further study network simplifications following an automatic pruning approach as in ThiNet [LZZ+19].

- Meanwhile, regarding the **3D-2D IR framework**, future research is planned to reduce the run time required by studying evolutionary multi-resolution IR methods, surrogate assisted approaches [HVC16], and computation on GPUs.

- In the **decision-making** problem, we plan to fully develop and validate the designed hierarchical decision support system. In particular, we firstly plan to develop and validate the fourth (criteria) and third (superimposition) levels using frontal sinuses radiographs. Once these two levels are successfully developed and validated, we will tackle the second level (skeletal structure) using several superimpositions of frontal sinuses. Afterwards, we will validate

these three levels with other skeletal structures (such as clavicles, patellae, etc). Lastly, we will tackled the first level (subject), which will aggregate all the information available of the same subject, considering multiple skeletal structures and superimpositions.

- Once the three stages have been developed and validated independently, we plan to **validate them jointly**.
- Finally, we plan to study the identification reliability of different bones and cavities (both separately and combined) for the CR task [PTB11] through a collaboration with the Israel National Centre of Forensic Medicine and the Hebrew University of Jerusalem.

VIII.3 Publications

This section contains the scientific publications produced in the course of this PhD dissertation.

JCR-SCI indexed journal papers published (2):

- Gómez, O., Mesejo, P., Ibáñez, O., Valsecchi, A. & Cordón, O. (2019) **"Deep architectures for high-resolution multi-organ chest X-ray image segmentation"**. NEURAL COMPUTING AND APPLICATIONS (JCR 2018; impact factor: 4.664. Cat.: COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE. Pos.: 21/134. Q1.). DOI: 10.1007/ s00521-019-04532-y. *Related with the Chapter V.*
- Gómez, O., Ibáñez, O., Valsecchi, A., Cordón, O., & Kahana, T. (2018) **"3D-2D Silhouette-based Image Registration for Comparative Radiography-based Forensic Identification"**. PATTERN RECOGNITION 83:469-480 (JCR 2018; impact factor: 5.898. Cat.: COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE. Pos.: 14/134. Q1. Cat.: ENGINEERING, ELECTRICAL & ELECTRONIC. Pos.: 25/266. D1. Q1.). DOI: 10.1016/ j.patcog. 2018.06.011. *Related with the Chapter VI.*

JCR-SCI indexed journal papers submitted (1):

- Gómez, O., Ibáñez, O., Valsecchi, A., Bermejo, E., Molina, E. & Cordón, O. **"Comparative Radiography by 3D-2D evolutionary image registration: performance analysis of the real-coded evolutionary algorithm"**. Submitted in May 2019 to APPLIED SOFT COMPUTING (JCR 2018; impact factor: 4.873. Cat.: COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE Pos.: 20/134. Q1. Cat.: COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS. Pos.: 11/106. Q1.). *Related with the Chapter VII.*

Patents (1):

- Gómez, O., Ibáñez, O., Mesejo, P., Cordón, O., Damas, S. & Valsecchi, A. (2018). **Procedimiento de identificación de imágenes óseas**. Ref. Number: P201831303. Submission date: 29/12/2018. Owner institutions: University of Granada, Panacea Cooperative Research. *Related with Chapters IV, V, VI and VII.*

Conferences (8):

- Gómez, O., Ibáñez, O., Valsecchi, A., Cordón, O. & Kahana, T. **"Soft Computing and Computer Vision for Comparative Radiography in Forensic Identification"**. 6th Congress of the International Society for Forensic Radiology and Imaging (ISFRI) and 12th Anniversary Meeting of the International Association of Forensic Radiographers (IAFR), Odense (Denmark), May 2017.

- Valsecchi, A., Gómez, O., Ibáñez, O., Cordón, O. & Kahana, T. **"A computer-aided method for comparative radiography for skeleton-based identification"**. 17th Meeting of the International Association for Craniofacial Identification (IACI), Brisbane (Australia), July 2017.

- Gómez, O., Ibáñez, O., Valsecchi, A., Cordón, O. & Kahana, T. **"Método asistido por ordenador para la identificación forense mediante comparación de radiografías"**. Actas de las IX Jornadas de la Asociación Española de Antropología y Odontología Forense (AEAOF), La Rábida, October 2017.

- Gómez, O., Ibáñez, O., Valsecchi, A., Cordón, O. & Kahana, T. **"Registrado de imágenes 3D-2D para identificación forense mediante comparación de radiografías"**. XIII Congreso Español de Metaheurísticas y Algoritmos Evolutivos y Bioinspirados (MAEB), Granada (Spain), October 2018.

- Gómez, O., Ibáñez, O., Mesejo, P., Valsecchi, A. & Cordón, O. **"Soft Computing y Visión por Ordenador para la Identificación Forense mediante Comparación de Radiografías"**. II Congreso Nacional/IV Jornadas de Investigadores en Formación Fomentando la interdisciplinariedad (JIFFI), Granada (Spain), Juny 2018.

- Gómez, O., Ibáñez, O., Valsecchi, A. & Cordón, O. **"Improving comparative radiography by multi- resolution 3D-2D evolutionary image registration"**. 14th International Conference on Hybrid Artificial Intelligent Systems (HAIS), Leon (Spain), September 2019.

- Gómez, O., Mesejo, P., Ibáñez, O., Valsecchi, A. & Cordón, O. **"Automatic Segmentation of Skeletal Structures in X-ray Images using Deep Learning: Towards a Computer-aided Decision Support System for Comparative Radiography"**. 8th Biennial Meeting of the International Association of Craniofacial Identification (IACI), Online, October 2019. DOI: 10.1007/978-3-030-29859-3_9.

- Gómez, O., Ibáñez, O., Mesejo, P., Valsecchi, A. & Cordón, O. **"Hacia un sistema de apoyo a la toma de decisiones asistida por ordenador para la radiografía comparativa"**. Actas de las XI Jornadas de la Asociación Española de Antropología y Odontología Forense (AEAOF), Pastrana, November 2019.

Works related to the application of soft computing and computer vision to forensic anthropology, but not related to comparative radiography (2):

- Gómez, O., Ibáñez, O. & Cordón, O. (2017) **"Improved image registration in skull–face overlay using expert knowledge".** PROGRESS IN ARTIFICIAL INTELLIGENCE 6 (4):285-298. DOI: 10.1007/s13748-017-0124-6.
- Gómez, O., Ibáñez, O. & Cordón, O. **"Evolutive Image Registration in Craniofacial Superimposition: Modeling and Incorporating Expert Knowledge"**. XI Congreso Español de Metaheurísticas y Algoritmos Evolutivos y Bioinspirados (MAEB), Salamanca (Spain), September 2016. DOI: 10.1007/978-3-319-44636-3_33.

VIII.4 Acknowledgements

This doctoral thesis has been supported by the Spanish MECD FPU grant [grant number FPU14/02380]. This research was supported by the Spanish Ministerio de Economía y Competividad under the NEWSOCO project [Grant Number TIN2015-67661-P], including European Development Regional Funds (EDRF). This work was also supported by the Spanish Ministry of Science, Innovation and Universities, and European Regional Development Funds (ERDF) under grant EXASOCO (PGC2018-101216-B-I00). We would also like to acknowledge the high performance computing server Alhambra from the University of Granada and the supercomputing center of Galicia (CESGA), that were utilized in this work. We would also like to thank to Tzipi Kahana from the The Hebrew University of Jerusalem for her professional advice on CR. We would also like to thank to the Institut National de Recherche en Informatique et en Automatique (INRIA) for their collaboration during the doctoral stay of this PhD dissertation. Lastly, we would also like to acknowledge to Andrea Cerezo Vallecillo and José Manuel Pérez Jiménez, MSc students from the Physical Anthropology lab (PAL) of the University of Granada, for segmenting frontal sinuses in radiographs and CTs. These segmentations have allowed us to carry out some of the experiments described in this PhD dissertation.

Chapter IX

Bibliography

[3SE] 3D scan expert. `https://3dscanexpert.com/realitycapture-photogrammetry-software-review/`. Accessed: 2019-07-07.

[AD18] Prem Anuja and Nagabhushana Doggalli. Software in forensic odontology. *Indian Journal of Multidisciplinary Dentistry*, 8(2):94, 2018.

[Ada03] Bradley J Adams. Establishing personal identification based on specific patterns of missing, filled, and unrestored teeth. *Journal of Forensic Sciences*, 48(3):487–496, 2003.

[AHS85] David H Ackley, Geoffrey E Hinton, and Terrence J Sejnowski. A learning algorithm for Boltzmann machines. *Cognitive Science*, 9(1):147–169, 1985.

[ARG+10] Santiago Crespo Alonso, Victor Cosialls Roca, Josep Castellà García, Olga Martínez Castillo, Albert Rousset Berrecosa, María Placer Cabaleiro Dieguez, and Jordi Medallo Muñiz. Identificación mediante estudio comparativo de radiografías craneales ante mortem y post mortem (in spanish). *Revista Española de Medicina Legal*, 36(1):35–40, 2010.

[ATY+19] Zahangir Alom, Tarek M Taha, Chris Yakopcic, Stefan Westberg, Paheding Sidike, Shamima Nasrin, Mahmudul Hasan, Brian C Van Essen, Abdul AS Awwal, and Vijayan K Asari. A state-of-the-art survey on deep learning theory and architectures. *Electronics*, 8(3):292, 2019.

[B+95] Christopher M Bishop et al. *Neural networks for pattern recognition*. Oxford university press, 1995.

[BCD+18] Enrique Bermejo, Manuel Chica, Sergio Damas, Sancho Salcedo-Sanz, and Oscar Cordón. Coral reef optimization with substrate layers for medical image registration. *Swarm and Evolutionary Computation*, 42:138–159, 2018.

[BDBP15] Sunil L Bangare, Amruta Dubal, Pallavi S Bangare, and ST Patil. Reviewing Otsu's method for image thresholding. *International Journal of Applied Engineering Research*, 10(9):21777–21783, 2015.

[Ber18] Enrique Bermejo. *New developments in evolutionary image registration for complex 3D scenarios.* PhD thesis, Universidad de Granada, 2018.

[Bey13] Hans-Georg Beyer. *The theory of evolution strategies.* Springer Science & Business Media, 2013.

[BFK18] Lei Bi, Dagan Feng, and Jinman Kim. Dual-path adversarial learning for fully convolutional network (FCN)-based medical image segmentation. *The Visual Computer*, 34(6):1043–1052, 2018.

[BFME97] Thomas Bäck, David B Fogel, and Zbigniew Michalewicz (Eds.). *Handbook of evolutionary computation.* CRC Press, 1997.

[BKP14] Haithem Boussaid, Iasonas Kokkinos, and Nikos Paragios. Discriminative learning of deformable contour models. In *ISBI*, pages 624–628. IEEE, 2014.

[BL00] Byron Gilliam Brogdon and Joel E Lichtenstein. Forensic radiology in historical perspective. *Critical Reviews in Diagnostic Imaging*, 41(1):13–42, 2000.

[BL13] Kenneth L Bontrager and John Lampignano. *Textbook of radiographic positioning and related anatomy.* Elsevier Health Sciences, 2013.

[BLMM12] Adrian Brady, Risteárd Ó Laoide, Peter McCarthy, and Ronan McDermott. Discrepancy and error in radiology: concepts, causes and consequences. *Ulster Med J*, 81(1):3, 2012.

[BMR+14] Alison L Brough, Bruno Morgan, Claire Robinson, Sue Black, Craig Cunningham, Catherine Adams, and Guy N Rutty. A minimum data set approach to post-mortem computed tomography reporting for anthropological biological profiling. *Forensic science, medicine, and pathology*, 10(4):504–512, 2014.

[Bow01] C Michael Bowers. Jurisprudence issues in forensic odontology. *Dental Clinics of North America*, 45(2):399–415, 2001.

[BPC+07] Gleb Beliakov, Ana Pradera, Tomasa Calvo, et al. *Aggregation functions: a guide for practitioners*, volume 221. Springer, 2007.

[BR10] Joanna L Besana and Tracy L Rogers. Personal identification using the frontal sinus. *Journal of Forensic Sciences*, 55(3):584–589, 2010.

[BSJ96] Lynne S Bell, Mark F Skinner, and Sheila J Jones. The speed of post mortem change to the human skeleton and its taphonomic significance. *Forensic Science International*, 82(2):129–140, 1996.

[BTE98] Mario Beauchemin, Keith PB Thomson, and G Edwards. On the Hausdorff distance used for the evaluation of segmentation results. *Can J Remote Sens*, 24(1):3–8, 1998.

[BWA15] Michael A Bruno, Eric A Walker, and Hani H Abujudeh. Understanding and confronting our mistakes: the epidemiology of error in radiology and strategies for error reduction. *Radiographics*, 35(6):1668–1676, 2015.

[CAICW18] Carmen Campomanes-Alvarez, Oscar Ibáñez, Oscar Cordón, and Caroline Wilkinson. Hierarchical information fusion for decision making in craniofacial superimposition. *Information Fusion*, 39:25–40, 2018.

[CBS17] Jodi Caple, John Byrd, and Carl N Stephan. Elliptical fourier analysis: fundamentals, applications, and value for forensic anthropology. *International Journal of Legal Medicine*, 131(6):1675–1690, 2017.

[CC17] Eugénia Cunha and Cristina Cattaneo. *Historical Routes and Current Practice for Personal Identification*, pages 398–411. Springer International Publishing, Cham, 2017.

[CDLB+15] Gianguido Cossellu, Stefano De Luca, Roberto Biagi, Giampietro Farronato, Mariano Cingolani, Luigi Ferrante, and Roberto Cameriere. Reliability of frontal sinus by cone beam-computed tomography (CBCT) for individual identification. *La Radiologia Medica*, 120(12):1130–1136, 2015.

[CDS06] Oscar Cordón, Sergio Damas, and Jose Santamaría. A fast and accurate approach for 3D image registration using the scatter search evolutionary algorithm. *Pattern Recognition Letters*, 27(11):1191–1200, 2006.

[CDS07] Oscar Cordón, Sergio Damas, and José Santamaría. A practical review on the applicability of different EAs to 3D feature-based registration. *Genetic and evolutionary computation in image processing and computer vision. EURASIP Book Series on SP&C*, pages 247–269, 2007.

[CFM+05] Roberto Cameriere, Luigi Ferrante, Dora Mirtella, Franco Ugo Rollo, and Mariano Cingolani. Frontal sinuses for identification: quality of classifications, possible error and potential corrections. *Journal of Forensic Sciences*, 50(4):770–773, 2005.

[CFS+02] JB Cornelison, TW Fenton, NJ Sauer, JL deJong, and BC Hunter. Comparative radiography of the lateral hyoid: a new method for

human identification. In *Proceedings of the Annual Meeting of the American Academy of Forensic Sciences*, pages 11–16, 2002.

[CGC11] Romina Ciaffi, Daniele Gibelli, and Cristina Cattaneo. Forensic radiology and personal identification of unidentified bodies: a review. *La Radiologia Medica*, 116(6):960–968, 2011.

[CH16] Angi M Christensen and Gary M Hatch. Quantification of radiologic identification (radid) and the development of a population frequency data repository. *Journal of Forensic Radiology and Imaging*, 7:14–16, 2016.

[CH18] Angi M Christensen and Gary M Hatch. Advances in the use of frontal sinuses for human identification. In *New Perspectives in Forensic Human Skeletal Identification*, pages 227–240. Elsevier, 2018.

[Cha08] Uday Chakraborty. *Advances in differential evolution*, volume 143. Springer, 2008.

[Chr04] Angi M Christensen. The impact of Daubert: Implications for testimony and research in forensic anthropology (and the use of frontal sinuses in personal identification). *Journal of Forensic Sciences*, 49(3):427–430, 2004.

[Chr05a] Angi M Christensen. Assessing the variation in individual frontal sinus outlines. *American Journal of Physical Anthropology*, 127(3):291–295, 2005.

[Chr05b] Angi M Christensen. Testing the reliability of frontal sinuses in positive identification. *Journal of Forensic Sciences*, 50(1):18–22, 2005.

[CL27] William Ledlie Culbert and Frederick M Law. Identification by comparison of roentgenograms: of nasal accessory sinuses and mastoid processes. *Journal of the American Medical Association*, 88(21):1634–1636, 1927.

[COL+11] Mariano Cabezas, Arnau Oliver, Xavier Lladó, Jordi Freixenet, and Meritxell Bach Cuadra. A review of atlas-based segmentation for magnetic resonance brain images. *Comput Methods Programs Biomed*, 104(3):e158 – e177, 2011.

[CPK+14] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv preprint arXiv:1412.7062*, 2014.

[CPK+16] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *CoRR*, abs/1606.00915, 2016.

[CPK+17] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):834–848, 2017.

[CPSA17] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.

[CSG+18] Angi M Christensen, Michael A Smith, Devora S Gleiber, Deborah L Cunningham, and Daniel J Wescott. The use of x-ray computed tomography technologies in forensic anthropology. *Forensic Anthropology*, 1(2):124, 2018.

[CYRC18] Peter Chondro, Chih-Yuan Yao, Shanq-Jang Ruan, and Li-Chien Chien. Low order adaptive region growing for lung segmentation on plain chest radiographs. *Neurocomputing*, 275:1002–1011, 2018.

[CZP+18] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 801–818, 2018.

[DBKK09] Elizabeth A DiGangi, Jonathan D Bethard, Erin H Kimmerle, and Lyle W Konigsberg. A new method for estimating age-at-death from the first rib. *American Journal of Physical Anthropology*, 138(2):164–176, 2009.

[DCI+11] Sergio Damas, Oscar Cordón, Oscar Ibáñez, Jose Santamaría, Inmaculada Alemán, Miguel Botella, and Fernando Navarro. Forensic identification by computer-aided craniofacial superimposition: a survey. *ACM Computing Surveys*, 43(4):1–27, 2011.

[DCS11] Sergio Damas, Oscar Cordón, and José Santamaría. Medical image registration using evolutionary computation: An experimental survey. *IEEE Computational Intelligence Magazine*, 6(4):26–42, 2011.

[DDC+19] Lucile Deloire, Idris Diallo, Romain Cadieu, Mathieu Auffret, Zarrin Alavi, Julien Ognard, and Douraïed Ben Salem. Post-mortem X-ray computed tomography (PMCT) identification using ante-mortem CT-scan of the sphenoid sinus. *Journal of Neuroradiology*, 46(4):248–255, 2019.

[DDS+09] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. IEEE, 2009.

[DF19] Summer J Decker and Jonathan M Ford. Forensic personal identification utilizing part-to-part comparison of CT-derived 3D lumbar models. *Forensic Science International*, 294:21–26, 2019.

[DFDCMT08] Ivanoe De Falco, Antonio Della Cioppa, Domenico Maisto, and Ernesto Tarantino. Differential evolution as a viable tool for satellite image registration. *Applied Soft Computing*, 8(4):1453–1462, 2008.

[DGBS17] Susan S D'alonzo, Pierre Guyomarc'h, John E Byrd, and Carl N Stephan. A Large-Sample Test of a Semi-Automated Clavicle Search Engine to Assist Skeletal Identification by Radiograph Comparison. *Journal of Forensic Sciences*, 62(1):181–186, 2017.

[DHG18] Sharon M Derrick, John A Hipp, and Priya Goel. The computer-assisted decedent identification method of computer-assisted radiographic identification. In *New Perspectives in Forensic Human Skeletal Identification*, pages 265–276. Elsevier, 2018.

[DJV+06] Richard Dirnhofer, Christian Jackowski, Peter Vock, Kimberlee Potter, and Michael J Thali. Virtopsy: minimally invasive, imaging-guided virtual autopsy. *Radiographics*, 26(5):1305–1333, 2006.

[DM46] Wilfrid J Dixon and Alexander M Mood. The statistical sign test. *Journal of the American Statistical Association*, 41(236):557–566, 1946.

[DS11] Swagatam Das and Ponnuthurai Nagaratnam Suganthan. Differential evolution: A survey of the state-of-the-art. *IEEE Transactions on Evolutionary Computation*, 15(1):4–31, 2011.

[DSD+13] Misha Denil, Babak Shakibi, Laurent Dinh, Nando De Freitas, et al. Predicting parameters in deep learning. In *Advances in neural information processing systems*, pages 2148–2156, 2013.

[dSPC+09] Rhonan Ferreira da Silva, Felippe Bevilacqua Prado, Isamara Geandra Cavalcanti Caputo, Karina Lopes Devito, Tessa de Luscena Botelho, and Eduardo Daruge Júnior. The forensic importance of frontal sinus radiographs. *Journal of Forensic and Legal Medicine*, 16(1):18–23, 2009.

[DSSF+17] Ivan Nunes Da Silva, Danilo Hernane Spatti, Rogerio Andrade Flauzino, Luisa Helena Bartocci Liboni, and Silas Franco dos Reis Alves. Artificial neural networks. *Cham: Springer International Publishing*, 2017.

[DTM+11] Parvathi Devi, VB Thimmarasa, Vishal Mehrotra, Vikas Singla, et al. Automated dental identification system: an aid to forensic odontology. *Journal of Indian Academy of Oral Medicine and Radiology*, 23(5):360, 2011.

[Dut44] Frank R Dutra. Identification of person and determination of cause of death from skeletal remains. *Arch Pathol*, 38:339–349, 1944.

[EEVG+15] Mark Everingham, SM Ali Eslami, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 111(1):98–136, 2015.

[Eng16] NHS England. Diagnostic imaging dataset annual statistical release 2015/2016. 2016.

[ERWS14] István Erlich, José L Rueda, Sebastian Wildenhues, and Fekadu Shewarega. Solving the IEEE-CEC 2014 expensive optimization test problems by using single-particle MVMO. In *2014 IEEE Congress on Evolutionary Computation (CEC)*, pages 1084–1091. IEEE, 2014.

[EVW10] István Erlich, Ganesh K Venayagamoorthy, and Nakawiro Worawat. A mean-variance optimization algorithm. In *2010 IEEE Congress on Evolutionary Computation (CEC)*, pages 1–6, 2010.

[FAB95] Jacques Feldmar, Nicholas Ayache, and Fabienne Betting. 3D-2D projective registration of free-form curves and surfaces. In *Proceedings of the Fifth International Conference on Computer Vision 1995*, pages 549–556. IEEE, 1995.

[FB04] Alexandre X Falcão and Felipe PG Bergo. Interactive volume segmentation with differential image foresting transforms. *IEEE Transactions on Medical Imaging*, 23(9):1100–1108, 2004.

[FD16] Jonathan M. Ford and Summer J. Decker. Computed tomography slice thickness and its effects on three-dimensional reconstruction of anatomical structures. *Journal of Forensic Radiology and Imaging*, 4:43 – 46, 2016. Special Issue: Papers from the ISFRI Conference 2015.

[FDGR11] Zacharias Fourie, Janalt Damstra, Peter O. Gerrits, and Yijin Ren. Evaluation of anthropometric accuracy and reliability using different three-dimensional scanning systems. *Forensic Science International*, 207(1):127 – 134, 2011.

[fFAS10] Scientific Working Group for Forensic Anthropology (SWGANTH). Personal identification. `https://www.nist.gov/sites/default/files/documents/2018/03/13/swganth_personal_identification.pdf`, 2010.

[FFM07] Juan Rogelio Falguera, A Pereira Sartori Falguera, and Aparecido Nilceu Marana. Frontal sinus recognition using image foresting transform and shape context. 2007.

[FFM08] Juan Rogelio Falguera, Fernanda Pereira Sartori Falguera, and Aparecido Nilceu Marana. Frontal sinus recognition for human identification. In *Biometric Technology for Human Identification V*, volume 6944, page 69440S. International Society for Optics and Photonics, 2008.

[FH99] Kenneth R Foster and Peter W Huber. *Judging science: Scientific knowledge and the federal courts*. Mit Press, 1999.

[FSF16] Daniel Franklin, Lauren Swift, and Ambika Flavel. 'Virtual anthropology' and radiographic imaging in the Forensic Medical Sciences. *Egyptian Journal of Forensic Sciences*, 6(2):31 – 43, 2016. Advances in Forensic Anthropology.

[GBC10] Costantino Grana, Daniele Borghesani, and Rita Cucchiara. Optimized block-based connected components labeling with decision trees. *IEEE Trans Image Process*, 19(6):1596–1609, 2010.

[GBC16] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[GCC+19] Daniele Gibelli, Michaela Cellina, Annalisa Cappella, Stefano Gibelli, Marta Maria Panzeri, Antonio Giancarlo Oliva, Giovanni Termine, Danilo De Angelis, Cristina Cattaneo, and Chiarella Sforza. An innovative 3D-3D superimposition for assessing anatomical uniqueness of frontal sinuses through segmentation on CT scans. *International Journal of Legal Medicine*, 133(4):1159–1165, 2019.

[GD18] Mairéad Grogan and Rozenn Dahyot. Shape registration with directional data. *Pattern Recognition*, 79:452–466, 2018.

[GDGS+16] Salvatore Gerbino, Domenico Maria Del Giudice, Gabriele Staiano, Antonio Lanzotti, and Massimo Martorelli. On the influence of scanning factors on the laser scanner-based 3D inspection process. *The International Journal of Advanced Manufacturing Technology*, 84(9-12):1787–1799, 2016.

[Geh65] Edmund A Gehan. A generalized wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika*, 52(1-2):203–223, 1965.

[GGOEO+17] Alberto Garcia-Garcia, Sergio Orts-Escolano, Sergiu Oprea, Victor Villena-Martinez, and Jose Garcia-Rodriguez. A review on deep learning techniques applied to semantic segmentation. *arXiv preprint arXiv:1704.06857*, 2017.

[GJT08] Bryan H. Derrickson Gerard J. Tortora. *Principles of Anatomy and Physiology [With A Brief Atlas of the Skeleton, Surface Anatomy]*. Wiley, 12 edition, 2008.

[GLL+18] Ke Gong, Xiaodan Liang, Yicheng Li, Yimin Chen, Ming Yang, and Liang Lin. Instance-level human parsing via part grouping network. *arXiv preprint arXiv:1808.00157*, 2018.

[GLO+16] Yanming Guo, Yu Liu, Ard Oerlemans, Songyang Lao, Song Wu, and Michael S Lew. Deep learning for visual understanding: A review. *Neurocomputing*, 187:27–48, 2016.

[Goo95] Philip C Goodman. The new light: discovery and introduction of the x-ray. *AJR. American Journal of Roentgenology*, 165(5):1041–1045, 1995.

[GPF+18] Dominic Gascho, Hinderberger Philipp, Patricia M Flach, Michael J Thali, and Sören Kottner. Standardized medical image registration for radiological identification of decedents based on paranasal sinuses. *Journal of Forensic and Legal Medicine*, 54:96–101, 2018.

[Han09a] Nikolaus Hansen. Benchmarking a BI-population CMA-ES on the BBOB-2009 function testbed. In *Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers*, pages 2389–2396. ACM, 2009.

[Han09b] Nikolaus Hansen. Benchmarking a BI-population CMA-ES on the BBOB-2009 noisy testbed. In *Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers*, pages 2397–2402. ACM, 2009.

[HCO+17] Alexandre Hacl, André Luiz Ferreira Costa, Juliane Mayara Oliveira, Maria José Tucunduva, José Raul Girondi, and Ana Carla Raphaelli Nahás-Scocate. Three-dimensional volumetric analysis of frontal sinus using medical software. *Journal of Forensic Radiology and Imaging*, 11:1–5, 2017.

[HDC+14] Gary M. Hatch, Fabrice Dedouit, Angi M. Christensen, Michael J. Thali, and Thomas D. Ruder. RADid: A pictorial review of radiologic identification using postmortem CT. *Journal of Forensic Radiology and Imaging*, 2(2):52 – 59, 2014.

[Her09] Gabor T Herman. *Fundamentals of computerized tomography: image reconstruction from projections*. Springer Science & Business Media, 2009.

[HK04] Nikolaus Hansen and Stefan Kern. Evaluating the CMA evolution strategy on multimodal test functions. In *International Conference on Parallel Problem Solving from Nature*, pages 282–291. Springer, 2004.

[HLM+16] Song Han, Xingyu Liu, Huizi Mao, Jing Pu, Ardavan Pedram, Mark A Horowitz, and William J. Dally. EIE: efficient inference engine on compressed deep neural network. *ISCA*, pages 243–254, 2016.

[HMK03] Nikolaus Hansen, Sibylle D Müller, and Petros Koumoutsakos. Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). *Evolutionary Computation*, 11(1):1–18, 2003.

[HMS18] Rahul Hooda, Ajay Mittal, and Sanjeev Sofat. An efficient variant of fully-convolutional network for segmenting lung fields from chest radiographs. *Wireless Personal Communications*, 101(3):1559–1579, 2018.

[HOT06] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural Computation*, 18(7):1527–1554, 2006.

[HSdJ+12] Laurens Hogeweg, Clara I Sánchez, Pim A de Jong, Pragnya Maduskar, and Bram van Ginneken. Clavicle segmentation in chest radiographs. *Med Image Anal*, 16(8):1490–1502, 2012.

[HVC16] Raphael T Haftka, Diane Villanueva, and Anirban Chaudhuri. Parallel surrogate-assisted global optimization with expensive functions–a survey. *Structural and Multidisciplinary Optimization*, 54(1):3–13, 2016.

[HW99] Myles Hollander and Douglas A Wolfe. *Nonparametric statistical methods*. Wiley-Interscience, 1999.

[HZ03] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, 2003.

[HZRS16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[ICD12] Oscar Ibáñez, Oscar Cordón, and Sergio Damas. A cooperative coevolutionary approach dealing with the skull–face overlay uncertainty in forensic identification by craniofacial superimposition. *Soft Computing*, 16(5):797–808, 2012.

[IFY+16] Morio Iino, Hideko Fujimoto, Motonori Yoshida, Hiroshi Matsumoto, and Masaki Q Fujita. Identification of a jawless skull by superimposing post-mortem and ante-mortem ct. *Journal of Forensic Radiology and Imaging*, 6:31–37, 2016.

[InBC+09] Oscar Ibáñez, Lucia Ballerini, Oscar Cordón, Sergio Damas, and José Santamaría. An experimental study on the applicability of evolutionary algorithms to craniofacial superimposition in forensic

identification. *Information Sciences*, 179(23):3998–4028, November 2009.

[Int18] Interpol. Disaster victim identification guide. *Available at: https://www.interpol.int/en/How-we-work/Forensics/Disaster-Victim-Identification-DVI*, 2018.

[IS15] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *ICML*, 1:448–456, 2015.

[Jac12] Paul Jaccard. The distribution of the flora in the alpine zone. 1. *New Phytologist*, 11(2):37–50, 1912.

[JAM96] Medico-Legal uses of the Roentgen Rays. *Journal of the American Medical Association (JAMA)*, XXVI(22):1084–1084, 1896.

[JBVH+06] Julien Jomier, Elizabeth Bullitt, Mark Van Horn, Chetna Pathak, and Stephen R Aylward. 3D/2D model-to-image registration applied to TIPS surgery. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 662–669. Springer, 2006.

[JC04] Anil K Jain and Hong Chen. Matching of dental x-ray images for human identification. *Pattern Recognition*, 37(7):1519–1532, 2004.

[JCA+14] Stefan Jaeger, Sema Candemir, Sameer Antani, Yì-Xiáng J Wáng, Pu-Xuan Lu, and George Thoma. Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. *Quant Imaging Med Surg*, 4(6):475, 2014.

[JH96] Anil Jain and Lin Hong. On-line fingerprint verification. In *Proceedings of 13th International Conference on Pattern Recognition*, volume 3, pages 596–600. IEEE, 1996.

[JJLL15] Zhaojie Ju, Xiaofei Ji, Jing Li, and Honghai Liu. An integrative framework of human hand gesture segmentation for human–robot interaction. *IEEE Systems Journal*, 11(3):1326–1336, 2015.

[JLJW14] Jia Jia, Juan Liu, Guofan Jin, and Yongtian Wang. Fast and effective occlusion culling for 3D holographic displays by inverse orthographic projection with low angular sampling. *Applied Optics*, 53(27):6287–6293, 2014.

[JLK06] Xudong Jiang, Manhua Liu, and Alex C Kot. Fingerprint retrieval for identification. *IEEE Transactions on Information Forensics and Security*, 1(4):532–542, 2006.

[Jor86] Michael Jordan. Attractor dynamics and parallelism in a connectionist sequential machine. In *Proc. of the Eighth Annual Conference of the Cognitive Science Society (Erlbaum, Hillsdale, NJ), 1986*, 1986.

[Kah09] Tzipi Kahana. *El aporte de la radiología al avance de la Antropología forense: perspectiva profesional (in Spanish).* PhD thesis, 2009.

[KB14] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[KGH02] Tzipi Kahana, L Goldin, and J Hiss. Personal identification based on radiographic vertebral features. *The American Journal of Forensic Medicine and Pathology*, 23(1):36–41, 2002.

[KH97] T Kahana and J Hiss. Identification of human remains: forensic radiology. *Journal of Clinical Forensic Medicine*, 4(1):7–15, 1997.

[KH02] Changick Kim and Jenq-Neng Hwang. Fast and automatic video object segmentation and tracking for content-based applications. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(2):122–129, 2002.

[Kha14] Muhammad Waseem Khan. A survey: Image segmentation techniques. *International Journal of Future Computer and Communication*, 3(2):89, 2014.

[KHS98] Tzipi Kahana, Jehuda Hiss, and Patricia Smith. Quantitative assessment of trabecular bone pattern identification. *Journal of Forensic Sciences*, 43(6):1144–1147, 1998.

[Kin12] Davis King. Dlib c++ library. *Available at: http://dlib.net*, 2012.

[Kir84] Robert Charles Kirkpatrick. *The New Photography: with Report of a Case in which a Bullet was Photographed in the Leg.* Montreal, 1984.

[KK11] Philipp Krähenbühl and Vladlen Koltun. Efficient inference in fully connected CRFs with gaussian edge potentials. In *Advances in neural information processing systems*, pages 109–117, 2011.

[KLP+13] Deog-Im Kim, U-Young Lee, Sang-Ouk Park, Dae-Soon Kwak, and Seung-Ho Han. Identification using frontal sinus by three-dimensional reconstruction from computed tomography. *Journal of Forensic Sciences*, 58(1):5–12, 2013.

[KPY+15] Yong-Deok Kim, Eunhyeok Park, Sungjoo Yoo, Taelim Choi, Lu Yang, and Dongjun Shin. Compression of deep convolutional neural networks for fast and low power mobile applications. *arXiv preprint arXiv:1511.06530*, 2015.

[KSF05] Michael G Koot, Norman J Sauer, and Todd W Fenton. Radiographic human identification using bones of the hand: A validation study. *Journal of Forensic Sciences*, 50(2):263–268, 2005.

[KSH12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[KTK08] Sudhir Kapoor, Akshay Tiwari, and Saurabh Kapoor. Primary tumours and tumorous lesions of clavicle. *Int Orthop*, 32(6):829, 2008.

[KWG02] Nigel J Kirk, Robert E Wood, and Marc Goldstein. Skeletal identification using the frontal sinus region: a retrospective study of 39 cases. *Journal of Forensic Sciences*, 47(2):318–323, 2002.

[LBD+89] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4):541–551, 1989.

[LBD+90] Yann LeCun, Bernhard E Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne E Hubbard, and Lawrence D Jackel. Handwritten digit recognition with a back-propagation network. In *Advances in neural information processing systems*, pages 396–404, 1990.

[LBH15] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436, 2015.

[LBOM12] Yann A LeCun, Léon Bottou, Genevieve B Orr, and Klaus-Robert Müller. Efficient backprop. In *Neural networks: Tricks of the trade*, pages 9–48. Springer, 2012.

[LCC+18] Jiamin Liu, Jinzheng Cai, Karthik Chellamuthu, Mohammadhadi Bagheri, Le Lu, and Ronald M Summers. Cascaded coarse-to-fine convolutional neural networks for pericardial effusion localization and segmentation on CT scans. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 1092–1095, April 2018.

[Lev44] Kenneth Levenberg. A method for the solution of certain nonlinear problems in least squares. *Quarterly of Applied Mathematics*, 2(2):164–168, 1944.

[LHL+18] Baiying Lei, Shan Huang, Ran Li, Cheng Bian, Hang Li, Yi-Hong Chou, and Jie-Zhi Cheng. Segmentation of breast anatomy for automated whole breast ultrasound images with boundary regularized convolutional encoder–decoder network. *Neurocomputing*, 321:178 – 186, 2018.

[LIDLC+16] Manuel López-Ibáñez, Jérémie Dubois-Lacoste, Leslie Pérez Cáceres, Mauro Birattari, and Thomas Stützle. The irace package: Iterated racing for automatic algorithm configuration. *Operations Research Perspectives*, 3:43–58, 2016.

[LJ05] Stan Z Li and Anil K Jain. *Handbook of Face Recognition*. Springer, 2005.

[LKB+17] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, 2017.

[LLC+18] Jonathan Laserson, Christine Dan Lantsman, Michal Cohen-Sfady, Itamar Tamir, Eli Goz, Chen Brestel, Shir Bar, Maya Atar, and Eldad Elnekave. TextRay: Mining Clinical Reports to Gain a Broad Understanding of Chest X-Rays. In *MICCAI*, pages 553–561, 2018.

[LMAPH18] Stéphane Lathuilière, Pablo Mesejo, Xavier Alameda-Pineda, and Radu Horaud. A comprehensive analysis of deep regression. *arXiv preprint arXiv:1803.08450*, 2018.

[LMB+14] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft COCO: Common objects in context. In *European Conference on Computer Vision*, pages 740–755. Springer, 2014.

[Los13] Ilya Loshchilov. CMA-ES with restarts for solving CEC 2013 benchmark problems. In *2013 IEEE Congress on Evolutionary Computation (CEC)*, pages 369–376. IEEE, 2013.

[LSD15] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440, 2015.

[LST+16] Geert Litjens, Clara I Sánchez, Nadya Timofeeva, Meyke Hermsen, Iris Nagtegaal, Iringo Kovacs, Christina Hulsbergen-Van De Kaa, Peter Bult, Bram Van Ginneken, and Jeroen Van Der Laak. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Scientific Reports*, 6:26286, 2016.

[LZZ+19] Jian-Hao Luo, Hao Zhang, Hong-Yu Zhou, Chen-Wei Xie, Jianxin Wu, and Weiyao Lin. ThiNet: Pruning CNN Filters for a Thinner Net. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(10):2525–2538, 2019.

[Man98] Robert W Mann. Use of bone trabeculae to establish positive identification. *Forensic Science International*, 98(1-2):91–99, 1998.

[Mar63] Donald W Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441, 1963.

[Mar91] James G March. Exploration and exploitation in organizational learning. *Organization Science*, 2(1):71–87, 1991.

[MBB14] Xanthé Mallett, Teri Blythe, and Rachel Berry. *Advances in forensic human identification*. CRC Press, 2014.

[MCA+05] Zulkepli Majid, Albert K. Chong, Anuar Ahmad, Halim Setan, and Abdul Rani Samsudin. Photogrammetry and 3D laser scanning as spatial data capture techniques for a national craniofacial database. *The Photogrammetric Record*, 20(109):48–68, 2005.

[MdlHVdDLB10] Stella Martin-de-las Heras, Aurora Valenzuela, Juan de Dios Luna, and Manuel Bravo. The utility of dental patterns in forensic dentistry. *Forensic Science International*, 195(1-3):166–e1, 2010.

[Mer15] Domingo Mery. *X-ray Testing*, pages 1–33. Springer International Publishing, Cham, 2015.

[Mes14] Pablo Mesejo. *Automatic segmentation of anatomical structures using deformable models and bio-inspired/soft computing*. PhD thesis, Università degli Studi di Parma. Dipartimento di Ingegneria dell'Informazione, 2014.

[MHS17] Ajay Mittal, Rahul Hooda, and Sanjeev Sofat. Lung field segmentation in chest radiographs: a historical review, current status, and expectations from deep learning. *IET Image Processing*, 11(11):937–952, 2017.

[MHS18] Ajay Mittal, Rahul Hooda, and Sanjeev Sofat. LF-SegNet: A fully convolutional encoder–decoder network for segmenting lung fields from chest radiographs. *Wireless Personal Communications*, 101(1):511–529, 2018.

[MICC16] Pablo Mesejo, Oscar Ibáñez, Oscar Cordón, and Stefano Cagnoni. A survey on image segmentation using metaheuristic-based deformable models: state of the art and critical analysis. *Appl Soft Comput*, 44:1–29, 2016.

[MLH18] Daniel Molina, Antonio LaTorre, and Francisco Herrera. An insight into bio-inspired and evolutionary algorithms for global optimization: Review, analysis, and lessons learnt over a decade of competitions. *Cognitive Computation*, 10(4):517–544, 2018.

[MNA16] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *3DV*, pages 565–571. IEEE, 2016.

[Mod04] Jan Modersitzki. *Numerical methods for image registration*. Oxford University Press on Demand, 2004.

[MR14] Ashley B Maxwell and Ann H Ross. A radiographic study on the utility of cranial vault outlines for positive identifications. *Journal of Forensic Sciences*, 59(2):314–318, 2014.

[MTLP12] Primoz Markelj, Dejan Tomaževič, Bostjan Likar, and Franjo Pernuš. A review of 3D/2D registration methods for image-guided interventions. *Medical Image Analysis*, 16(3):642–661, 2012.

[MTPL08] Primoz Markelj, Dejan Tomazevic, Franjo Pernus, and Bostjan Likar. Robust gradient-based 3-D/2-D registration of CT and MR to x-ray images. *IEEE Transactions on Medical Imaging*, 27(12):1704–1714, 2008.

[MTWL16] Shun Miao, Ahmet Tuysuzoglu, Z Jane Wang, and Rui Liao. Real-time 6DoF pose recovery from x-ray images using library-based DRR and hybrid optimization. *International Journal of Computer Assisted Radiology and Surgery*, 11(6):1211–1220, 2016.

[MVIA18] Rubén Martos, Andrea Valsecchi, Oscar Ibáñez, and Inmaculada Alemán. Estimation of 2D to 3D dimensions and proportionality indices for facial examination. *Forensic Science International*, 287:142 – 152, 2018.

[MVM+18] Sai Phani Kumar Malladi, Bijju Kranthi Veduruparthi, Jayanta Mukherjee, Partha Pratim Das, Saswat Chakrabarti, and Indranil Mallick. Robust 3D registration of CBCT images aggregating multiple estimates through random sampling. *Pattern Recognition Letters*, 108:8–14, 2018.

[MWH77] William Martel, Jeffrey D. Wicks, and Robert C. Hendrix. The accuracy of radiologic identification of humans using skeletal landmarks: A contribution to forensic pathology. *Radiology*, 124(3):681–684, 1977.

[NAM05] Omaima Nomir and Mohamed Abdel-Mottaleb. A system for human identification from x-ray dental radiographs. *Pattern Recognition*, 38(8):1295–1305, 2005.

[NAM07] Omaima Nomir and Mohamed Abdel-Mottaleb. Human identification from dental x-ray images based on the shape and appearance of the teeth. *IEEE Transactions on Information Forensics and Security*, 2(2):188–197, 2007.

[NLM+18] Alexey A Novikov, Dimitrios Lenis, David Major, Jiří Hladůvka, Maria Wimmer, and Katja Bühler. Fully convolutional architectures for multiclass segmentation in chest radiographs. *IEEE Trans Med Imaging*, 37(8):1865–1876, 2018.

[NSGF16] Emily Niespodziewanski, Carl N Stephan, Pierre Guyomarc'h, and Todd W Fenton. Human identification via lateral patella radiographs: A validation study. *Journal of Forensic Sciences*, 61(1):134–140, 2016.

[NSH08] Ehsan Nadernejad, Sara Sharifzadeh, and Hamid Hassanpour. Edge detection techniques: Evaluations and comparisons. *Applied Mathematical Sciences*, 2(31):1507–1520, 2008.

[NTG+18] Silviya Nikolova, Diana Toneva, Ivan Georgiev, Angel Dandov, and Nikolai Lazarov. Morphometric analysis of the frontal sinus: application of industrial digital radiography and virtual endocast. *Journal of Forensic Radiology and Imaging*, 12:31–39, 2018.

[NTGL18] Silviya Nikolova, Diana Toneva, Ivan Georgiev, and Nikolai Lazarov. Digital radiomorphometric analysis of the frontal sinus and assessment of the relation between persistent metopic suture and frontal sinus development. *American Journal of Physical Anthropology*, 165(3):492–506, 2018.

[NW06] Jorge Nocedal and Stephen Wright. *Numerical optimization*. Springer Science & Business Media, 2006.

[OLC10] Yew-Soon Ong, Meng Hiot Lim, and Xianshun Chen. Memetic computation—past, present & future. *IEEE Computational Intelligence Magazine*, 5(2):24–31, 2010.

[OSP02] Sébastien Ourselin, Radu Stefanescu, and Xavier Pennec. Robust registration of multi-modal images: towards real-time clinical applications. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 140–147. Springer, 2002.

[OT14] Francisco PM Oliveira and Joao Manuel RS Tavares. Medical image registration: a review. *Computer Methods in Biomechanics and Biomedical Engineering*, 17(2):73–93, 2014.

[OW08] Chris O'Donnell and Noel Woodford. Post-mortem radiology—a new sub-speciality? *Clinical Radiology*, 63(11):1189–1194, 2008.

[Pea95] Karl Pearson. Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, 58:240–242, 1895.

[PHK04] Steve Pieper, Michael Halle, and Ron Kikinis. 3D slicer. In *IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2004.*, pages 632–635. IEEE, 2004.

[PIH+15] Azin Parsa, Norliza Ibrahim, Bassam Hassan, Paul van der Stelt, and Daniel Wismeijer. Bone quality evaluation at dental implant site using multislice CT, micro-CT, and cone beam CT. *Clinical Oral Implants Research*, 26(1):e1–e7, 2015.

[PMS12] Jahagirdar B Pramod, Anand Marya, and Vidhii Sharma. Role of forensic odontologist in post mortem person identification. *Dental Research Journal*, 9(5):522, 2012.

[PMV03] Josien PW Pluim, JB Antoine Maintz, and Max A Viergever. Mutual-information-based registration of medical images: a survey. *IEEE Transactions on Medical Imaging*, 22(8):986–1004, 2003.

[Pow09] Michael JD Powell. *The BOBYQA algorithm for bound constrained optimization without derivatives*. Cambridge NA Report NA2009/06, University of Cambridge, Cambridge, UK, 2009.

[Pre01] Iain A. Pretty. A look at forensic dentistry-part 1: The role of teeth in the determination of human identity. *British Dental Journal*, 190(7):359–366, 2001.

[PSB12] M Mary Synthuja Jain Preetha, L Padma Suresh, and M John Bosco. Image segmentation using seeded region growing. In *2012 International Conference on Computing, Electronics and Electrical Technologies (ICCEET)*, pages 576–583. IEEE, 2012.

[PTB11] Mark Page, Jane Taylor, and Matt Blenkin. Uniqueness in the forensic identification sciences–fact or fiction? *Forensic Science International*, 206(1):12–18, 2011.

[PVD+07] Matthias Pfaeffli, Peter Vock, Richard Dirnhofer, Marcel Braun, Stephan A Bolliger, and Michael J Thali. Post-mortem radiological CT identification based on classical ante-mortem x-ray examinations. *Forensic Science International*, 171(2-3):111–117, 2007.

[PWL+98] Graeme P Penney, Jürgen Weese, John A Little, Paul Desmedt, Derek LG Hill, and David J. Hawkes. A comparison of similarity measures for use in 2-D-3-D medical image registration. *IEEE Transactions on Medical Imaging*, 17(4):586–595, 1998.

[PXP00] Dzung L. Pham, Chenyang Xu, and Jerry L. Prince. Current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, 2(1):315–337, 2000.

[PZZ13] Bo Peng, Lei Zhang, and David Zhang. A survey of graph theoretical approaches to image segmentation. *Pattern Recognition*, 46(3):1020 – 1038, 2013.

[QFS+96] Gérald Quatrehomme, Pierre Fronty, Michel Sapanet, Gilles Grévin, Paul Bailet, and Amédée Ollier. Identification by frontal sinus pattern in forensic anthropology. *Forensic Science International*, 83(2):147–153, 1996.

[QL13] A Kai Qin and Xiaodong Li. Differential evolution on the CEC-2013 single-objective continuous optimization testbed. In *2013 IEEE Congress on Evolutionary Computation (CEC)*, pages 1099–1106, 2013.

[QSMG17] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3D classification and segmentation. *Proc. Computer Vision and Pattern Recognition (CVPR), IEEE*, 1(2):4, 2017.

[RA04] Tracy L Rogers and Travis T Allard. Expert testimony and positive identification of human remains through cranial suture patterns. *Journal of Forensic Sciences*, 49(2):1–5, 2004.

[RBC+16a] Thomas D Ruder, Cédric Brun, Angi M Christensen, Michael J Thali, Dominic Gascho, Wolf Schweitzer, and Gary M Hatch. Comparative radiologic identification with CT images of paranasal sinuses–development of a standardized approach. *Journal of Forensic Radiology and Imaging*, 7:1–9, 2016.

[RBC+16b] Thomas D. Ruder, Cédric Brun, Angi M. Christensen, Michael J. Thali, Dominic Gascho, Wolf Schweitzer, and Gary M. Hatch. Comparative radiologic identification with CTimages of paranasal sinuses – development of a standardized approach. *Journal of Forensic Radiology and Imaging*, 7:1 – 9, 2016.

[RE13] José L Rueda and Istvan Erlich. Hybrid mean-variance mapping optimization for solving the IEEE-CEC 2013 competition problems. In *2013 IEEE Congress on Evolutionary Computation (CEC)*, pages 1664–1671, 2013.

[RE15] José L Rueda and István Erlich. MVMO for bound constrained single-objective computationally expensive numerical optimization. In *2015 IEEE Congress on Evolutionary Computation (CEC)*, pages 1011–1017. IEEE, 2015.

[RE18] José L Rueda and István Erlich. Hybrid single parent-offspring MVMO for solving CEC2018 computationally expensive problems. In *2018 IEEE Congress on Evolutionary Computation (CEC)*, pages 1–8. IEEE, 2018.

[REM+08] Claire Robinson, Roos Eisma, Bruno Morgan, Amanda Jeffery, Eleanor AM Graham, Sue Black, and Guy N Rutty. Anthropological measurement of lower limb and foot bones using multi-detector computed tomography. *Journal of Forensic Sciences*, 53(6):1289–1295, 2008.

[RFB15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *MICCAI*, pages 234–241, 2015.

[RKG+12] Thomas D Ruder, Markus Kraehenbuehl, Walther F Gotsmy, Sandra Mathier, Lars C Ebert, Michael J Thali, and Gary M Hatch. Radiologic identification of disaster victims: a simple and reliable method using CT of the paranasal sinuses. *European Journal of Radiology*, 81(2):e132–e138, 2012.

[RKP05] Jani Rönkkönen, Saku Kukkonen, and Kenneth V Price. Real-parameter optimization with differential evolution. In *Congress on Evolutionary Computation*, pages 506–513, 2005.

[RLM16] Ann H Ross, Alicja K Lanfear, and Ashley B Maxwell. Establishing standards for side-by-side radiographic comparisons. *The American Journal of Forensic Medicine and Pathology*, 37(2):86–94, 2016.

[Roe95] Wilhelm Conrad Roentgen. On a new kind of rays. *Br J Radiol*, 4:32–153, 1895.

[RRM+05] Daniel B Russakoff, Torsten Rohlfing, Kensaku Mori, Daniel Rueckert, Anthony Ho, John R Adler, and Calvin R Maurer. Fast generation of digitally reconstructed radiographs using attenuation fields with application to 2D-3D image registration. *IEEE Transactions on Medical Imaging*, 24(11):1441–1454, 2005.

[RSFI15] Hrafnhildur L Runolfsdottir, Gunnar Sigurdsson, Leifur Franzson, and Olafur S Indridason. Gender comparison of factors associated with age-related differences in bone mineral density. *Archives of Osteoporosis*, 10(1):1–9, 2015.

[RTE16] José L Rueda Torres and Istvan Erlich. Solving the CEC2016 real-parameter single objective optimization problems through MVMO-PHM. In *2016 IEEE World Congress on Computational Intelligence*, pages 1–10, 2016.

[RWC+99] Philip J Robinson, Daniel Wilson, A Coral, Ad Murphy, and P Verow. Variation between experienced observers in the interpretation of accident and emergency radiographs. *The British Journal of Radiology*, 72(856):323–330, 1999.

[SAANM03] WG Shadeed, Dia I Abu-Al-Nadi, and Mohammad Jamil Mismar. Road traffic sign detection in color images. In *10th IEEE International Conference on Electronics, Circuits and Systems, 2003. ICECS 2003. Proceedings of the 2003*, volume 2, pages 890–893. IEEE, 2003.

[SAL+16] Caio Belém Rodrigues Barros Soares, Manuella Santos Carneiro Almeida, Patrícia de Medeiros Loureiro Lopes, Ricardo Villar Beltrão, Andrea dos Anjos Pontual, Flávia Maria de Moraes Ramos-Perez, José Naral Figueroa, and Maria Luiza dos Anjos Pontual. Human identification study by means of frontal sinus imaginological aspects. *Forensic Science International*, 262:183–189, 2016.

[SAT+14] Carl N Stephan, Brett Amidan, Harold Trease, Pierre Guyomarc'h, Trenton Pulsipher, and John E Byrd. Morphometric comparison of clavicle outlines from 3D bone scans and 2D chest

radiographs: a shortlisting tool to assist radiographic identification of human skeletons. *Journal of Forensic Sciences*, 59(2):306–313, 2014.

[SCD11] Jose Santamaría, Oscar Cordón, and Sergio Damas. A comparative study of state-of-the-art evolutionary image registration methods for 3D modeling. *Computer Vision and Image Understanding*, 115(9):1340–1354, 2011.

[SCDI09] Jose Santamaría, Oscar Cordón, Sergio Damas, and O Ibánez. Tackling the coplanarity problem in 3D camera calibration by means of fuzzy landmarks: a performance study in forensic craniofacial superimposition. In *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, pages 1686–1693. IEEE, 2009.

[Sch21] Arthur Schuller. Das rontgenogram der stirnhohle: ein hilfsmittel fur die identitatsbestimmung von schadeln. *Monatenschrift fur Ohrenheilkunde*, 55:1617–1620, 1921.

[Sch15] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117, 2015.

[Sco45] Charles C Scott. X-ray pictures as evidence. *Mich. L. Rev.*, 44:773, 1945.

[SDGTC12] José Santamaría, Sergio Damas, José M. García-Torres, and Oscar Cordón. Self-adaptive evolutionary image registration using differential evolution and artificial immune systems. *Pattern Recognition Letters*, 33(16):2065–2070, 2012.

[SDW+18] Carl N Stephan, Susan S D'Alonzo, Emily K Wilson, Pierre Guyomarc'h, Gregory E Berg, and John E Byrd. Skeletal identification by radiographic comparison of the cervicothoracic region on chest radiographs. In *New Perspectives in Forensic Human Skeletal Identification*, pages 277–292. Elsevier, 2018.

[SF18] Emily Streetman and Todd W. Fenton. Chapter 22 - comparative medical radiography: Practice and validation. In Krista E. Latham, Eric J. Bartelink, and Michael Finnegan, editors, *New Perspectives in Forensic Human Skeletal Identification*, pages 251 – 264. Academic Press, 2018.

[SFB+15] Erik Smistad, Thomas L. Falch, Mohammadmehdi Bozorgi, Anne C. Elster, and Frank Lindseth. Medical image segmentation on GPUs – a comprehensive review. *Med Image Anal*, 20(1):1 – 18, 2015.

[SG14] Carl N. Stephan and Pierre Guyomarc'h. Quantification of perspective-induced shape change of clavicles at radiography and 3D scanning to assist human identification. *Journal of Forensic Sciences*, 59(2):447–453, 2014.

[SGG+14] Yeqin Shao, Yaozong Gao, Yanrong Guo, Yonghong Shi, Xin Yang, and Dinggang Shen. Hierarchical lung field segmentation with joint shape and appearance sparse learning. *IEEE Trans Med Imaging*, 33(9):1761–1780, 2014.

[SGW+12] Jakob Spoerk, Christelle Gendrin, Christoph Weber, Michael Figl, Supriyanto Ardjo Pawiro, Hugo Furtado, Daniella Fabri, Christoph Bloch, Helmar Bergmann, Eduard Gröller, et al. High-performance GPU-based rendering for real-time, rigid 2D/3D-image registration and motion prediction in radiation oncology. *Zeitschrift für Medizinische Physik*, 22(1):13–20, 2012.

[SHK+14] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res*, 15(1):1929–1958, 2014.

[SHNY17] Norihiro Shinkawa, Toshinori Hirai, Ryuichi Nishii, and Nobuhiro Yukawa. Usefulness of 2D fusion of postmortem CT and antemortem chest radiography studies for human identification. *Japanese Journal of Radiology*, 35(6):303–309, 2017.

[Sin51] Arthur C Singleton. The roentgenological identification of victims of the "noronic" disaster. *The American Journal of Roentgenology and Radium Therapy*, 66(3):375—384, 1951.

[SKI+00] Junji Shiraishi, Shigehiko Katsuragawa, Junpei Ikezoe, Tsuneo Matsumoto, Takeshi Kobayashi, Ken-ichi Komatsu, Mitate Matsui, Hiroshi Fujita, Yoshie Kodera, and Kunio Doi. Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. *AJR Am J Roentgenol*, 174(1):71–74, 2000.

[SL10] Ruslan Salakhutdinov and Hugo Larochelle. Efficient learning of deep Boltzmann machines. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 693–700, 2010.

[SLJ+15] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9, 2015.

[Smo86] Paul Smolensky. Information processing in dynamical systems: Foundations of harmony theory. Technical report, Colorado Univ at Boulder Dept of Computer Science, 1986.

[Sol11] Angela Soler. Positive identification through comparative panoramic radiography of the maxillary sinuses: a validation

study. In *Proceedings of the 63rd Annual Meeting of the American Academy of Forensic Sciences*, pages 23–26, 2011.

[Son11] Byung Cheol Song. A fast normalized cross correlation-based block matching algorithm using multilevel cauchy-schwartz inequality. *ETRI Journal*, 33(3):401–406, 2011.

[Sør48] Thorvald Sørensen. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on danish commons. *Kongelige Danske Videnskabernes Selskab*, 5:1–34, 1948.

[Sör15] Kenneth Sörensen. Metaheuristics—the metaphor exposed. *International Transactions in Operational Research*, 22(1):3–18, 2015.

[SP97] Rainer Storn and Kenneth Price. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11(4):341–359, 1997.

[SSMBV17] Sancho Salcedo-Sanz, Jesús Muñoz-Bulnes, and Mark JA Vermeij. New coral reefs-based approaches for the model type selection problem: a novel method to predict a nation's future energy demand. *International Journal of Bio-inspired Computation*, 10(3):145–158, 2017.

[SVZ13] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*, 2013.

[SWCT11] Carl N Stephan, Allysha P Winburn, Alexander F Christensen, and Andrew J Tyrrell. Skeletal identification by radiographic comparison: Blind tests of a morphoscopic method using antemortem chest radiographs. *Journal of Forensic Sciences*, 56(2):320–332, 2011.

[SZ14] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[Sze10] Richard Szeliski. *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010.

[TB06] Tim Thompson and Sue Black. *Forensic human identification: An introduction*. CRC press, 2006.

[TBD03] Michael J. Thali, Marcel Braun, and Richard Dirnhofer. Optical 3D surface digitizing in forensic medicine: 3D documentation of skin and bone injuries. *Forensic Science International*, 137(2):203–208, 2003.

[TBV02] Michael J Thali, BG Brogdon, and Mark D Viner. *Forensic radiology*. CRC Press, 2002.

[TF13] Ryoji Tanabe and Alex S Fukunaga. Success-history based parameter adaptation for differential evolution. In *2013 IEEE Congress on Evolutionary Computation (CEC)*, pages 71–78, 2013.

[TF14] Ryoji Tanabe and Alex S Fukunaga. Improving the search performance of shade using linear population size reduction. In *2014 IEEE Congress on Evolutionary Computation (CEC)*, pages 1658–1665, 2014.

[The17] The CGAL Project. *CGAL User and Reference Manual*. CGAL Editorial Board, 4.9.1 edition, 2017.

[TI11] Khang Siang Tan and Nor Ashidi Mat Isa. Color image segmentation using histogram thresholding – fuzzy C-means hybrid approach. *Pattern Recognition*, 44(1):1 – 15, 2011.

[TKWK08] Zbislaw Tabor, Dariusz Karpisz, Leszek Wojnar, and Piotr Kowalski. An automatic recognition of the frontal sinus in x-ray images of skull. *IEEE Transactions on Biomedical Engineering*, 56(2):361–368, 2008.

[TLSP03] Dejan Tomazevic, Bostjan Likar, Tomaz Slivnik, and Franjo Pernus. 3-D/2-D registration of CT and MR to x-ray images. *IEEE Transactions on Medical Imaging*, 22(11):1407–1416, Nov 2003.

[TOA+07] Ertugrul Tatlisumak, Gulgun Yilmaz Ovali, Asim Aslan, Mahmut Asirdizer, Yildiray Zeyfeoglu, and Serdar Tarhan. Identification of unknown bodies by using CTimages of frontal sinus. *Forensic Science International*, 166(1):42–48, 2007.

[TSNF73] Jun-Ichiro Toriwaki, Yasuhito Suenaga, Toshio Negoro, and Teruo Fukumura. Pattern recognition of chest X-ray images. *Comput Vision Graph*, 2(3):252 – 271, 1973.

[TVCC05] Marco Tomassini, Leonardo Vanneschi, Philippe Collard, and Manuel Clergue. A study of fitness distance correlation as a difficulty measure in genetic programming. *Evolutionary Computation*, 13(2):213–239, 2005.

[TYS+03] Michael J Thali, Kathrin Yen, Wolf Schweitzer, Peter Vock, Chris Boesch, Christoph Ozdoba, Gerhard Schroth, Michael Ith, Martin Sonnenschein, Tanja Doernhoefer, et al. Virtopsy, a new imaging horizon in forensic pathology: virtual autopsy by postmortem multislice computed tomography (MSCT) and magnetic resonance imaging (MRI)-a feasibility study. *Journal of Forensic Sciences*, 48(2):386–403, 2003.

[UGK+05] Selma Uysal, Dilek Gokharman, Mahmut Kacar, Isıl Tuncbilek, and Ugur Kosar. Estimation of sex by 3D CT measurements of the foramen magnum. *Journal of Forensic Sciences*, 50(6):JFS2005058–5, 2005.

[UVL16] Dmitry Ulyanov, Andrea Vedaldi, and Victor S. Lempitsky. Instance normalization: The missing ingredient for fast stylization. *CoRR*, abs/1607.08022, 2016.

[VBDC18] Andrea Valsecchi, Enrique Bermejo, Sergio Damas, and Oscar Cordón. Metaheuristics for medical image registration. *Handbook of Heuristics*, pages 1079–1101, 2018.

[VDDP18] Athanasios Voulodimos, Nikolaos Doulamis, Anastasios Doulamis, and Eftychios Protopapadakis. Deep learning for computer vision: A brief review. *Computational Intelligence and Neuroscience*, 2018.

[vdKPT+05] Everine B van de Kraats, Graeme P Penney, Dejan Tomazevic, Theo Van Walsum, and Wiro J Niessen. Standardized evaluation methodology for 2-D-3-D registration. *IEEE Transactions on Medical Imaging*, 24(9):1177–1189, 2005.

[VGRV01] Bram Van Ginneken, BM Ter Haar Romeny, and Max A Viergever. Computer-aided diagnosis in chest radiography: a survey. *IEEE Trans Med Imaging*, 20(12):1228–1241, 2001.

[VGSL06] Bram Van Ginneken, Mikkel B Stegmann, and Marco Loog. Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database. *Med Image Anal*, 10(1):19–40, 2006.

[VH01] Remco C. Veltkamp and Michiel Hagedoorn. Principles of visual information retrieval. chapter State of the Art in Shape Matching, pages 87–119. Springer-Verlag, London, UK, 2001.

[Vic05] A Leah Vickers. Daubert, critique and interpretation: What empirical studies tell us about the application of daubert. *USFL Rev.*, 40:109, 2005.

[Vit] Vitrea® advanced visualization. `https://www.vitalimages.com/enterprise-imaging-solution/advanced-visualization-2/`. Accessed: 2019-07-07.

[VR17] Mark D Viner and John Robson. Post-mortem forensic dental radiography-a review of current techniques and future developments. *Journal of Forensic Radiology and Imaging*, 8:22–37, 2017.

[Wan17] Chunliang Wang. Segmentation of multiple structures in chest radiographs using multi-task fully convolutional networks. In Puneet Sharma and Filippo Maria Bianchi, editors, *SCIA*, pages 282–289, 2017.

[WB09] Zhou Wang and Alan C Bovik. Mean squared error: Love it or leave it? a new look at signal fidelity measures. *IEEE Signal Processing Magazine*, 26(1):98–117, 2009.

[WK07] Daniel J Withey and Zoltan J Koles. Medical image segmentation: Methods and software. In *NFSI-ICFBI*, pages 140–143, 2007.

[WR10] Lelia Watamaniuk and Tracy Rogers. Positive personal identification of human remains based on thoracic vertebralmargin morphology. *Journal of Forensic Sciences*, 55(5):1162–1170, 2010.

[WS77] Harry Wechsler and Jack Sklansky. Finding the rib cage in chest radiographs. *Pattern Recognition*, 9(1):21 – 30, 1977.

[WVBWK15] Zhizheng Wu, Cassia Valentini-Botinhao, Oliver Watts, and Simon King. Deep neural networks employing multi-task learning and stacked bottleneck features for speech synthesis. In *ICASSP*, pages 4460–4464, 2015.

[XSM+17] Junfeng Xiong, Yeqin Shao, Jingchen Ma, Yacheng Ren, Qian Wang, and Jun Zhao. Lung field segmentation using weighted sparse shape composition with robust initialization. *Medical Physics*, 44(11):5916–5929, 2017.

[YG10] Xinjie Yu and Mitsuo Gen. *Introduction to evolutionary algorithms*. Springer Science & Business Media, 2010.

[YK15] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015.

[YLL+18] Wei Yang, Yunbi Liu, Liyan Lin, Zhaoqiang Yun, Zhentai Lu, Qianjin Feng, and Wufan Chen. Lung field segmentation in chest radiographs from boundary maps by a structured edge detector. *IEEE J Biomed Health Inform*, 22(3):842–851, 2018.

[YMSS87] Mineo Yoshino, Sachio Miyasaka, Hajime Sato, and Sueshige Seta. Classification system of frontal sinus patterns by radiography. its application to identification of unknown skeletal remains. *Forensic Science International*, 34(4):289–299, 1987.

[ZLLT14] Zhanpeng Zhang, Ping Luo, Chen Change Loy, and Xiaoou Tang. Facial landmark detection by deep multi-task learning. In *ECCV*, pages 94–108, 2014.

[ZOZF16] Zhiyuan Zhang, Sim Heng Ong, Xin Zhong, and Kelvin W.C. Foong. Efficient 3D dental identification via signed feature histogram and learning keypoint detection. *Pattern Recognition*, 60:189–204, 2016.

[ZPW+09] Ying Zhu, Simone Prummer, Peng Wang, Terrence Chen, Dorin Comaniciu, and Martin Ostermeier. Dynamic layer separation for coronary DSA and enhancement in fluoroscopic sequences. In *MICCAI*, pages 877–884, 2009.

[ZYF+11] Xin Zhong, Deping Yu, Kelvin WC Foong, Terence Sim, Yoke San Wong, and Ho-Lun Cheng. Towards automated pose invariant

3D dental biometrics. In *2011 International Joint Conference on Biometrics (IJCB)*, pages 1–7. IEEE, 2011.

[ZYQ19] Zhi-Hua Zhou, Yang Yu, and Chao Qian. *Evolutionary Learning: Advances in Theories and Algorithms*. Springer, 2019.

[ZYW+13] Xin Zhong, Deping Yu, Yoke San Wong, Terence Sim, Wen Feng Lu, Kelvin Weng Chiong Foong, and Ho-Lun Cheng. 3D dental biometrics: Alignment and matching of dental casts for human identification. *Computers in Industry*, 64(9):1355–1370, 2013. Special Issue: 3D Imaging in Industry.