Carpal Bone Analysis using Geometric and Deep Learning Models

Mahsa Rezaei

A Thesis
in
The Concordia Institute
for
Information Systems Engineering

Presented in Partial Fulfillment of the Requirements
for the Degree of Master of Applied Science (Quality Systems Engineering) at
Concordia University
Montreal, QC, Canada

November 2020

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Mahsa Rezaei

Entitled: **Carpal Bone Analysis using Geometric and Deep Learning Models.**

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science (Quality Systems Engineering)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____ Chair
Dr. A. Youssef

_____ Examiner
Dr. J. Bentahar

_____ Examiner
Dr. A. Youssef

_____ Supervisor
Dr. A. Ben Hamza

Approved by _____
Dr. A. Ben Hamza, Director
Concordia Institute for Information Systems Engineering

_____
Dr. M. Debbabi, Interim Dean
Faculty of Engineering and Computer Science

Date _____

Abstract

Carpal bone analysis using geometric and deep learning models

Mahsa Rezaei, M.A. Sc.
Concordia University, 2020

The recent trend for analyzing 3D shapes in medical application has arisen new challenges for a vast amount of research activities. Quantitative shape comparison is a fundamental problem in computer vision, geometry processing and medical imaging. This thesis is motivated by the availability of carpal bone shape dataset to develop efficient techniques for diagnosis of a variety of wrist diseases and examine human skeletal.

This study is conducted in two sections. First, we propose a spectral graph wavelet approach for shape analysis of carpal bones of the human wrist. More precisely, we employ spectral graph wavelets to represent the cortical surface of a carpal bone via the spectral geometric analysis of the Laplace-Beltrami operator in the discrete domain. We then propose global spectral graph wavelet (GSGW) descriptor that is isometric invariant, efficient to compute and combines the advantages of both low-pass and band-pass filters. Subsequently, we perform experiments on shapes of the carpal bones of ten women and ten men from a publicly-available database of wrist bones. Using one-way multivariate analysis of variance (MANOVA) and permutation testing, we show through extensive experiments that the proposed GSGW framework gives a much better performance compared to the global point signature (GPS) embedding approach for comparing shapes of the carpal bones across populations.

In the second section, we evaluate bone age to assess children's biological maturity and to diagnose any growth disorders in children. Manual bone age assessment (BAA) methods are time-consuming and prone to observer variability by even expert radiologists. These drawbacks motivate us for proposing an accurate computerized BAA method based on human wrist bones X-ray images. We also investigate automated BAA methods using state-of-the-art deep learning models that estimate the bone age more accurate than the manual methods by eliminating human observation

variations. The presented approaches provide faster assessment process and cost reduction in the hospitals/clinics. The accuracy of our experiments is evaluated using mean absolute error (MAE), and the results demonstrate that exploiting InceptionResNet-V2 model in our architecture achieves higher performance compared to the other used pre-trained models.

Acknowledgements

I would like to express my gratitude and appreciation to my supervisor Prof. A. Ben Hamza, you have been a tremendous mentor for me. I would like to thank you for encouraging my research and for allowing me to grow as a data analyst. Your advice on both my research as well as my career have been invaluable.

I am forever grateful to my husband Dr. Majid Masoumi, and also dedicate my thesis to him for his unconditional love, constant support and encouragement.

Table of Contents

List of Tables

List of Figures

List of Acronyms

BAA Bone Age Assessment

GP Greulich and Pyle

TW Tanner and Whitehouse

GR Gilsanz and Ratibin

RUS Radius, Ulna, and Short Bones

MAE Mean Absolute Error

AdaGrad Adaptive Gradient Algorithm

RMSProp Root Mean Square Propagation

MSE Mean Squared Error

LBO Laplace-Beltrami Operator

HKS Heat Kernel Signature

WKS Wave Kernel Signature

BoF Bag-of-Features

SS-BoF Spatially Sensitive Bags of Features

GA-BoF Geodesic-Aware Bags of Features

SGWC Spectral Graph Wavelet Codes

MANOVA One-Way Multivariate Analysis of Variance

SGWC-BoF Spectral Graph Wavelet Codes Bag-of-Features

GSGW Global Spectral Graph Wavelet

GPS Global Point Signature

NMSE Normalised Mean Square Error

CHAPTER 1

Introduction

In this chapter, we present the motivation behind our work, the problem statement, objectives of the study and followed by literature reviews.

1.1 Framework and Motivation

With the advent of X-ray technology, X-ray images facilitate the estimation of bone age for a child's hand, where the extent of growth and development of hand bones have been visualized by wrist radiographs. An X-ray image indicates the best visualization of a pattern of ossification in a child's hand and wrist bones until the elongation of the bone is completed. Thus, the bone age is derived by comparing the maturation level of hand and wrist bones with normal age levels and can be fairly predicted in the period of time. Recently, bone age assessment (BAA) has gained remarkable attention from academia and medicine. Various manual techniques have been proposed to compute bone age score from the radiographs, such as the Greulich Pyle (GP) [5, 6] and Tanner Whitehouse (TW) [7].

Due to the significant variation of calculated bone age and time-consuming procedure of these techniques, the results are not acceptable by senior radiologists [8]. To address the limitations, automated methods based on deep learning techniques have been proposed for BAA.

In this thesis, we also analyze 3D carpal bones among men and women using a discriminative shape descriptor and then explore the application of deep learning to predict children's wrist bones ages.

1.2 Problem Statement

Shape and regression analyses are critical challenges in computer vision, geometry processing and medical imaging. In this thesis, a discriminative shape descriptor is proposed for quantitative shape comparison of the human wrist between genders. In addition, for developing automated BAA methods, we introduce regression analysis in conjunction with deep learning models to evaluate biological maturity of children from their chronological ages (real ages) on the RSNA dataset, including X-ray images of their left hand wrists.

1.2.1 Shape analysis

Global spectral graph wavelet (GSGW) [9, 10] approach is a global descriptor, proposed for shape analysis on carpal bones of the human wrist. By applying GSGW descriptor, the spectral geometric analysis of the Laplace-Beltrami operator in the discrete domain is exploited to represent the cortical surface of a carpal bone. The proposed descriptor takes advantages of both low-pass and band-pass filters to be more discriminated, which is also computationally efficient and isometric invariant. In order to assess the analysis of carpal bone shapes of both genders, two standard metrics are used, i.e. one-way multivariate analysis of variance (MANOVA) and permutation testing.

1.2.2 Regression analysis

The goal of regression analysis is to efficiently predict age of hand bones as a continuous variable. We split RSNA dataset into two disjoint subsets, training set for learning process, and test set for testing. In a bid to improve the performance, various deep learning pre-trained models are investigated to extract high-level features of the radiographs, followed by applying a regression model to predict age of children's bones. Finally, the assessment of a regression model is performed on the test set by comparing the predicated bone age with the real age using mean absolute error (MAE) metric.

1.3 Objectives

This research aims to propose a GSGW framework that represents cortical surface of a carpal bone using a global shape descriptor, in which the descriptor is defined as an area-weighted sum of all local spectral graph wavelet signatures at each surface point. In fact, The global descriptor takes into account the spectral geometric analysis on a shape, generated by the Laplace-Beltrami operator in the discrete domain. Also, considering the wavelet transform enables us to perform

local analysis and a multi-resolution analysis as well. The objective of introducing GSGW is to understand and analyze variations of carpal bones shapes to better diagnose wrist pathologies compared to the existing approaches.

Moreover, in this thesis, we investigate an automated algorithm for BAA based on publicly available RSNA dataset using deep learning pre-trained models to develop the regression model's performance. In fact, the deep learning pre-trained models provide high-level features learned from different large datasets, that is positively correlated with improvement of predicted bone age as an output of our model.

1.4 Literature Review

The human wrist (or carpus) is a complex joint that connects the hand to the forearm and is composed of eight carpal bones arranged in two rows of four bones each. Each carpal bone has a unique shape and plays a significant functional role in wrist stability and mobility. Changes in the shape of a carpal bone may be a sign of wrist injuries or disorders, such as arthritis and carpal tunnel syndrome, and hence understanding and analyzing variations of bone shapes is essential to the diagnosis of wrist pathologies. Quantitative shape analysis of carpal bones not only helps identify unique phenotypes across populations, but also allows for the detection of abnormal wrist pathologies as well as the investigation of biomechanical properties of the wrist joints [11]. This analysis has become possible thanks in large part to the availability of databases of normal and abnormal pathologies [12].

Quantitative analysis of variations in shape has been successfully applied to other medical imaging applications [13, 14], including the diagnosis and understanding of cardiovascular diseases. Xu *et al.* [13] exploited kinematic features from a 3D shape throughout the cardiac cycle to analyze the function of the human right ventricle, which is anatomically consistent and provides direct comparison across populations of individuals. Wu *et al.* [14] presented a computational approach for statistical analysis of sets of watertight surfaces produced from untagged medical images in two main steps, namely surface parametrization and proper orthogonal decomposition.

The recent surge of interest in the spectral analysis of the Laplace-Beltrami operator (LBO) has resulted in a considerable number of spectral shape signatures that have been successfully applied to a broad range of areas, including shape analysis [15–21], multimedia protection [22], and medical imaging [23]. The diversified nature of these applications is a powerful testimony of the practical usage of spectral shapes signatures, which are usually defined as feature vectors representing local and/or global characteristics of a shape and may be broadly classified into two main categories: local and global descriptors. Local descriptors (also called point signatures) are

defined on each point of the shape and often represent the local structure of the shape around that point, while global descriptors are usually defined on the entire shape capturing its global structure.

Most point signatures may easily be aggregated to form global descriptors by integrating over the entire shape. One of the simplest global spectral shape signatures is Shape-DNA [15], which is an isometry-invariant global descriptor defined as a truncated sequence of the LBO eigenvalues arranged in increasing order of magnitude. Gao *et al.* [24] developed a variant of Shape-DNA, referred to as compact Shape-DNA (cShape-DNA), which is an isometry-invariant signature resulting from applying the discrete Fourier transform to the area-normalized eigenvalues of the LBO. Chaudhari *et al.* [23] proposed a global point signature (GPS) embedding for quantifying the overall bone shape, and it is obtained by setting the LBO eigenfunctions in the GPS signature [16] to unity. More precisely, the GPS embedding is defined as a truncated sequence of inverse square roots of the area-normalized eigenvalues of the LBO. In addition to providing an efficient representation for comparing shapes of the carpal bones across populations, the GPS embedding has several desirable properties for shape analysis of carpal bones, including invariance to Euclidean and isometric transformations.

While the GPS signature [16] is invariant under isometric deformations of the shape, it suffers from the problem of eigenfunctions' switching whenever the associated eigenvalues are close to each other. This problem was lately well handled by the heat kernel signature (HKS) [25], which is a temporal descriptor defined as an exponentially-weighted combination of the LBO eigenfunctions. HKS is a local shape descriptor that has a number of desirable properties, including robustness to small perturbations of the shape, efficiency and invariance to isometric transformations. From the graph Fourier perspective, it can be seen that HKS is highly dominated by information from low frequencies, which correspond to macroscopic properties of a shape. To give rise to substantially more accurate matching than HKS, the wave kernel signature (WKS) [26] was proposed as an alternative in an effort to allow access to high-frequency information.

Another alternative approach to transforming local descriptors into global representations is the bag-of-features (BoF) paradigm [17]. The BoF model represents each object in the dataset as a collection of unordered feature descriptors extracted from local areas of the shape, just as words are local features of a document. A baseline BoF approach quantizes each local descriptor to its nearest cluster center using K-means clustering and then encodes each shape as a histogram over cluster centers by counting the number of assignments per cluster. These cluster centers form a visual vocabulary or codebook whose elements are often referred to as visual words or codewords. Although the BoF paradigm has been shown to provide significant levels of performance, it does not, however, take into consideration the spatial relations between features, which may have an

adverse effect not only on its descriptive ability but also on its discriminative power. To account for the spatial relations between features, Bronstein *et al.* introduced a generalization of a bag of features, called spatially sensitive bags of features (SS-BoF) [17]. The SS-BoF is a global descriptor defined in terms of mid-level features and the heat kernel, and can be represented by a square matrix whose elements represent the frequency of nearby codewords in the vocabulary. In the same spirit, Bu *et al.* [27] recently proposed the geodesic-aware bags of features (GA-BoF) for 3D shape classification by replacing the heat kernel in SS-BoF with a geodesic exponential kernel.

More recently, vertex-frequency analysis on graphs via the Fourier transform in the spectral graph-theoretic setting has received a great deal of interest [28, 29]. While the Fourier transform has been widely used as a reliable tool in signal processing applications for many years, wavelet analysis has been shown to provide some key advantages over the Fourier transform, making it an interesting alternative for many applications. In particular, unlike the Fourier transform, wavelet analysis is able to perform local analysis and also makes it possible to perform a multiresolution analysis. Classical wavelets are constructed by translating and scaling a mother wavelet, which is used to generate a set of functions through the scaling and translation operations. The wavelet transform coefficients are then obtained by taking the inner product of the input function with the translated and scaled waveforms.

The application of wavelets to graphs (or triangle meshes in geometry processing) is, however, problematic and not straightforward due in part to the fact that it is unclear how to apply the scaling operation on a signal (or function) defined on the mesh vertices. To tackle this problem, Coifman *et al.* [30] introduced the diffusion wavelets, which generalize the classical wavelets by allowing for multiscale analysis on graphs. The construction of diffusion wavelets interacts with the underlying graph through repeated applications of a diffusion operator, which induces a scaling process. Hammond *et al.* [31] showed that the wavelet transform can be performed in the graph Fourier domain, and proposed a spectral graph wavelet transform that is defined in terms of the eigensystem of the graph Laplacian matrix.

Recently, a spectral graph wavelet signature (SGWS) was introduced in [32], and it has shown superior performance over HKS and WKS in 3D shape analysis. SGWS is a multiresolution local descriptor that is not only isometric invariant, but also compact, easy to compute and combines the advantages of both band-pass and low-pass filters.

Bone age is described as a measurement of skeletal maturity using radiography of the ossification center. Despite a large amount of scientific research on BAA, a lack of consensus on the accuracy of bone age prediction techniques exists for clinical environments [33]. In fact, in the practical application of BAA methods in clinical environments and courts of justice, the most accurate

results are essential. These methods are reduced the subjective evaluation as the primary reason for achieving the inaccurate result by eliminating role of human observer [34].

The Greulich and Pyle (GP) atlas [35] was introduced based on the radiographic atlas of skeletal development of the hand and wrist. The method is one of the most frequent used atlases for measurement of bone age by radiologists in the West and Pakistan [36]. The atlas includes the reference images of the hand and the left wrist for both genders from birth until the age of 18 for women and the age of 19 for men, in which each image provides an explanation of gradual age-related changes in the bone structure. Bone age is calculated by comparing the subject's left hand wrist X-ray with the closest matching reference X-rays in the atlas, which are standardized for ages of the atlas [36]. The Tanner-Whitehouse (TW) [7] was proposed based on maturity level instead of age for 20 selected regions of interest (ROI) in specific hand and wrist bones for each range of age. The specific stage labeled $(A, B, C, D, \ldots, I)$ are determined for the development level of each ROI, which represents a numerical score for each stage of development for individual bone. Then, a total maturity score is computed by summing up all the ROI scores and correlated with bone age separately for both genders. This method benefits from higher accuracy but it is relatively time-consuming and complex compared to GP method. A new version of atlas called Gilsanz and Ratibin (GR) atlas [37] was later developed to idealized and artificial images specific to age and sex standards of skeletal maturity, produced by analyses of the size, shape, morphology and density of ossification centers for hand X-rays of healthy children to generate images with typical developmental features in each ossification center. The new GR images are more accurate from the view of radiologists and of better quality compared to those of the old GP atlas [38]. The new GR atlas standards also have a regular interval of six months between the ages of 2 and 6 and an annual interval between the ages of 7 and 17. Regardless of nearly the same result in determining bone age in GP and GR methods, the GR contains more outliers and can be replaced with the older GP atlas [37].

The majority of automatic BAA techniques are based on X-ray images of the left hand wrist to indicate skeletal maturity. The analysis of wrist bones age is a challenge even for radiologists due to rapid changes in the shapes of wrist bones over time, and some bones also overlap with maturation [39]. Most computer-based approaches utilize TW because of its ability to score for skeletal maturity, and to evaluate the radiograph of a known hand [40].

The importance of the automation methods such as hand segmentation for bone age estimation, is recognized by researchers [41, 42]. Computerized approaches in BAA are projected to reduce the BAA cost by reducing the time spent on BAA process by radiologists [43]. The first semi-automated BAA system called HANDX [44] was proposed to enable automatic segmentation of

wrist bones on X-ray images. The system includes three parts: pre-processing, segmentation and measurement, in which decreases the observer variability. The output of system is helpful for detecting skeletal growth abnormalities in children. The system process is as follows: the X-rays are normalized in the first phase to feed in the second phase. The segmentation phase identifies the individual bones in the hand and isolates the edges of the bone. Finally, quantitative parameters are obtained. This method lacks sufficient accuracy when the hand image has been fused and has never been assessed for a large scale hand image.

The PROI-based technique [45] was introduced based on the region including the phalanges and epiphyses. In the first phase, a horizontal line is scanned to evaluate the bone age before the soft tissue between the thumb and fist finger is detected and the lower boundary of the PROI is found. In the following phase, the upper boundary with a horizontal line is scanned at the edge of the third finger. Then, the segmentation phase begins when the upper, lower, left and right boundaries of the PROI have been identified. For bone segmentation, a gradient image is used, and the output threshold for determining bone edges is based on empirical analysis. At the end of the region, the density of the value of pixels is higher than the center section. This technique is measured for the boundary between the third distal, middle, and proximal phalanges using the standard table designed by the Garn group [46], in which phalangeal length is converted into skeletal age. The system was assessed with 50 computer radiographs (CR) of patients and compared with a radiologist's results, in which the mean difference from the assessment was 0.02 mm with a measurement error of 0.08 mm [47].

Phalanges, Epiphyses, and Carpals is a computer-based system for BAA that was presented based on the third digit by the National Tsing-Hwa University, in which the left hand was extracted from the X-ray image [48]. In the pre-processing stage, thresholding methods and heuristic searches were used to rotate the X-ray. The approach operates the segmented phalangeal region of interest (PROI) with Gabor filters for smoothing, canny edge detector and local variance method for finding the edge, and refinement. The PROI segmentation takes advantage of low error rate in BAA, however, poor image processing techniques are considered as its drawback.

Another computerized system for BAA is the neural network classifiers (NNC), developed by Liu et al [49], which is based on two geometric features of carpal bones and the RUS (radius, ulna, and short bones). For the segmentation of the bones, a large dataset of samples along with particle swarm algorithm was exploited. Two individual classifiers are then applied in this method on the RUS bone and the carpal bone for samples under nine years old. The method has the advantage of a small standard deviation of the differences between the system and observer and reduces the carpal bone-based system's variability in comparison with previous systems. However, it suffers

from high image processing loading than other methods.

CHAPTER 2

Global spectral graph wavelet signature for surface analysis of carpal bones

2.1 Introduction

The human wrist is composed of eight carpal bones which each bone has a unique shape. The bones shapes play a key role in diagnosis of any wrist abnormality, such as wrist instability, osteoarthritis [50]. Also, further applications of analyzing bone shape may be in development of patient-specific prostheses [51] examination of skeletal growth and maturity pattern [52, 53], reconstructive [54] as well as researches on carpal bones of different species [55]. Therefore, shape analysis of carpal bones can be provided by detailed investigation of visual representation (features) of wrist bones shapes and led to advancement in orthopedic science.

In recent years, utility of spectral analysis of the Laplace-Beltrami operator (LBO) has been considerably developed to build a number of spectral shape signatures or descriptors. In fact, LBO represents a shape via a compact and concise descriptor and can be applied to facilitate the tasks in various fields, such as shape analysis [15–17, 19–21, 56], multimedia protection [22], medical imaging [23, 57], and food processing industry [58].

The shape representation are categorized into local and global descriptors. As a matter of fact, local descriptors are usually defined on each point of the shape, which the category of local descriptors include as following: the global point signature (GPS) [16], heat kernel signature (HKS) [25], wave kernel signature (WKS) [26], and spectral graph wavelet signature (SGWS) [19]. On the other hand, global descriptors are composed of point signatures by integrating over the entire

shape. For example, Shape-DNA is defined as a truncated sequence of the LBO eigenvalues arranged in increasing order of magnitude. A new version of the GPS descriptor was introduced by Chaudhari et al. [23] that set the LBO eigenfunctions to unity. A developed version of Shape-DNA proposed by Gao et al. [24], referred to as compact Shape-DNA (cShape-DNA), in which an isometry-invariant signature was constructed by applying the discrete Fourier transform to the area-normalized eigenvalues of the LBO.
More recently, Fourier transform has been a powerful tool in signal processing. Unlike, the advantages of the Fourier transform, wavelet analysis has been proven the superiority than the Fourier transform due to its power of localization and multi-resolution analysis. Recently, a spectral graph wavelet signature [32] (SGWS) has been shown to provide higher performance as a multi-resolution local descriptor than HKS and WKS on 3D shape Analysis. The descriptor is isometric invariant and computationally efficient and also contains the information given by both band-pass and low-pass filters.

In this chapter, we introduce a global spectral graph wavelet (GSGW) framework that represents the shape of the cortical surface of a carpal bone by a global shape descriptor defined as an area-weighted sum of all local spectral graph wavelet signatures at each surface point. The resulting global descriptor is not only isometric invariant, but also efficient to compute and requires less memory storage. Using one-way multivariate analysis of variance (MANOVA) and permutation testing, we show through experiments on a publicly-available database that our proposed GSGW approach yields better performance compared to existing methods in providing an efficient way for comparing shapes of the carpal bones across populations.

The rest of this chapter is organized as follows. In Section 2.2, we briefly provide a brief overview the Laplace-Beltrami operator and its spectral analysis in the discrete domain. In Section 2.3, we introduce a global spectral graph wavelet framework, and we discuss in detail its main algorithmic steps. Experimental results are presented in Sections 2.4. Finally, we conclude in Section 2.5 and point out some future work directions.

2.2 Background

A 3D shape is usually modeled as a triangle mesh $\mathbb{M}$ whose vertices are sampled from a Riemannian manifold. A triangle mesh $\mathbb{M}$ may be defined as a graph $\mathbb{G} = (\mathcal{V}, \mathcal{E})$ or $\mathbb{G} = (\mathcal{V}, \mathcal{T})$, where $\mathcal{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is the set of vertices, $\mathcal{E} = \{e_{ij}\}$ is the set of edges, and $\mathcal{T} = \{\mathbf{t}_1, \ldots, \mathbf{t}_g\}$ is the set of triangles. Each edge $e_{ij} = [\mathbf{v}_i, \mathbf{v}_j]$ connects a pair of vertices $\{\mathbf{v}_i, \mathbf{v}_j\}$. Two distinct vertices $\mathbf{v}_i, \mathbf{v}_j \in V$ are adjacent (denoted by $\mathbf{v}_i \sim \mathbf{v}_j$ or simply $i \sim j$) if they are connected by an edge, i.e. $e_{ij} \in E$.

2.2.1 Laplace-Beltrami Operator

Given a compact Riemannian manifold $\mathbb{M}$, the space $L^2(\mathbb{M})$ of all smooth, square-integrable functions on $\mathbb{M}$ is a Hilbert space endowed with inner product $\langle f_1, f_2 \rangle = \int_{\mathbb{M}} f_1(\mathbf{x}) f_2(\mathbf{x})\, da(\mathbf{x})$, for all $f_1, f_2 \in L^2(\mathbb{M})$, where $da(x)$ (or simply dx) denotes the measure from the area element of a Riemannian metric on $\mathbb{M}$. Given a twice-differentiable, real-valued function $f : \mathbb{M} \to \mathbb{R}$, the Laplace-Beltrami operator (LBO) is defined as $\Delta_{\mathbb{M}} f = -\mathrm{div}(\nabla_{\mathbb{M}} f)$, where $\nabla_{\mathbb{M}} f$ is the intrinsic gradient vector field and div is the divergence operator [59, 60]. The LBO is a linear, positive semi-definite operator acting on the space of real-valued functions defined on $\mathbb{M}$, and it is a generalization of the Laplace operator to non-Euclidean spaces.

Discretization: A real-valued function $f : \mathcal{V} \to \mathbb{R}$ defined on the mesh vertex set may be represented as an m-dimensional vector $\mathbf{f} = (f(i)) \in \mathbb{R}^m$, where the ith component $f(i)$ denotes the function value at the ith vertex in $\mathcal{V}$. Using a mixed finite element/finite volume method on triangle meshes [61], the value of $\Delta_{\mathbb{M}} f$ at a vertex $\mathbf{v}_i$ (or simply i) can be approximated using the cotangent weight scheme as follows:

$$\Delta_{\mathbb{M}} f(i) \approx \frac{1}{a_i} \sum_{j \sim i} \frac{\cot \alpha_{ij} + \cot \beta_{ij}}{2} \big(f(i) - f(j)\big), \tag{2.1}$$

where α_{ij} and β_{ij} are the angles $\angle(\mathbf{v}_i \mathbf{v}_{k_1} \mathbf{v}_j)$ and $\angle(\mathbf{v}_i \mathbf{v}_{k_2} \mathbf{v}_j)$ of two faces $\mathbf{t}^\alpha = \{\mathbf{v}_i, \mathbf{v}_j, \mathbf{v}_{k_1}\}$ and $\mathbf{t}^\beta = \{\mathbf{v}_i, \mathbf{v}_j, \mathbf{v}_{k_2}\}$ that are adjacent to the edge $[i, j]$, and a_i is the area of the Voronoi cell at vertex i. It should be noted that the cotangent weight scheme is numerically consistent and preserves several important properties of the continuous LBO, including symmetry and positive semi-definiteness [62].

Spectral Analysis: The $m \times m$ matrix associated to the discrete approximation of the LBO is given by $\mathbf{L} = \mathbf{A}^{-1}\mathbf{W}$, where $\mathbf{A} = \mathrm{diag}(a_i)$ is a positive definite diagonal matrix (mass matrix), and $\mathbf{W} = \mathrm{diag}(\sum_{k \neq i} c_{ik}) - (c_{ij})$ is a sparse symmetric matrix (stiffness matrix). Each diagonal element a_i is the area of the Voronoi cell at vertex i, and the weights c_{ij} are given by

$$c_{ij} = \begin{cases} \dfrac{\cot \alpha_{ij} + \cot \beta_{ij}}{2} & \text{if } i \sim j \\ 0 & \text{o.w.} \end{cases} \tag{2.2}$$

where α_{ij} and β_{ij} are the opposite angles of two triangles that are adjacent to the edge $[i, j]$.

The eigenvalues and eigenvectors of $\mathbf{L}$ can be found by solving the generalized eigenvalue problem $\mathbf{W}\boldsymbol{\varphi}_\ell = \lambda_\ell \mathbf{A}\boldsymbol{\varphi}_\ell$ using, for instance, the Arnoldi method of ARPACK[1], where λ_ℓ are the

[1] ARPACK (ARnoldi PACKage) is a MATLAB library for computing the eigenvalues and eigenvectors of large matrices.

eigenvalues and $\boldsymbol{\varphi}_\ell$ are the unknown associated eigenfunctions (i.e. eigenvectors which can be thought of as functions on the mesh vertices). We may sort the eigenvalues in ascending order as $0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_m$ with associated orthonormal eigenfunctions $\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \ldots, \boldsymbol{\varphi}_m$, where the orthogonality of the eigenfunctions is defined in terms of the $\mathbf{A}$-inner product, i.e.

$$\langle \boldsymbol{\varphi}_k, \boldsymbol{\varphi}_\ell \rangle_{\mathbf{A}} = \sum_{i=1}^{m} a_i \varphi_k(i) \varphi_\ell(i) = \delta_{k\ell}, \quad \text{for all } k, \ell = 1, \ldots, m. \tag{2.3}$$

We may rewrite the generalized eigenvalue problem in matrix form as $\mathbf{W}\boldsymbol{\Phi} = \mathbf{A}\boldsymbol{\Phi}\boldsymbol{\Lambda}$, where $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_m)$ is an $m \times m$ diagonal matrix with the λ_ℓ on the diagonal, and $\boldsymbol{\Phi}$ is an $m \times m$ orthogonal matrix whose ℓth column is the unit-norm eigenvector $\boldsymbol{\varphi}_\ell$.

The successful use of the LBO eigenvalues and eigenfunctions in shape analysis is largely attributed to their isometry invariance and robustness to noise. Moreover, the eigenfunctions associated to the smallest eigenvalues capture well the large-scale properties of a shape. As shown in Figure 2.1, the (non-trivial) eigenfunctions of the LBO encode important information about the intrinsic global geometry of a shape. Notice that the eigenfunctions associated with larger eigenvalues oscillate more rapidly. Blue regions indicate negative values of the eigenfunctions and red colors regions indicate positive values, while green and yellow regions in between.

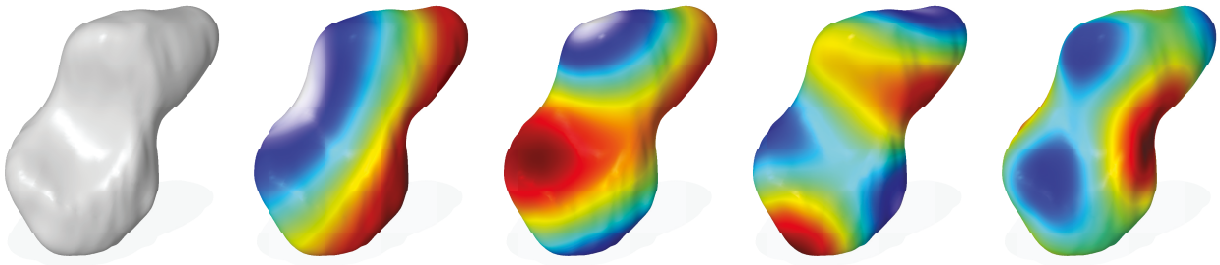


Figure 2.1: From left to right: 3D scaphoid bone and selected eigenfunctions (2nd, 5th, 7th and 10th) of the LBO mapped into the surface of the bone.

2.3 Method

In this section, we provide a detailed description of our GSGW framework for the analysis of the cortical surface of a carpal bone using spectral graph wavelets. We start by defining the spectral graph wavelet transform on a Riemannian manifold. We show how to build local descriptors from spectral graph wavelets and its subcomponent functions. Then, we propose a novel global shape descriptor defined as an area-weighted sum of all local spectral graph wavelet signatures at each mesh vertex. Finally, we provide the main algorithmic steps of our carpal bone analysis framework.

2.3.1 Local Descriptors

Graph Fourier Transform: For any graph signal $f : \mathcal{V} \to \mathbb{M}$, the forward and inverse graph Fourier transforms (also called manifold harmonic and inverse manifold harmonic transforms) are defined as

$$\hat{f}(\ell) = \langle f, \varphi_\ell \rangle = \sum_{i=1}^{m} a_i f(i) \varphi_\ell(i), \quad \ell = 1, \ldots, m \tag{2.4}$$

and

$$f(i) = \sum_{\ell=1}^{m} \hat{f}(\ell) \varphi_\ell(i) = \sum_{\ell=1}^{m} \langle f, \varphi_\ell \rangle \varphi_\ell(i), \quad i \in \mathcal{V}, \tag{2.5}$$

respectively, where $\hat{f}(\ell)$ is the value of f at eigenvalue λ_ℓ (i.e. $\hat{f}(\ell) = \hat{f}(\lambda_\ell)$). In particular, the graph Fourier transform of a delta function δ_j centered at vertex j is given by

$$\hat{\delta}_j(\ell) = \sum_{i=1}^{n} a_i \delta_j(i) \varphi_\ell(i) = \sum_{i=1}^{n} a_i \delta_{ij} \varphi_\ell(i) = a_j \varphi_\ell(j),$$

The forward and inverse graph Fourier transforms may be expressed in vector form as follows:

$$\hat{\mathbf{f}} = \mathbf{\Phi}^\mathsf{T} \mathbf{A} \mathbf{f} \quad \text{and} \quad \mathbf{f} = \mathbf{\Phi} \hat{\mathbf{f}}, \tag{2.6}$$

where $\mathbf{f} = (f(i))$ and $\hat{\mathbf{f}} = (\hat{f}(\ell))$ are m-dimensional vectors, whose elements are given by (2.4) and (2.5), respectively. The vector $\hat{\mathbf{f}}$ represents the signal's graph Fourier series expansion in the area-weighted eigenvector basis and describes the frequency components of the graph signal $\mathbf{f}$.

The inverse graph Fourier transform reconstructs the graph signal by combining graph frequency components, as shown in Figure 2.2, which demonstrate the ability of the LBO eigenfunctions in rendering the shape-based features. As can be seen, the lower-order eigenfunctions capture the global structure of shape, while by increasing the number of eigenfunctions more details of the curvature of the bone are captured.

Figure 2.2: From left to right: 3D hamate bone in a healthy man and its graph Fourier reconstruction using 5, 20, 50 and 100 eigenfunctions of the LBO.

The normalized mean squared error between the original bone surface and its graph Fourier reconstruction is shown in Figure 2.3, where the x-axis is the number of eigenfunctions of the LBO. As can be seen, a relatively small number of eigenfunctions (i.e. between 20 and 30) would be enough to efficiently capture the features of the carpal bone surface to analyze shape differences in a population study. By features, we mean the points on the carpal bones that contain salient information about the shape (e.g. protrusions). The extracted features should be robust to transformations. Our global spectral graph wavelet descriptor is a dense descriptor that makes use of the LBO eigenvalues and eigenfunctions, which are invariant to isometric transformation. Hence, our framework provides a robust descriptor for describing the carpal bones that helps facilitate the statistical analysis among bone shapes. Rendering a carpal bone surface in a lower-dimension has some advantages, including the ability of being invulnerable to tessellation noise or image segmentation

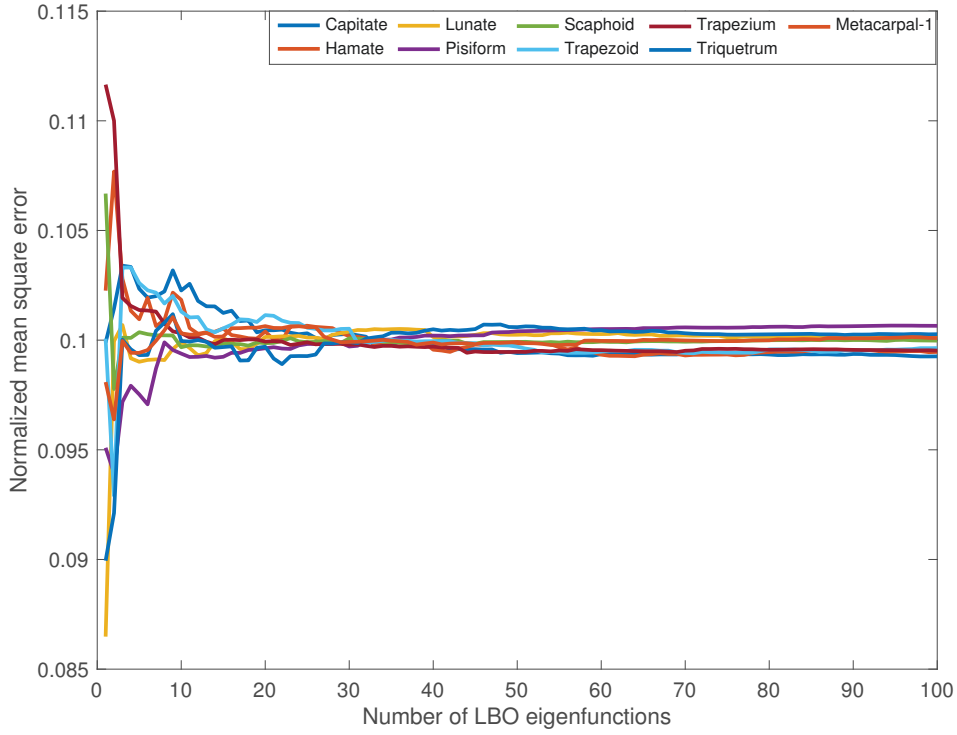


Figure 2.3: Normalized mean squared error between the original carpal bone and its graph Fourier reconstruction as a function of the LBO eigenfunctions.

Spectral Graph Wavelet Transform: Wavelets are useful in describing functions at different levels of resolution. To characterize the localized context around a mesh vertex $j \in \mathcal{V}$, we assume that the signal on the mesh is a unit impulse function, that is $f(i) = \delta_j(i)$ at each mesh vertex

$i \in \mathcal{V}$. The spectral graph wavelet coefficients are expressed as

$$W_{\delta_j}(t, j) = \langle \boldsymbol{\delta}_j, \boldsymbol{\psi}_{t,j} \rangle = \sum_{\ell=1}^{m} a_j^2 g(t\lambda_\ell) \varphi_\ell^2(j), \tag{2.7}$$

and that the coefficients of the scaling function are

$$S_{\delta_j}(j) = \sum_{\ell=1}^{m} a_j^2 h(\lambda_\ell) \varphi_\ell^2(j). \tag{2.8}$$

Following the multiresolution analysis, the spectral graph wavelet and scaling function coefficients are collected to form the spectral graph wavelet signature at vertex j as follows:

$$\mathbf{s}_j = \{\mathbf{s}_L(j) \mid L = 1, \ldots, R\}, \tag{2.9}$$

where R is a resolution parameter, and $\mathbf{s}_L(j)$ is the shape signature at resolution level L given by

$$\mathbf{s}_L(j) = \{W_{\delta_j}(t_k, j) \mid k = 1, \ldots, L\} \cup \{S_{\delta_j}(j)\}. \tag{2.10}$$

The wavelet scales t_k ($t_k > t_{k+1}$) are selected to be logarithmically equispaced between maximum and minimum scales t_1 and t_L, respectively. Thus, the resolution level L determines the resolution of scales to modulate the spectrum. At resolution $R = 1$, the spectral graph wavelet signature $\mathbf{s}_j$ is a 2-dimensional vector consisting of two elements: one element, $W_{\delta_j}(t_1, j)$, of spectral graph wavelet function coefficients and another element, $S_{\delta_j}(j)$, of scaling function coefficients. And at resolution $R = 2$, the spectral graph wavelet signature $\mathbf{s}_j$ is a 5-dimensional vector consisting of five elements (four elements of spectral graph wavelet function coefficients and one element of scaling function coefficients). In general, the dimension of a spectral graph wavelet signature $\mathbf{s}_j$ at vertex j can be expressed in terms of the resolution R as follows:

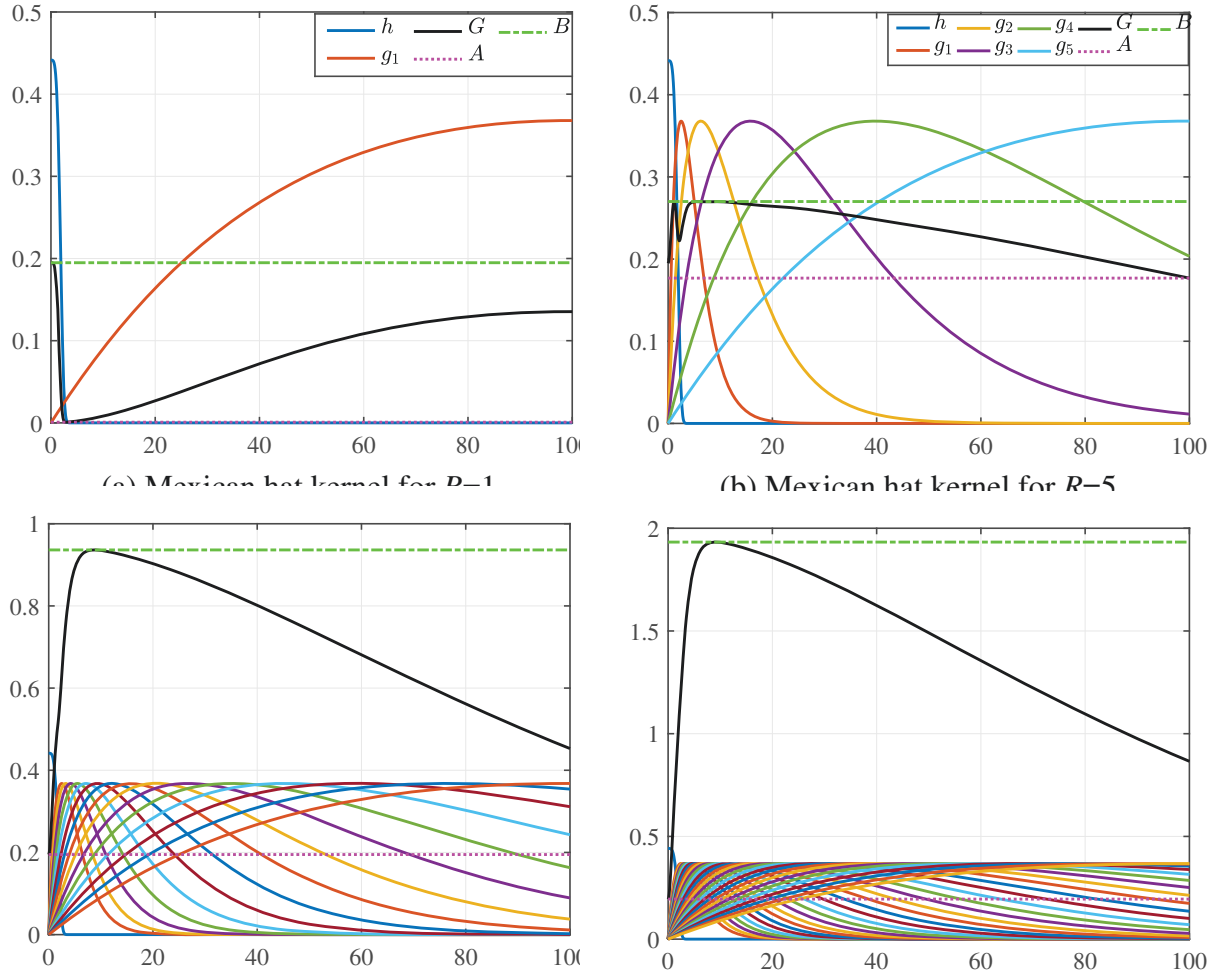
$$p = \frac{(R+1)(R+2)}{2} - 1. \tag{2.11}$$

Hence, for a p-dimensional signature $\mathbf{s}_j$, we define a $p \times m$ spectral graph wavelet signature matrix as $\mathbf{S} = (\mathbf{s}_1, \ldots, \mathbf{s}_m)$, where $\mathbf{s}_j$ is the signature at vertex j and m is the number of mesh vertices. In our implementation, we used the Mexican hat wavelet as a kernel generating function g. In addition, we used the scaling function h given by

$$h(x) = \gamma \exp\left(-\left(\frac{x}{0.6\lambda_{\min}}\right)^4\right), \tag{2.12}$$

where $\lambda_{\min} = \lambda_{\max}/20$ and γ is set such that $h(0)$ has the same value as the maximum value of g. The maximum and minimum scales are set to $t_1 = 2/\lambda_{\min}$ and $t_L = 2/\lambda_{\max}$, where $\lambda_{\min}$ and $\lambda_{\max}$ are the smallest and largest LBO eigenvalues, respectively.

The geometry captured at each resolution R of the spectral graph wavelet signature can be viewed as the area under the curve G shown in Figure 2.4. For a given resolution R, we can understand the information from a specific range of the spectrum as its associated areas under G. As the resolution R increases, the partition of spectrum becomes tighter, and thus a larger portion of the spectrum is highly weighted.



(a) Mexican hat kernel for R=1 (b) Mexican hat kernel for R=5

(c) Mexican hat kernel for R=15 (d) Mexican hat kernel for R=30

Figure 2.4: Spectrum modulation using different kernel functions at various resolutions. The dark line is the squared sum function G, while the dash-dotted and the dotted lines are upper and lower bounds (B and A) of G, respectively.

2.3.2 Global Descriptor

A commonly used methodology for building a global shape descriptor is by aggregating local signatures using the BoF paradigm. However, a major drawback of the BoF model is that it only

considers the distribution of the codewords and disregards all information about the spatial relations between features, and hence the descriptive ability and discriminative power of the BoF paradigm may be negatively impacted. In addition, the BoF process is time-consuming since it requires different steps such as constructing dictionary, feature coding and feature pooling. To circumvent these limitations, we represent a shape $\mathbb{M}$ by a p-dimensional vector

$$\mathbf{x} = \mathbf{S}\mathbf{a} = \sum_{i=1}^{m} a_i \mathbf{s}_i, \tag{2.13}$$

where $\mathbf{S} = (\mathbf{s}_1, \ldots, \mathbf{s}_m)$ is a $p \times m$ matrix of local spectral graph wavelet signatures and $\mathbf{a} = (a_1, \ldots, a_m)^\mathsf{T}$ is an m-dimensional vector of mesh vertex areas (i.e. each element a_i is the area of the Voronoi cell at mesh vertex i).

We refer to the p-dimensional vector $\mathbf{x}$ as the global spectral graph wavelet (GSGW) descriptor of the carpal bone surface. The GSGW descriptor enjoys a number of desirable properties including simplicity, compactness, invariance to isometric deformations, and computational feasibility. Moreover, GSGW combines the advantages of both band-pass and low-pass filters.

2.3.3 Proposed Algorithm

Our proposed carpal bone analysis algorithm consists of two main steps. In the first step, we represent each bone in the dataset by a spectral graph wavelet signature matrix, which is a feature matrix consisting of local descriptors. More specifically, let $\mathcal{D}$ be a dataset of n carpal bones modeled by triangle meshes $\mathbb{M}_1, \ldots, \mathbb{M}_n$. We represent each surface $\mathbb{M}_i$ in the dataset $\mathcal{D}$ by a $p \times m$ spectral graph wavelet signature matrix $\mathbf{S}_i$, whose columns are p-dimensional local signatures and m is the number of mesh vertices.

In the second step, we compute the p-dimensional global spectral graph wavelet descriptor $\mathbf{x}_i = \mathbf{S}_i\mathbf{a}_i$ of each carpal bone $\mathbb{M}_i$, for $i = 1, \ldots, n$. Subsequently, the feature vectors $\mathbf{x}_i$ of all n shapes in the dataset are arranged into a $n \times p$ data matrix $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^\mathsf{T}$.

To assess the performance of the proposed GSGW framework, we employed two commonly-used evaluation criteria, namely MANOVA and permutation testing. MANOVA is a multivariate data analysis technique used to determine whether there are any statistical differences between independent groups on more than one continuous dependent variable, while a permutation test in a non-parametric test that resamples the observed data many times in order to determine a p-value for the test. The p-value is the probability of obtaining an effect at least as extreme as the one in our observed data when the null hypothesis is true, and it basically measures how compatible our data are with the null hypothesis. A small p-value provides enough evidence that we can reject the null hypothesis. Algorithm 1 summarizes the main algorithm steps of our GSGW approach.

loa 1: GSGW approach

Input: Dataset $\mathcal{D} = \{\mathbb{M}_1, \ldots, \mathbb{M}_n\}$ of n carpal bones

1: **for** $i = 1$ to n **do**
2: Compute the $p \times m$ spectral graph wavelet matrix $\mathbf{S}_i$ for each carpal bone $\mathbb{M}_i$, where m is the number of vertices
3: Compute the p-dimensional vector $\mathbf{x}_i = \mathbf{S}_i \mathbf{a}_i$, where $\mathbf{a}_i$ is an m-dimensional vector of vertex areas
4: **end for**
5: Arrange all the feature vectors $\mathbf{x}_i$ into a $n \times p$ data matrix $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^\mathsf{T}$
6: Perform MANOVA and permutation test on $\mathbf{X}$ to quantify the statistical differences between carpal bones

Output: p-values for MANOVA and permutation test.

2.4 Experimental Results

In this section, we evaluate the performance of our proposed GSGW approach on the analysis of carpal bone surfaces via extensive experiments. The effectiveness of our method is validated by performing a comprehensive comparison with the global point signature embedding approach [23].

Datasets: In order to evaluate the performance of our GSGW framework on carpal bone surfaces, a total of 20 men and women with average age of 25 years old from a publicly-available benchmark [12] have been chosen. In this dataset, the bones of the wrist are obtained from CT volume images. More precisely, the carpal bones undergo segmentation by detecting the 2D outer cortical bone contours in each image slice. For each bone, the contours are identified and then aggregated into a single 3D point cloud. Finally, a triangular mesh is constructed using the acquired points. Each triangular mesh consists of an edge set (i.e. connectivity list) and vertex locations (i.e. vertex set). We also performed uniform sampling on triangular meshes to have an equal number of vertices. As shown in Figure 2.5, the carpal bones of the right wrist in a healthy male are the capitate, hamate, lunate, pisiform, scaphoid, trapezium, trapezoid, triquetrum. Since the trapeziometacarpal joint of the thumb is a common site of osteoarthritis, the first metacarpal bone is also considered in our analysis. The forearm's radius and ulna bones, which support the many muscles that manipulate the bones of the hand and wrist, are also depicted in Figure 2.5.

Performance Evaluation Measures To compare the shapes of the carpal bones in women versus men, we computed the GSGW descriptor for each carpal bone (eight in total) and the first metacarpal bone for each subject for both the right and left wrists. For fair comparison between the proposed GSGW approach and GPS embedding method, we followed the same settings described in [12]. In order to quantify the difference between the sexes for the carpal bone shapes,

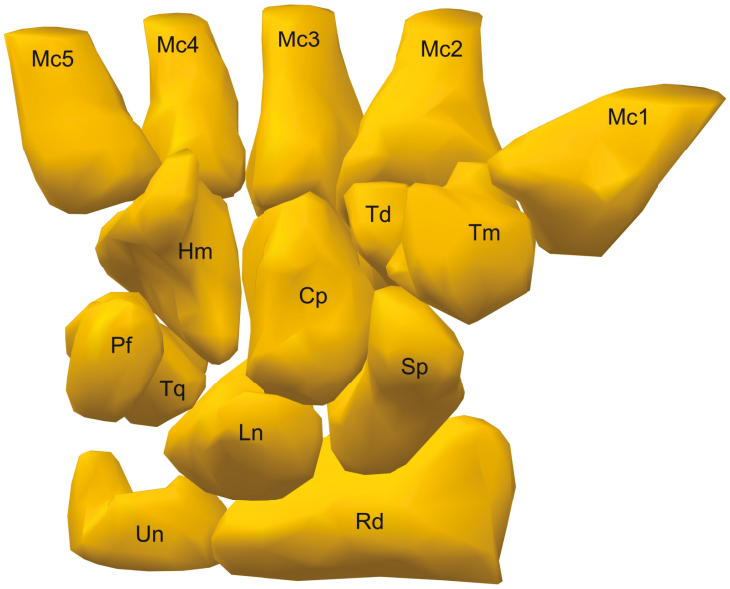


Figure 2.5: Carpal bone anatomy of a healthy male from a palmar view. The carpus consists of eight carpal bones, which are arranged in proximal and distal rows. The proximal row contains scaphoid (Sp), lunate (Ln), triquetrum (Tq) and pisiform (Pf), while the distal row contains trapezium (Tm), trapezoid (Td), capitate (Cp), and hamate (Hm). The distal row adjoins the five metacarpals (Mc1-5) of the wrist. The radius (Rd) and ulna (Un) are also shown.

we compared GSGW to GPS embedding for each bone of the right and left wrist separately for the two groups (ten women versus ten men) using MANOVA and permutation testing.

For permutation testing, gender labels of the samples are randomly shuffled for 1000 times to get the correct distribution of a test statistic under a null hypothesis. We report the p-values generated by MANOVA and permutation testing. For $p < 0.05$, there would be a statistically significant difference between the two groups.

Baseline Method: For the wrist benchmark [12] used for experimentation, we report the comparison results of our method against the GPS embedding approach [23].

Implementation Details: The experiments were conducted on a desktop computer with an Intel Core i5 processor running at 3.10 GHz and 8 GB RAM; and all the algorithms were implemented in MATLAB. The appropriate dimension (i.e. length or number of features) of a shape signature is problem-dependent and usually determined experimentally. For fair comparison, we used the same parameters that have been employed in the baseline method, and in particular the dimensions of shape descriptors. In our setup, a total of 31 eigenvalues and associated eigenfunctions of the LBO were computed. We also set the resolution parameter to $R = 30$, resulting in a 495-dimensional GSGW descriptor.

2.4.1 Carpal Bone Dataset

The carpal bone dataset consists of 360 mesh models from 20 classes [12]. The bones of the wrist are obtained from the CT volume images, and then the carpal bones are rendered and represented as triangular mesh models. Each class contains 18 objects with distinct postures. Moreover, each model in the dataset has approximately $m = 1502$ vertices.

Results In our GSGW approach, each surface in the carpal bone dataset is represented by a 495×1502 matrix of spectral graph wavelet signatures, resulting in a data matrix $\mathbf{X}$ of size 495×360. Figure 2.6 shows the spectral graph wavelet descriptors of two carpal bones (capitate and lunate) from two different classes of the carpal bone dataset. As can be seen, the global descriptors are quite different and hence may be used to efficiently discriminate between surfaces in statistical analysis tasks.

We compared the proposed GSGW method to the GPS embedding approach by performing MANOVA and non-parametric permutation testing. The results are summarized in Table 2.7 for the right wrist and Table 2.8 for the left wrist. In these tables, the numbers marked with an asterisk indicate that the p-value exceeds the significance level of 0.05.

As can be seen, our method achieves better analytical performance than the GPS embedding method for both right and left wrist. For the right wrist, the GSGW approach yields the lower p-value compared to GPS embedding for six carpal bones out of nine using MANOVA, and for seven bones out of nine using permutation testing. In addition, the p-value in the MANOVA test for some bones (e.g. hamate) has decreased to 6.5×10^{-3}. For the left wrist, our GSGW approach significantly improves the results by yielding a lower p-value in the MANOVA test for all bones, except for Metacarpal-1. Also, unlike GPS embedding, our method achieved a much lower p-value in the permutation test for all carpal bones. Moreover, the p-value in the MANOVA test for some bones (e.g. hamate) has plummeted to 2.5×10^{-3}. To speed-up experiments, all shape descriptors were computed offline, albeit their computation is quite inexpensive due in large part to the fact that only a relatively small number of the LBO eigenfunctions need to be computed.

To further assess the discriminative power of our approach, we computed the GSGW descriptors of carpal bone surfaces from the same class. As shown in Figures 2.9 and 2.10, even for very similar carpal bones with a slightly difference, the proposed GSGW approach is able to distinguish between the shapes.

Parameter Sensitivity: The proposed approach depends on two key parameters that affect its overall performance. The first one is the resolution parameter R of the spectral graph wavelet signature. The second one is the number of eigenfunctions of the LBO, which plays an important role in the GSGW descriptor. As shown in Figure 2.11, the best MANOVA result on the carpal

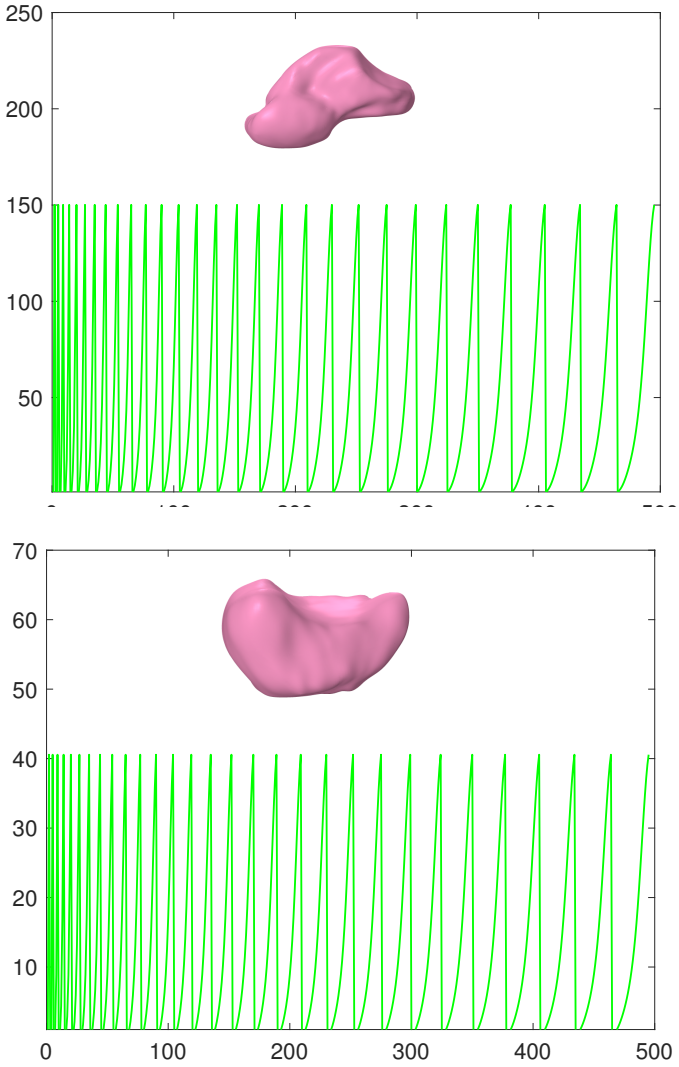


Figure 2.6: Global spectral graph wavelet descriptors of two carpal bones: capitate (Top) and lunate (Bottom).

bones dataset is achieved using 30 eigenfunctions and a resolution parameter $R = 30$. In addition, the performance of proposed method using MANOVA is satisfactory for a wide range of parameter values, indicating the robustness of the GSGW framework to the choice of these parameters.

Bone	MANOVA		Permutation Test	
	GPS	GSGW	GPS	GSGW
Capitate	**0.0226**	0.0319	0.0493	**0**
Hamate	0.0065	**2.37e-11**	0.0097	**0.0020**
Lunate	0.0379	**0.0069**	0.0210	**0**
Pisiform	**0.0428**	0.0441	0.0365	**0.0040**
Scaphoid	0.0135	**0.0003**	0.0255	**0**
Trapezoid	**0.0004**	0.0088	0.0087	**0**
Trapezium	0.0007	**1.80e-5**	**0.0101**	0.0200
Triquetrum	0.0015	**1.96e-5**	**0.0124**	$0.200^{\star}$
Metacarpal-1	0.0137	**0.0009**	0.0245	**0.0040**

Figure 2.7: Comparison of carpal bone surfaces of the right wrist between males and females using MANOVA and permutation test. Boldface numbers indicate the better performance, while the numbers marked with an asterisk indicate that the p-value exceeds 0.05

Bone	MANOVA		Permutation Test	
	GPS	GSGW	GPS	GSGW
Capitate	0.0148	**0.0065**	0.0416	**0.0100**
Hamate	0.0048	**6.64e-8**	0.0089	**0.0020**
Lunate	$0.0532^{\star}$	**2.41e-5**	$0.0766^{\star}$	**0.0020**
Pisiform	0.0102	**2.56e-5**	0.0201	**0**
Scaphoid	0.0012	**3.19e-5**	0.0120	**0.0060**
Trapezoid	0.0022	**0.0013**	0.0136	**0.0020**
Trapezium	0.0023	**0.0002**	0.0140	**0**
Triquetrum	0.0449	**0.0005**	$0.0836^{\star}$	**0**
Metacarpal-1	**0.0015**	0.0482	0.0129	**0.0100**

Figure 2.8: Comparison of carpal bone surfaces of the left wrist between males and females using MANOVA and permutation test. Boldface numbers indicate the better performance, while the numbers marked with an asterisk indicate that the p-value exceeds 0.05.

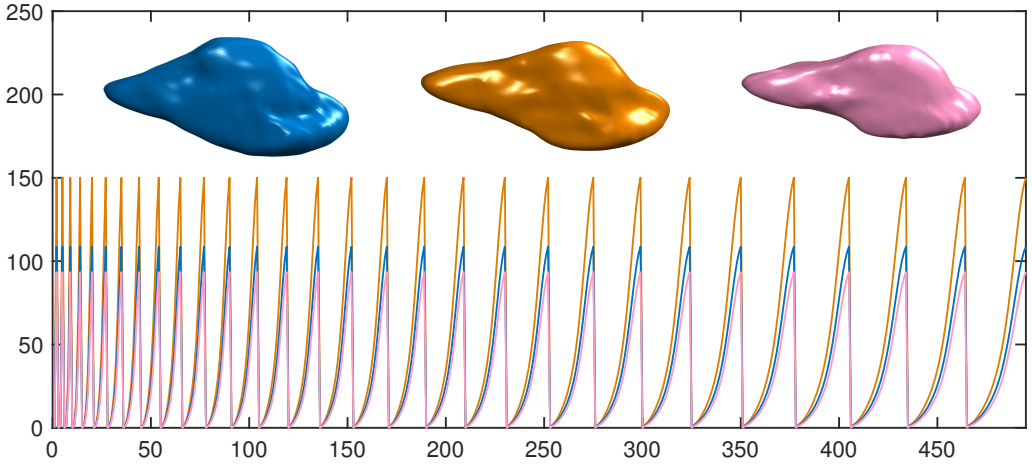


Figure 2.9: Global spectral graph wavelet descriptors for three capitate bones of women's left wrists.

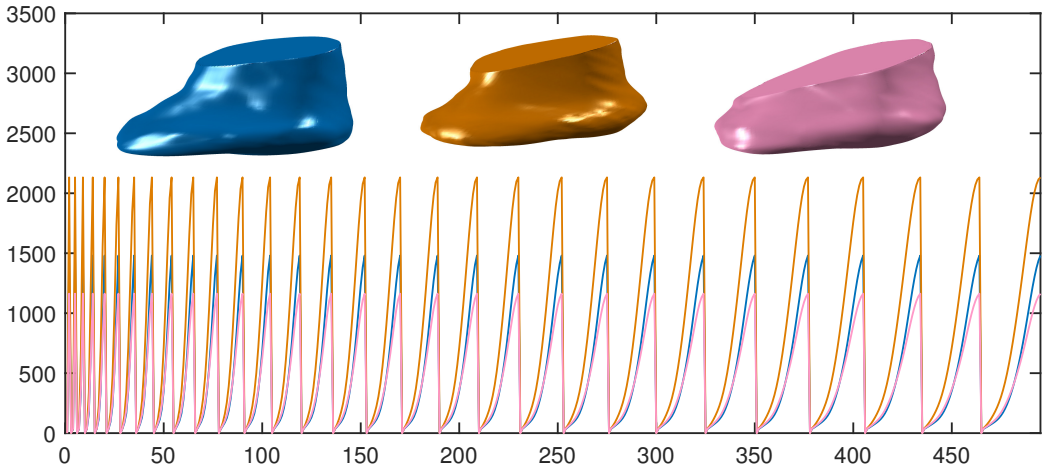


Figure 2.10: Global spectral graph wavelet descriptors for three metacarpal bones of women's left wrists.

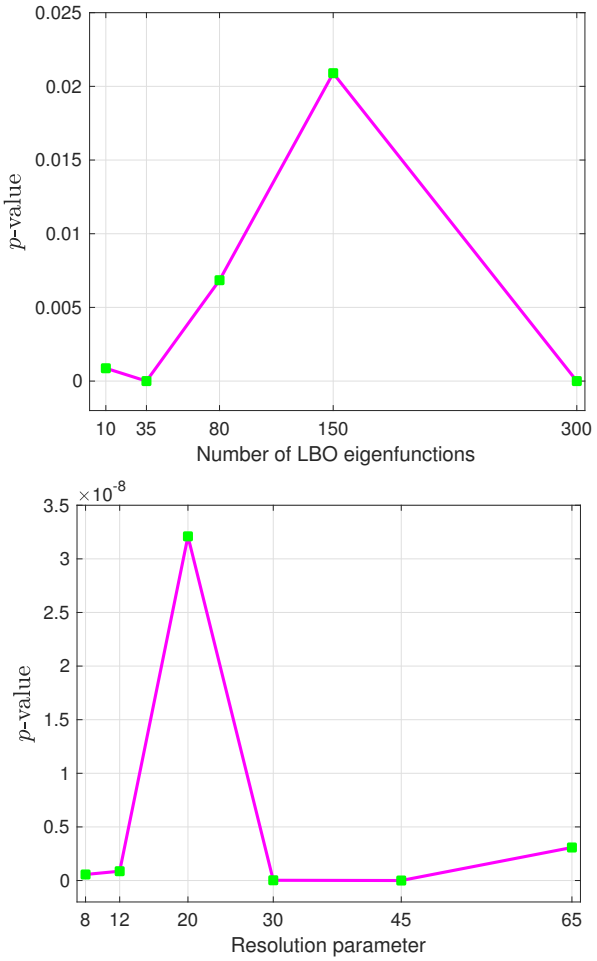


Figure 2.11: From top to bottom: effect of the parameters on MANOVA results for the hamate bone in terms of number of eigenfunctions and resolution parameter, respectively.

2.5 Conclusion

In this chapter, we introduced a spectral-geometric based framework for population study of carpal bones. We proposed an efficient representation of the cortical surface of the carpal bone via a global spectral graph wavelet descriptor, and then performed statistical analysis using MANOVA and permutation testing on the carpal bones of the human wrist in an effort to compare shapes of the carpal bones across populations. The proposed GSGW descriptor enjoys a number of desirable properties including simplicity, compactness, invariance to isometric deformations, and computational feasibility. Moreover, our approach not only captures the similarity between feature descriptors, but also substantially outperforms existing methods.

CHAPTER 3

Bone age assessment using deep learning for 2D X-ray images

3.1 Introduction

In this chapter, we describe the background of bone age estimation and the relevant deep learning models employed as pre-trained models for regression analysis. Then, we show the proposed architecture on RSNA dataset, followed by training and fine-tuning processes to implement the regression task and consequently discuss the results.

3.2 Background and Relevant Work

BAA is a clinical method for determining age discrepancy between one's skeletal bone age and their chronological age (real age from birthdate). It has always been considered as an applicable skill for various fields, such as assessment of bone age by observing the eruption of the second molar on young men for calling them for the Roman empire military service [63]. In the 19th century, tooth eruption was estimated children's age as a reliable method by dentists, and the minimum bone age was considered 7 years old in Britain [64]. In 1846, unreliability of bone age estimation based on the only observance of tooth eruption was announced in [65]. In 1886, an idea of the age estimation in young people was introduced for the first time [66], and then the first systematic assessment of age variations in carpal bones was proposed in 1887 [67]. In 1895, discovery of X-rays triggered a revolution in the estimation of bone age, in which the skeleton radiography was

applied as a complementary approach for tooth eruption [65]. This was the beginning of a new era for the researchers to determine the age of the subjects based on wrist bone maturity depicted on the radiographs [68]. Between 1950 and 1980, the most important traditional manual methods, Greulich and Pyle (GP) [69] and Tanner-Whitehouse (TW) [70] were defined for the BAA based on radiological analysis of carpus bone. However, the disadvantages of these methods were resulted in reducing the accuracy of BAA. In fact, the methods are time-consuming and prone to inter- and intra-observer variability [71] that cause problems in their clinical application; the comparison between subjects and follow-up of patients. Hence, these drawbacks motivate an automatic BAA system to be introduced in the past 3 decades [72, 73].

Recently, the importance of BAA for living subjects has been much attracted in Europe, particularly for a precise census population to verify citizen's age. By increasing a large number of immigrants to European countries, who do not have any birth certificate to show their real ages, the problem of counting the young immigrant's population annually has arisen. In 2003 and 2006, the census of immigrants in Spain revealed about 6,000 people from sub-Saharan Africa, who had traveled by trucks and dangerous boats via the Mediterranean Sea to Spain with unknown real ages. At that time, no standard method for the BAA existed for illegal immigrants in European countries. Some countries, such as France and the UK, only held an interview without any expert examination with those who do not have any identification documents. In 2010, a multi-factorial method has introduced by Austria in three levels, including evaluation by a doctor, a dental examination and X-ray analysis with low confidence. Thus, to address the unaccompanied minors issue, an accurate and effective forensic method is needed to provide an age estimation report in a criminal proceeding and public courts in European countries [74].

Recently, deep-learning methods have shown improvements in performance over conventional machine learning approaches for many biomedical problems. In the medical imaging field, convolutional neural networks (CNN) have been successfully applied for a wide variety of medical problems, such as bone diseases prediction, breast cancer histology image analysis and diabetic retinopathy screening. Therefor, we tried to apply a fully automated deep learning approach for the BAA problem on the Radiological Society of North America (RSNA) dataset. This chapter aims to investigate several automated methods for the evaluation of bone age based on left-hand radiographs for the living subjects. In this chapter, we review some deep learning models and discuss the methodology and the implementation of the system.

3.3 Deep Pre-trained Models

A pre-trained model is trained on more than a million images and can classify images into 1000 object categories, such as many animals, mug, coffee and keyboard. The model includes deep learning model weights that have already learned to extract informative and high-level feature representations from natural images in the image classification task, which is utilized as a starting point to learn a new task. Due to computational and time restrictions, building a model from scratch is impossible. Therefore, a pre-trained model is exploited for promoting the existing models or testing on other models much faster and easier than training a network from scratch. To evaluate the performance of the recent most powerful CNN architectures for regression analysis, we focus on six different CNN architectures, i.e. VGG16, VGG19, Inception-V3, MobileNet-V2, NasNetlarge, and InceptionResNet-V2.

3.3.1 VGG16 and VGG19

The visual geometry group (VGG) network architecture is a very deep convolutional network for large scale image recognition as shown in Figure 3.1 [1]. The network is characterized by its simplicity, using only 3×3 convolutional layers (Conv) stacked on top of each other in increasing depth. In the convolutional layer, each neuron is connected to a small region of the input volume (3×3 region), but extends or slides along the full depth of the input volume. As the filter slides across the input volume, the dot product of the entries of the filter and the input are calculated. Then, reducing volume size is handled by max-pooling. Two fully-connected layers, each with $4,096$ neurons are then followed by a softmax classifier. The number of 16 and 19 stand for the number of weight layers in these two networks as shown in Table 3.1.

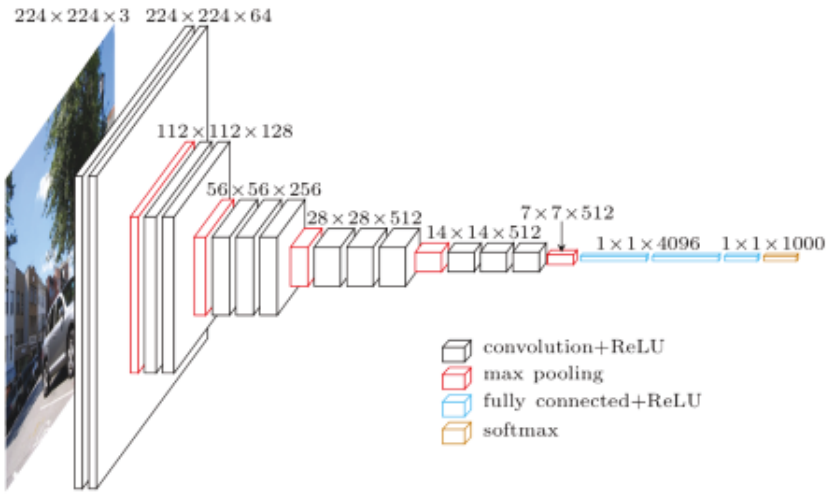


Figure 3.1: The VGG16 architecture [1]

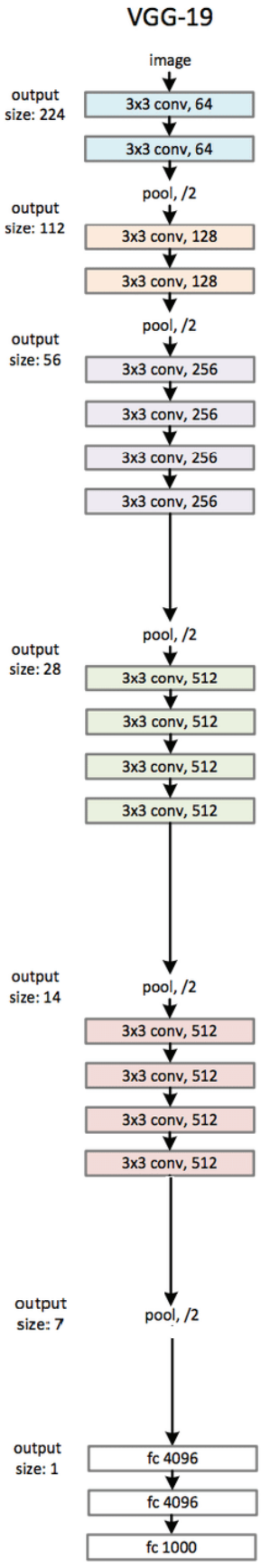


Figure 3.2: the VGG 19 architecture [1]

Table 3.1: Structure of VGG16 (column A) and VGG19 (column B) pre-trained models [1].

ConvNet Configuration	
A	**B**
16 weight layer	19 weight layer
Input(224×224 RGB image)	
conv3-64 conv3-64	conv3-64 conv3-64
maxpool	
conv3-128 conv3-128	conv3-128 conv3-128
maxpool	
conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool	
conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool	
FC-4096	
FC-4096	
FC-1000	
soft-max	

In 2014, 16 and 19 layer networks were considered very deep (although we now have the ResNet architecture which can be successfully trained at depths of 50-200 for ImageNet and over 1,000 for CIFAR-10). The VGG designers found training VGG16 and VGG19 challenging (specifically regarding convergence on the deeper networks). Therefore, to make the training process easier, they first trained smaller versions of VGG with fewer weight layers (columns A and C).

Two major drawbacks of VGGNet include the low speed of the training process and the large disk or bandwidth being used due to the large number of weights in the network architecture. Becuase of its depth and number of fully-connected nodes, VGG is over 533 MB for VGG16 and 574 MB for VGG19, the application of VGG is a tiresome task.

3.3.2 Inception-V3

Inception-V3 is a convolutional neural network that has 48 layers depth and was introduced by factorization idea [2]. The inception-V3 architecture has similar complexity as the VGG model. The most conspicuous feature of this architecture is the use of fewer parameters, where Inception-V1 with 7 million parameters has a lower number of parameters than AlexNet with 60 million parameters and VGGNet with three times more parameters than AlexNet. Furthermore, this model is computationally efficient regarding fewer parameters. Due to the lower error rate, the model is known as the 1st runner up for image classification in ImageNet large scale visual recognition competition (ILSVRC) in 2015.

Inception-V3 Architecture The model has a mixture of symmetric and asymmetric building blocks, including convolutions, average pooling, max-pooling, concats, dropouts, and fully connected layers. Furthermore, the loss is calculated using softmax. The Figure 3.16 depicts the architecture of the model at a high level, which is decomposed of different parts:

Factorizing Convolutions In order to reduce the number of parameters/connections without decreasing the efficiency of the network, factorizing convolutions are applied to smaller convolutions. Two 3×3 convolutions replaces one 5×5 convolution as shown in Figure 3.4 such that once a layer of 5×5 filter is applied, the number of parameters equals 25. By using 2 layers of 3×3 filters, number of parameters is $2 \times (3 \times 3) = 18$. Thus, number of parameters is reduced by 28% as depicted in Figure 3.5. To factorize into asymmetric convolutions, a 3×1 convolution followed by a 1×3 convolution is replaced by as 3×3 convolution as shown in Figure 3.6.

Intuitively, by using 3×3 filter, the number of parameters is $3 \times 3 = 9$, while using 3×1 and 1×3 filters, leads to a number of parameters as $3 \times 1 + 1 \times 3 = 6$. Thus, the number of parameters is reduced by 33% as shown in Figure 3.7.

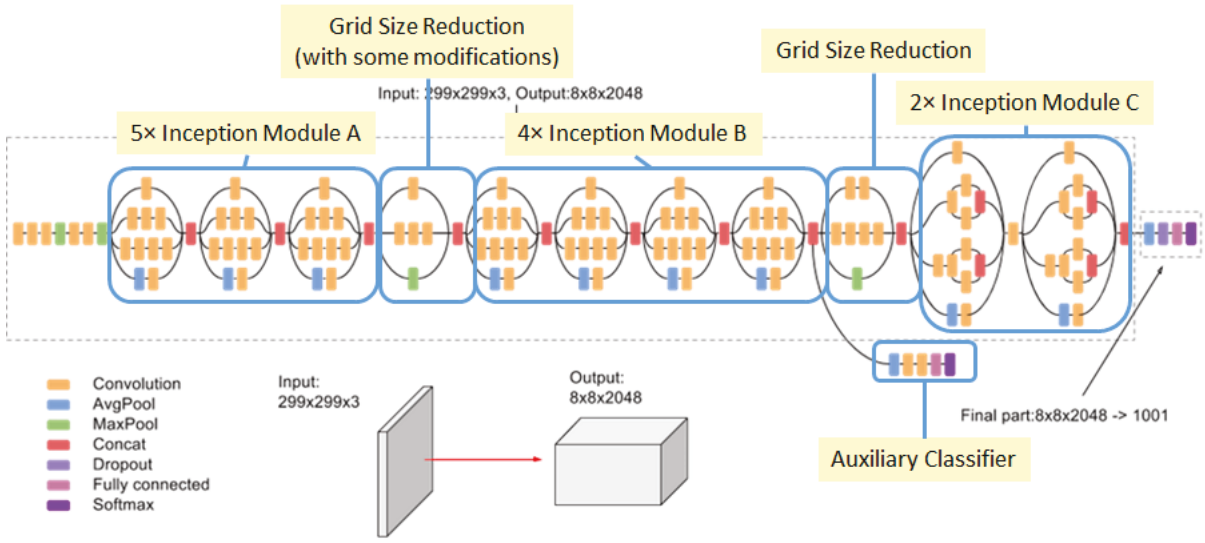


Figure 3.3: Inception-V3 Architecture [2]

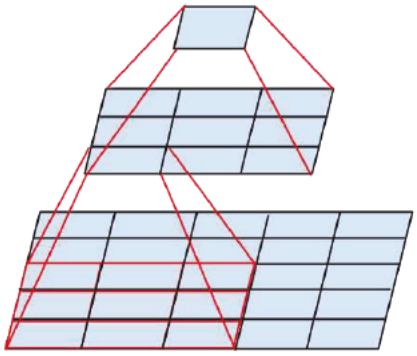


Figure 3.4: Mini-network (Two 3×3 convolutions) replacing the 5×5 convolutions [2].

To improve high dimensional representations, Inception module C is introduced as shown in Figure 3.8. Therefore, the advantage of using these three Inception modules is to reduce the number of parameters and the possibility of over-fitting for the whole network.

Auxiliary Classifier Only an auxiliary classifier is employed on the top of the last 17×17 layer. It is worth noting that the auxiliary classifier acts as a regularizer.

Efficient Grid Size Reduction

The task of max-pooling is to downsize the feature map. However, if the max-pooling is followed by a conv layer, the network becomes too greedy. Moreover, if the conv layer is followed by a max-pooling, the network becomes computationally expensive. Thereby, a technique for efficient reduction of the grid size is proposed as depicted Figure 3.10.

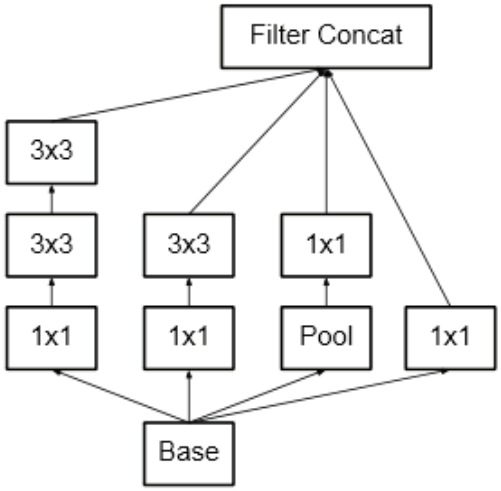


Figure 3.5: Inception Module **A** using factorization [2]

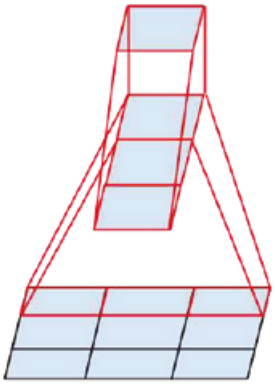


Figure 3.6: Replacing one 3×1 convolution followed by one 1×3 convolution with one 3×3 convolution [2]

By considering the efficient grid size reduction, 320 feature maps are created by Conv with stride of 2, and also 320 feature maps are generated by max-pooling. Then, the result of concatenating these 2 sets of feature maps is 640 feature maps as input for the next level of the inception module. Thus, the advantage of the used architecture is to having a less expensive and still efficient network.

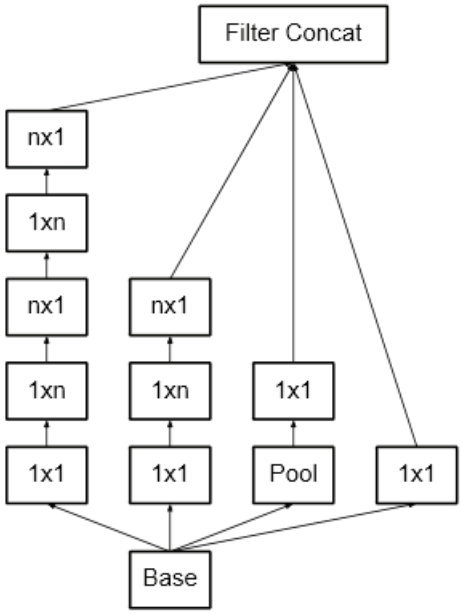


Figure 3.7: Inception Module B using asymmetric factorization of the $n \times n$ convolutions, which any $n \times n$ convolution can be replaced by a $1 \times n$ convolution followed by a $n \times 1$ convolution, where $n = 7$ [2]

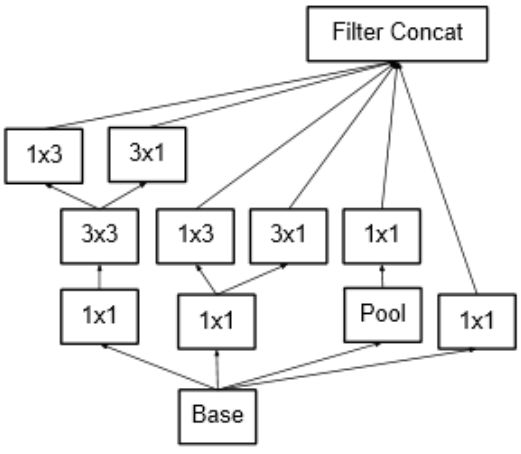


Figure 3.8: Inception Module C using asymmetric factorization [2]

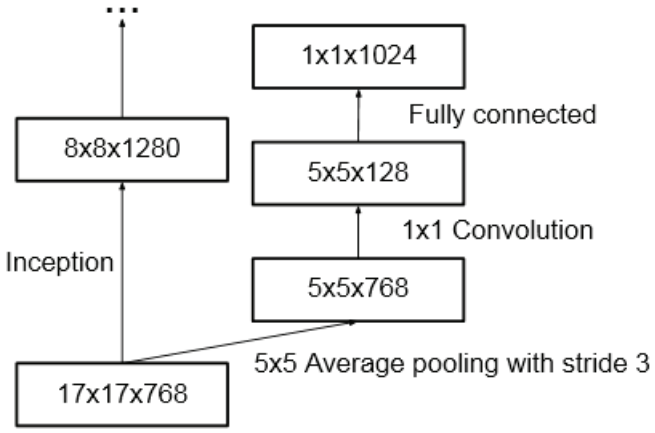


Figure 3.9: One auxiliary classifier shows at the right-most architecture and acts as a regularization and the left-most shows the main branch of Inception-v3 [2]

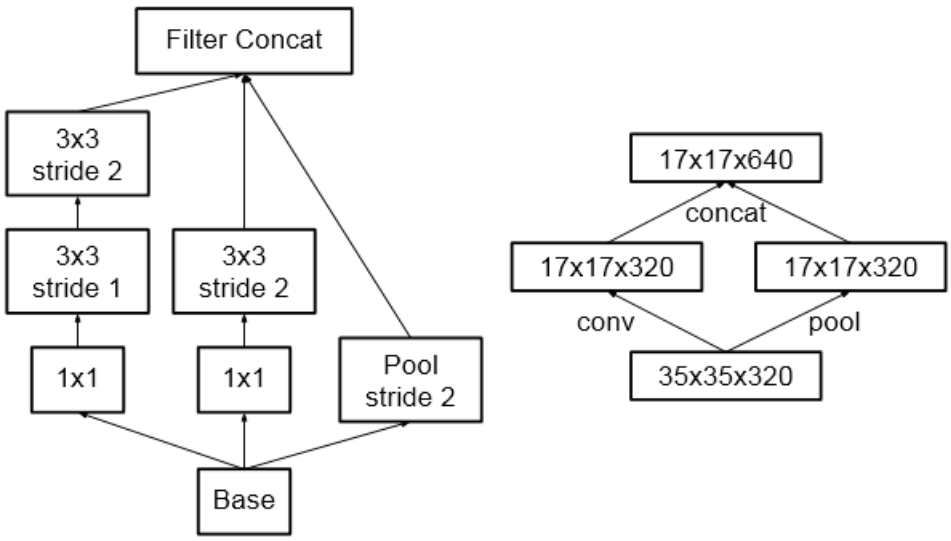


Figure 3.10: Detailed architecture of efficient grid size reduction (Left), efficient grid size reduction (Right) [2].

3.3.3 MobileNet-V2 Architecture

MobileNet architecture has been designed for mobile devices by Google. The model contains pre-trained weights on the popular ImageNet database, which is a database containing millions of images belonging to more than $20,000$ classes. MobileNet-V2 is one of the MobileNet categories that is a small, low-latency, and low-power model parameterized to meet the resource constraints in different cases, such as classification, detection, embeddings and segmentation [75].

MobileNet-V2 has two types of blocks, including residual blocks with the stride of 1 and stride of 2 for downsizing. The idea is based on replacing the standard convolution layers with several separate layers, which means a full standard 3×3 convolution layer replaces with a 1×1 convolution layer (pointwise convolution) and a 3×3 depthwise convolution layer. Pointwise convolution is used to create a linear combination of the output of the depthwise layer, and depthwise convolution uses a single filter per each input channel (input depth). Also, ReLU is used in both layers.

A full standard convolution is always the combination of two operations, including the data organization and the feature extraction, in which operation of data organization should be determined by the nature of data. Thus, to improve the computation efficiency of these two operations, the 1×1 convolution and the depthwise convolutions are employed for organizing data and extracting features, respectively. The structure of the MobileNet-V2 module is shown in Figure 3.11.

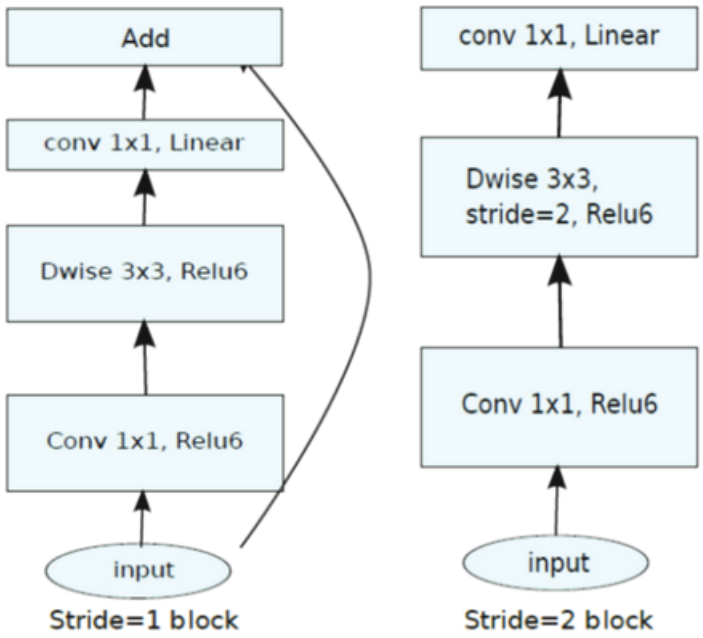


Figure 3.11: MobileNet-V2 model with the stride of 1 and 2. In both architectures, the model has 3 layers. the first layer is 1×1 convolution (pointwise convolution) with ReLU6. The next layer is the depthwise convolution and the last layer is another 1×1 convolution but without any non-linearity. It should be noted that if ReLU is used again, the deep networks only have the power of a linear classifier on the non-zero volume part of the output domain. [3]

3.3.4 NASNetLarge Architecture

NASNetLarge model was introduced by Zoph et al. [76] that is a convolutional neural network trained on over one million images from the ImageNet dataset. Neural architecture search (NAS) is a widely used model in the field of machine learning for automating the design of artificial neural networks (ANN). The model enables us to classify images into 1000 object categories, and learn the model architectures directly on the dataset of interest, which has been proposed to design networks that are on par or outperform hand-designed architectures. This technique results in extracting better discriminative information (high feature representation) from a wide range of images.

3.3.5 InceptionResNet-V2 Architecture

In image classification, InceptionResNet-V2 [4] is a state-of-the-art convolutional neural network (CNN) that achieves better accuracy among previous models on the ILSVRC. More precisely, InceptionResNet-V2 is a lightweight package for defining, training and evaluating models, which borrows some ideas from Microsoft's ResNet papers [77, 78] to upgrade the earlier Inception-V3 model. This model propose residual connections in its design, allowing shortcuts in the model by simplifying the Inception blocks with fewer parallel towers than the previous Inception-V3 as shown in Figures 3.12 and 3.13. Considering the ability, researchers are able to successfully train even deeper neural networks. Therefore, This ability is led to even better performance in recognition tasks.

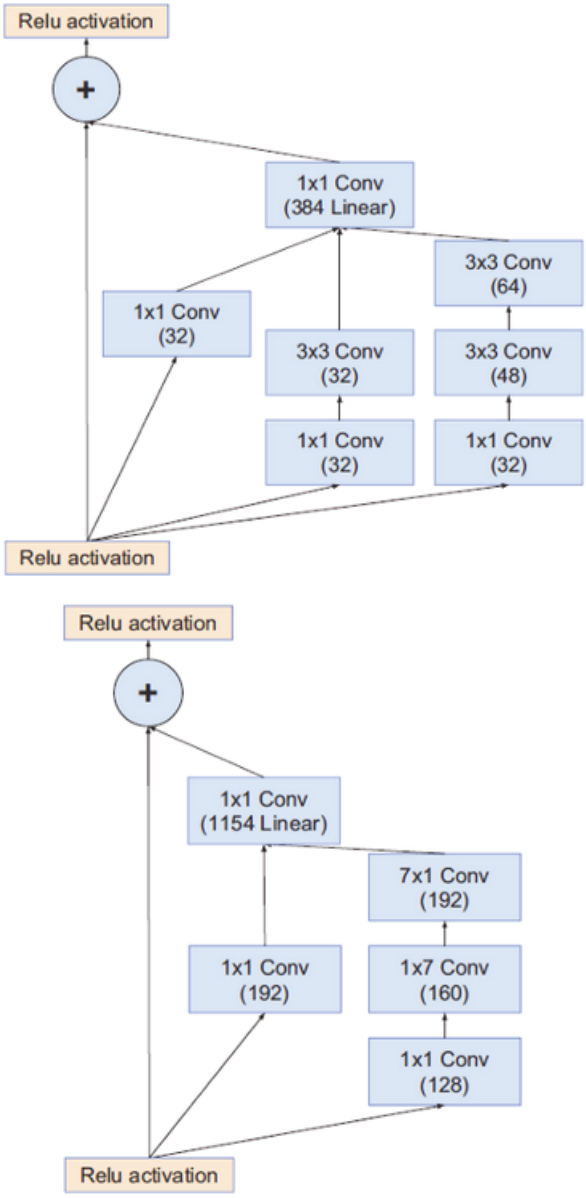


Figure 3.12: from Top to bottom: Architecture of Inception-ResNet-A and Inception-ResNet-B [4]

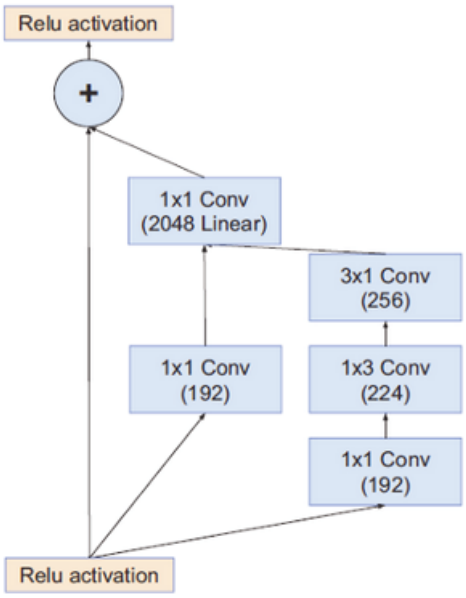


Figure 3.13: Architecture of Inception-ResNet-C [4]

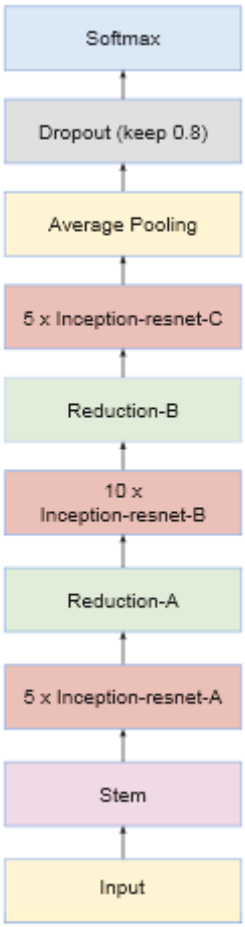


Figure 3.14: Schema for Inception-ResNet-v2 networks [4]

3.3.6 Size Comparison of the Used Pre-trained Models

A summary of the size comparison between the used pre-trained models are shown in Table 3.2. MobileNet has the smallest size among the models about $16MB$, such that once we use the model, running time is significantly reduced. However, the largest model size belongs to NASNetLarge model with $343MB$, having the higher computational time and fairly higher accuracy compared to other methods.

Table 3.2: Comparison of the used pre-trained models in terms of its sizes

Models	**Size MB**
VGG19	549
MobileNet	16
Inception-V3	92
NASNetLarge	343
IncetionResNet-V2	215

3.4 RSNA Dataset

Radiological Society of North America (RSNA) dataset was released by Stanford Children's Hospital and Colorado Children's Hospital and pediatric bone age challenge was organized in 2017, in which the radiographs have been taken in different conditions and times via various hardware.

Based on visual comparison to Greulich and Pyle's radiographic atlas of skeletal development of the hand and wrist, professional pediatric radiologists were able to interpret these images and then documented bone age in radiology reports. According to the reports, the skeleton ages were employed for training the model as the ground truth.

The RSNA dataset contains $12,611$ training dataset, which $5,778$ images are allocated for female radiographs and $6,833$ for male radiographs, $1,425$ images for validation, and 200 images for test images. Due to the small size of the test dataset and its unknown labels, we tested the model on 1000 radiographs from the training set, which is called validation set. The skeleton ages are ranging from 1 to 228 months, in which most subjects were 5-15 years old children. The examples of RSNA dataset are shown in Figure 3.15.

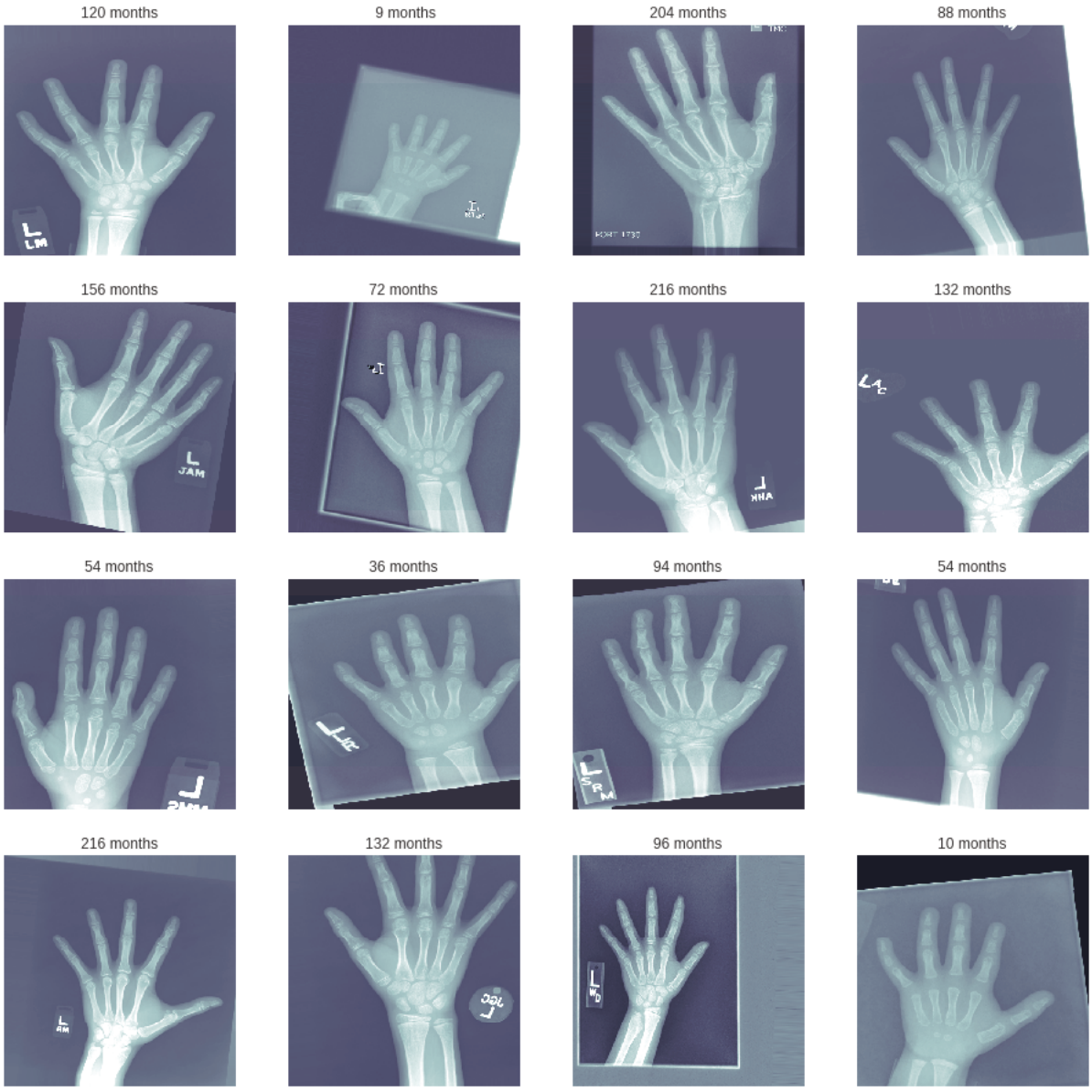


Figure 3.15: Examples of RSNA dataset

3.4.1 Proposed Neural Network Architectures

For training our model using the knowledge learned from a pre-trained network, various layers are taken into account in architecture, including batch normalization, learned weights from different pre-trained models, pooling layer (global average pooling2D), and dropout layer. In the following, we are going to explain these layers:

Batch normalization layer The layer normalizes the activation of the previous layer at each batch so that a transformation is applied to set the mean activation close to zero and the activation

standard deviation close to 1.

Global average pooling 2D Neural networks also contain pooling layers. Pooling layers are used to deliberately decrease the spatial size of the output representation, reducing the number of parameters and to control over-fitting of the model. In essence, the pooling layer helps simplify the spatial representation of the output volume from the preceding layer, followed by applying a function, such as max-pooling, average-pooling, or ℓ_2-norm pooling.

Dropout layer The aim of using dropout layer is to prevent over-fitting by reducing neurons. This layer is composed of a fraction rate of input units to be randomly set as zero at each update during training time. The fraction rate is a value between 0 and 1, which is a fraction of the input units to drop. In our study, we employed a drop rate of 0.5.

Dense layer A dense layer is a regular layer of neurons in a neural network, in which each neuron receives input from all the neurons in the previous layer, thus the neurons are densely connected.

3.5 ImageNet dataset

ImageNet dataset contains over 15 million labeled high-resolution images and around $22,000$ categories. ImageNet large scale visual recognition competition (ILSVRC) utilizes a subset of ImageNet of around 1000 images in each of 1000 categories, including roughly 1.2 million training images, $50,000$ validation images and $100,000$ testing images.

3.6 Transfer Learning

Transfer learning is a common approach in deep learning applications. The approach is promoted learning in new tasks by leveraging knowledge from a pre-trained network. We consider a pre-trained network and its corresponding trained weights as a starting point to learn a new task so that the learned features are transferred to a new task using a smaller number of training images. A network is usually fine-tuned by transfer learning much easier and faster than a network with randomly initialized weights from scratch.

The pre-trained models are powerful tool that enable us to generalize them for images outside of the ImageNet dataset. The most prominent reason for using transfer learning is the computational cost of running algorithms on hardware. Although, RAM is available in plenty and has an affordable price for every user, running a complex machine learning problem entails hundreds of GBs of RAM or GPUs that are expensive for most ML users.

3.7 Transfer Learning with Fine-tuning

The proposed pre-trained models have strong potential to be generalized to other existed images of the ImageNet dataset using transfer learning. For this purpose, fine-tuning is exploited by some modifications of the pre-existing models. There are three types of fine-tuning a model. The first approach employs a pre-trained model as a features extraction mechanism, and then utilizes the entire network as a fixed features extractor for a new dataset. The second way is to use the architecture of the pre-trained model, in which all the weights are initialized randomly to train the model according to a new dataset. The last way is to train some layers while freezing others. In this way, a part of a pre-trained model is used to keep the weights of initial layers of the model frozen while only higher layers are retrained. It is worth noting that the number of layers to be frozen or to be trained can be varied in fine-tuning.

3.7.1 Use of Different Pre-trained Models to Assess Bone Age

In this work, various architectures trained on ImageNet dataset are used. By taking advantage of the first way to fine-tune a network as discussed above, a fixed features extractor based on the type of a pre-trained model is applied.

Considering the low number of the training images in RSNA dataset ($12,611$ images) and low similarity between RSNA dataset and ImageNet dataset, we retrained the output of dense layers of the used pre-trained models as features extractors. These layers are used as the inputs of our proposed architectures. The features are trained according to the new dataset (RSNA) and then are fed to a dense layer as an output layer. More precisely, the output layer of a pre-trained model, e.g. VGG16 model, is a softmax activation function with 1000 categories for the classification task. However, the last layer is removed in our architectures and replaced by a dense layer to solve a regression problem and predict the age of hand bone for RSNA dataset.

3.7.2 Pre-processing on RSNA dataset

RSNA dataset contains $12,611$ X-ray images with portable network graphics (PNG) format. These images with the given information of the bone age, genders and labels were loaded into memory at running time. Afterward, the dataset is split into training and validation sets, in which the validation set includes 25% of the training set, i.e. $9,458$ images for training and $3,153$ for validation sets.

Since the number of images per each age category and gender was not equal, the distribution in the training set was adjusted to a new training size of $10,000$ instead of the old size of $9,458$, and the number of samples for each age was changed to $1,000$.

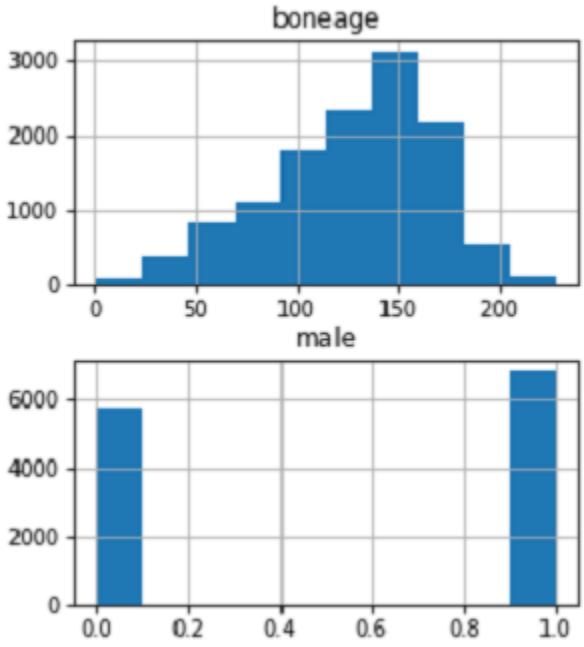


Figure 3.16: From top to bottom: the distribution of age and gender, respectively. Age is shown based on months.

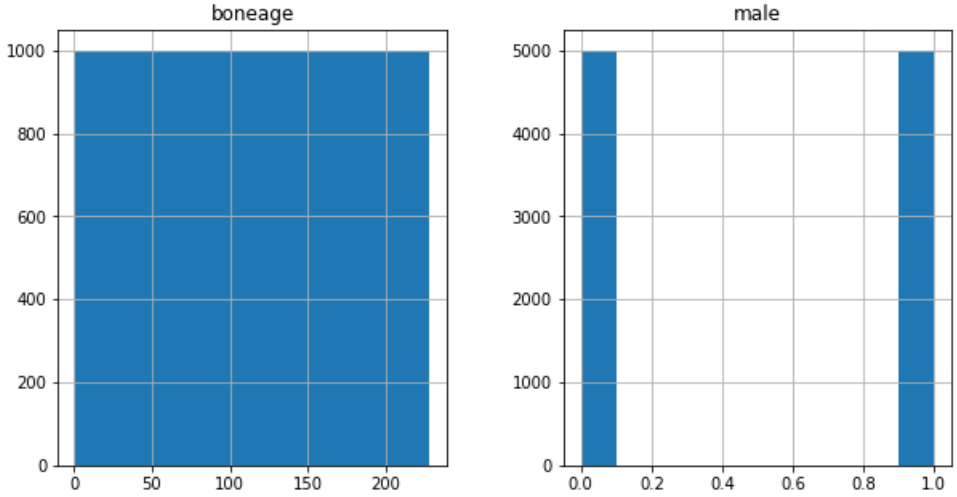


Figure 3.17: Balancing the distribution in the training set

To meet the right size of the input image in different pre-trained networks, each input image in RSNA is resized. For the used pre-trained models in this work, we consider 299×299 for InceptionResNet-V2, 331×331 for NasNetLarge, 331×331 for Inception-V3 , 224×224 for Mobilenet , 224×224 for VGG16, and 224×224 for VGG19 model.

Also, data augmentation is performed on the entire dataset via Keras ImageDataGenerator. In

the training process, we set up samplewise center for each sample to zero, samplewise standard deviation normalization to divide each input by its standard deviation, horizontal flip in which the input is randomly flipped horizontally, rotation range of 10 degrees, and zoom range of 15%. These values were chosen based on radiologist gestalt, which play important roles in the interpretation of other types of radiologic examinations [79], and is realistic variances from image to image in practice.

A real-time augmentation drastically improves learning process by multiplying the dataset, but each of the selected transformation such as zooming, rotation, and position in the dataset enhances generalizability. This ability of augmentation helps the model to learn features that are intrinsic for the patients rather than the imaging technique.

3.7.3 Training

In a bid to expedite the training process of the network, we used a smaller batch size with less memory, where the batch size defines the number of training examples in a forward or backward pass (an iteration). So, the training and validation sets are both buffered into batches of 32 items with greyscale color mode and the same size of the used pre-trained model input. In testing step, we used the validation set with a big batch size of 100 for evaluating the regression task.

For instance, given MobileNet-V2 as our pre-trained model, all the weights are initialized randomly to retrain our model on RSNA training set. The training time of our model took approximately 3 hours (fairly fast learning), in which MAE loss function was attempted to reach the minimum value using Adam optimizer when the predicted child's age is exactly equal to the true age of the child. Subsequently, the learning rate is reduced when the validation loss is plateaued.

In addition, we need to monitor the model throughout the training process for improving MAE as a model's metric. Therefore, a Keras callback function is specified for monitoring a quantity of validation loss (Val loss), that is the value of error after running the validation set through the trained network, and once no improvement is observed for a number of epochs (patience=5), the learning rate is decreased. It is worth noting that, when the learning process is stagnated, the learning rate will be reduced by a factor of 0.8 and then the value of cooldown is set, which is the number of epochs to wait before resuming normal operation after the learning rate has been reduced. We assumed a value of 0.0001 for setting the minimum learning rate as a lower bound on the learning rate.

Finally, the predicted results are stored in a $hdf5$ format. It should be noted that as the epochs increase, both validation and training loss (error) are dropped. If the training loss continues to drop, it means the network is learning the data better and better while increasing the validation loss

leads to over-fitting.

3.8 Loss Function

In our study we used mean squared error (MSE) as our loss function, that is a procedure for estimating an unobserved quantity, measuring the average of the squared errors. The MSE is defined as the average squared difference between the real values and the estimated value. The MSE is given as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2 \tag{3.1}$$

where, y_i and $\tilde{y}_i$ are ground truth label and predicted label for the ith training example, respectively. The actual optimized objective is the mean of the output array across all training examples n. The value of MSE is almost always strictly positive and none zero.

3.9 Linear Regression

Regression analysis is a statistical tool for the investigation of relationships between variables to identify, estimate and validate the relationship. The estimated relationship is then used to predict one variable from the value of the other variable(s).

Suppose that we have n pairs of examples (x_i, y_i). A preliminary step prior to statistical analysis of the relationship between the variables x and y is to draw a scatter plot, representing whether a linear relationship between dependent and independent variables exists. Variable x is known as predictor or independent variable while variable y is a response or dependent variable. The simplest type of regression analysis is linear regression. We assume that each observed value x_i of the response variable y_i can be described as follows:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad , i = 1, ..., n \tag{3.2}$$

where y_i is the dependent variable that we wish to predict, x_i is the independent variable used to predict. The regression coefficients β_0 and β_1 are unknown parameters, which also called intercept and slope of the line, respectively. ϵ_i is a random error with unknown variance. The term linear is used because equation 3.2 is a linear function of the unknown parameters β_0 and β_1.

3.10 Optimizer

The choice of optimization algorithm for a deep learning model plays a key role in speeding up the obtained results. Extensive stochastic gradient descent is the Adam optimization algorithm used for deep learning applications in computer vision. Since the optimizer fast approaches the solution, the algorithm is a great substitution of classical stochastic gradient descent procedure to update weights based on the training data.

Some of the attractive benefits of using Adam algorithm on non-convex optimization problems includes being computationally efficient, straightforward to implement, low memory requirements, well-suited for problems with large data and parameter, being invariant to diagonal rescale of the gradients, and suitable for problems with very noisy or sparse gradients.

The Adam optimizer is different from the classic Stochastic gradient descent. Moreover, the Adam algorithm is the combination of two other extensive stochastic gradient descent, including adaptive gradient algorithm (AdaGrad) and root mean square propagation (RMSProp), which benefits from both aforementioned algorithms. AdaGrad maintains a per-parameter learning rate so that performances on problems with sparse gradients are improved while RMSProp retains per-parameter learning rate is adapted based on the average of recent magnitudes of the gradients for the weight. Therefore, the algorithm has a great performance in online and non-stationary problems such as noise. Furthermore, Adam optimizer adapts the learning rate parameters based on the average first moment (the mean) as in RMSProp, and also advantages of the average of the second moments of the gradients (the uncentered variance). Specifically, an exponential moving average of the gradient and the squared gradient are obtained using Adam. The parameters β_1 and β_2 control the decay rates of these moving averages. If the recommended initial values of the moving averages, β_1 and β_2 are close to 1.0, a bias of moment approaches toward zero.

3.11 Regression metrics

3.11.1 Mean absolute error

Mean absolute error (MAE) is a metric to measure the difference between two continuous variables. Variables X and Y as paired observations denote observed and predicted examples, respectively. Considering a scatter plot with n points, where each point i has a coordinates (x_i, y_i) points, MAE is vertical distance between each point and identity line. The MAE is calculated by:

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} = \frac{\sum_{i=1}^{n} |e_i|}{n} \tag{3.3}$$

3.12 Implementation

In this section, extensive experiments have been conducted through different architectures to train the proposed models using different pre-trained models on RSNA dataset. Finally, we analyzed the result of each architecture on the regression task.

To implement our deep learning framework, we exploited Keras library with Tensorflow, running as a backend. All the used codes were written in Python 3.7 along with some scientific libraries such as Numpy and Pandas. We also utilized GPU, i.e. Tesla K80 GPU, provided by Google's free cloud service to speed up the computation time. It is worth noting that to use the GPU capability, Google Colab is available for all the users that can apply Keras, Tensorflow and Pytorch libraries.

All the used pre-trained models are available in Keras framework, which allows us to extract features using Keras library.

After the basic pre-processing steps on RSNA dataset, a simple MLP model (multilayer perceptron) was proposed with the following architecture in Table 3.3. The architecture consists of six hidden layers, including batch normalization, a pre-trained model weights, another batch normalization layer, global average pooling 2D, dropout layer and finally a dense layer. Also, the network architectures and parameters are depicted in Table 3.3, in which six different models (architectures) are shown in terms of number and type of the used layers, output of each layer and number of parameters in each layer. In first layer, the input image layer varies for the various used architectures, which the selected input size for each model was considered as the input size of the used pre-trained model, which means that 299×299 for InceptionResNet-V2 model, 331×331 for NasNetLarge, 331×331 for Inception-V3 , 224×224 for Mobilenet , 224×224 for VGG16 and 224×224 for VGG19 model, respectively. In this layer, we have used batch normalization, in which the layer normalizes an input layer (i.e. $\mu = 0$, $\sigma = 1$).

The output shape of the first layer has the same size of input layer with 1 channel because we used greyscale images. A None dimension in a shape tuple means that the network will be able to accept inputs of any dimension, i.e. no pre-defined number. Therefore, we could assign different batch sizes during training the network. In this layer, we just obtained four trained parameters for each used model.

In layer 2, we added a pre-trained model architecture and took the output shape of our first layer as an input to feed the second layer. Finally, we trained our model with a minibatch size of 16 for 10 epochs. For instance, given MobileNet-V2 as our pre-trained model, all the weights are initialized randomly to retrain our model on RSNA training set and MAE is used as our loss function. Our goal is to find the point that minimizes Adam optimizer implemented in Keras with default values of 0.001 for learning rate, 0.9 for β_1 and 0.999 for β_2. The training of MobileNet-V2

model architecture took approximately 3 hours that is fairly fast enough.

Then, the output shape of the second layer using VGG16 and VGG19 architectures would be 7×7 with 512 units (neurons), where the number of trainable parameters in VGG19 was greater than VGG16 with $14,713,536$. For Inception-V3 and InceptionResNet-V2 models, the size of the output is 9×9, with $2,048$ and 1536 units, respectively. Also, the number of trainable parameters of InceptionResNet-V2 are 2.5 times greater than Inception-V3 parameters. For the output of NasLargeNet and Mobilenet models, we had a 11×11 with $4,032$ units and 16×16 with $1,024$ units, respectively. In this layer, MobileNet-V2 architecture had a smaller number of trained parameters while NasLargeNet model was assigned the larger trained parameters.

In the next layer, we again applied a batch normalization, in which NasNetLarge with the size of 11×11 and $4,032$ units, was gained the higher number of parameters about $16,128$. Furthermore, for both VGG16 and VGG19 we observed the lower number of parameters 7×7 with 512 units as the output of the third layer, and the layer created $2,048$ trained parameters.

Afterward, global average pooling was used on the output of previous layer to optimize the performance and prevent over-fitting. In this layer, the number of units in our model using NasNet-Large architecture was the largest value among the other architectures about $4,032$. Also, for our models using VGG16 and VGG19 architecture, the number of units was equal to 512 that is considered the smallest units between other architectures. Notably, none of the parameters (weights) were trained in this layer and even in the next layer, which is dropout. In the fifth layer, regularization of the neural network was done using dropout, which randomly zeros out the interdependent neurons in training time. The fraction rate of dropout was chosen 0.5, which means the neurons were ignored by a random fraction of 0.5 for each training sample and iteration. Finally, a dense layer was applied in the last hidden layer with no activation function. The output layer had only one neuron, indicating the predicated bone age for each input image.

Table 3.3: Architectures of six models created from various pre-trained models weights for training network in regression task: (a)Architecture used weights of VGG16, (b) VGG19, (c)MobileNet , (d)Inception-V3 , (e)NasNetLarge and (f)InceptionResNet-V2

# layers	Models	Layer(type)	Output shape	# Params
1	(a)	Batchnormalization(Batch)	(None, 224, 224, 1)	4
	(b)	Batchnormalization(Batch)	(None, 224, 224, 1)	4
	(c)	Batchnormalization(Batch)	(None, 224, 224, 1)	4
	(d)	Batchnormalization(Batch)	(None, 331, 331, 1)	4
	(e)	Batchnormalization(Batch)	(None, 331, 331, 1)	4
	(f)	Batchnormalization(Batch)	(None, 229, 299, 1)	4
2	(a)	VGG16(Model)	(None, 7, 7, 512)	14, 713, 536
	(b)	VGG19(Model)	(None, 7, 7, 512)	20, 024, 384
	(c)	MobileNet(Model)	(None, 16, 16, 1024)	3, 228, 288
	(d)	Inception − V3(Model)	(None, 9, 9, 2048)	21, 802, 784
	(e)	NasLargeNet(Model)	(None, 11, 11, 4032)	84, 915, 090
	(f)	InceptionResNet − V2(Model)	(None, 9, 9, 1536)	54, 336, 160
3	(a)	Batchnormalization(Batch)	(None, 7, 7, 512)	2, 048
	(b)	Batchnormalization(Batch)	(None, 7, 7, 512)	2, 048
	(c)	Batchnormalization(Batch)	(None, 16, 16, 1024)	4, 096
	(d)	Batchnormalization(Batch)	(None, 9, 9, 2048)	8, 192
	(e)	Batchnormalization(Batch)	(None, 11, 11, 4032)	16, 128
	(f)	Batchnormalization(Batch)	(None, 9, 9, 1536)	6, 144
4	(a)	Globalaveragepooling2D	(None, 512)	0
	(b)	Globalaveragepooling2D	(None, 512)	0
	(c)	Globalaveragepooling2D	(None, 1024)	0
	(d)	Globalaveragepooling2D	(None, 2048)	0
	(e)	Globalaveragepooling2D	(None, 4032)	0
	(f)	Globalaveragepooling2D	(None, 1536)	0
5	(a)	Dropout(Dropout)	(None, 512)	0
	(b)	Dropout(Dropout)	(None, 512)	0
	(c)	Dropout(Dropout)	(None, 1024)	0
	(d)	Dropout(Dropout)	(None, 2048)	0
	(e)	Dropout(Dropout)	(None, 4032)	0
	(f)	Dropout(Dropout)	(None, 1536)	0
6	(a)	Dense(Dense)	(None, 1)	513
	(b)	Dense(Dense)	(None, 1)	513
	(c)	Dense(Dense)	(None, 1)	1025
	(d)	Dense(Dense)	(None, 1)	2049
	(e)	Dense(Dense)	(None, 1)	4033
	(f)	Dense(Dense)	(None, 1)	1537

After training our models according to the models' summaries shown in Table 3.3, we trained our models as discussed in section 3.7.3 by running them in 10 epochs and 331 steps per epochs. We evaluated the models on the validation set of 1000 X-ray images of hand bones.We used various architectures and compared them using MAE metric in our experimental results. As Table 3.4 shows the model with InceptionResNet-V2 weights achieved MAE of 14.28 months on the validation set and chosen as the best predictive model. This high accuracy can be due to the deeper architecture of this pre-trained model (164 hidden layers) than the others, i.e. MobileNet-V2 with 53 layers, Inception-V3 with 48 layers, VGG19 and VGG16 with 19 and 16 layers, respectively. The performance of the used MobileNet-V2, NasNetLarge and Inception-V3 weights achieved substantially better results. Moreover, the model based on VGG16 and VGG19 weights resulted in the worst performance of MAE, with 20.89 and 19.94 months, respectively.

Table 3.4: Comparison of various models (a) Architecture used VGG16 weights, (b) VGG19 weights, (c) MobileNet-V2 weights, (d) Inception-V3 weights, (e) NasNetLarge weights and (f) InceptionResNet-V2 weights on X-ray images of hand bones for both genders in terms of mean absolute error.

Models	**MAE-months**
(a)VGG16	**20.86**
(b)VGG19	**19.94**
(c)MobileNet − V2	**17.49**
(d)Inception − V3	**18.40**
(e)NasNetLarge	**18.47**
(f)InceptionResNet − V2	**14.28**

In next step, we demonstrate the evaluation of our trained models for regression task in Figure 3.18. As can be seen, linear regression models are represented on six different proposed models, where red dots are the samples of the validation set on the graph, and each point has a x-y coordinate. Each prediction line is shown with a blue line, which is a line passing through all the points and fits them in the best way.

By drawing a scatter plot, MAE can be visualized by the vertical distance between each point and its prediction (identity) line. The plot aims to find a line that minimizes the squared distances among the points to reveal the relationship between actual ages $(\mathbf{x})$ and predicted ages $(\mathbf{y})$. In Figure 3.18, linear regression models are depicted for the proposed six architectures. As can be seen, the best MAE results belong to the proposed InceptionResNet-V2 and MobileNet-V2, respectively, while we observed low density of the points closed to the identity line for the VGG16

and VGG19 weights.

In addition, several validation samples (radiographs) were randomly chosen from the whole validation set for each architecture to visualize the predicted and the real bone ages (Figures 3.19 to 3.24).

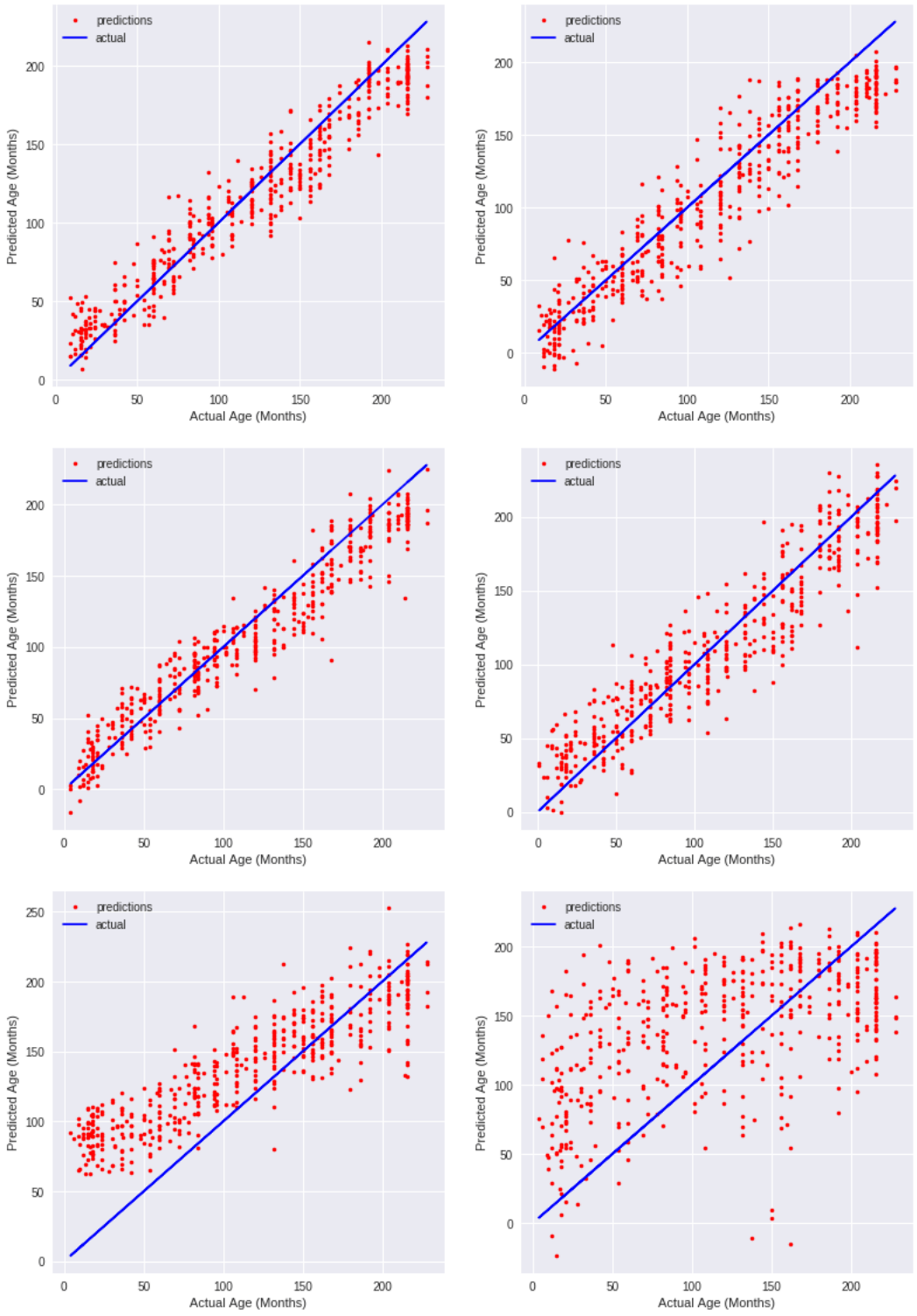


Figure 3.18: Linear regression models for the used various architectures (from left to right, top to bottom: InceptionResNet-V2, NasNetLarge, Inception-V3, Mobilenet, VGG19 and VGG16 weights).

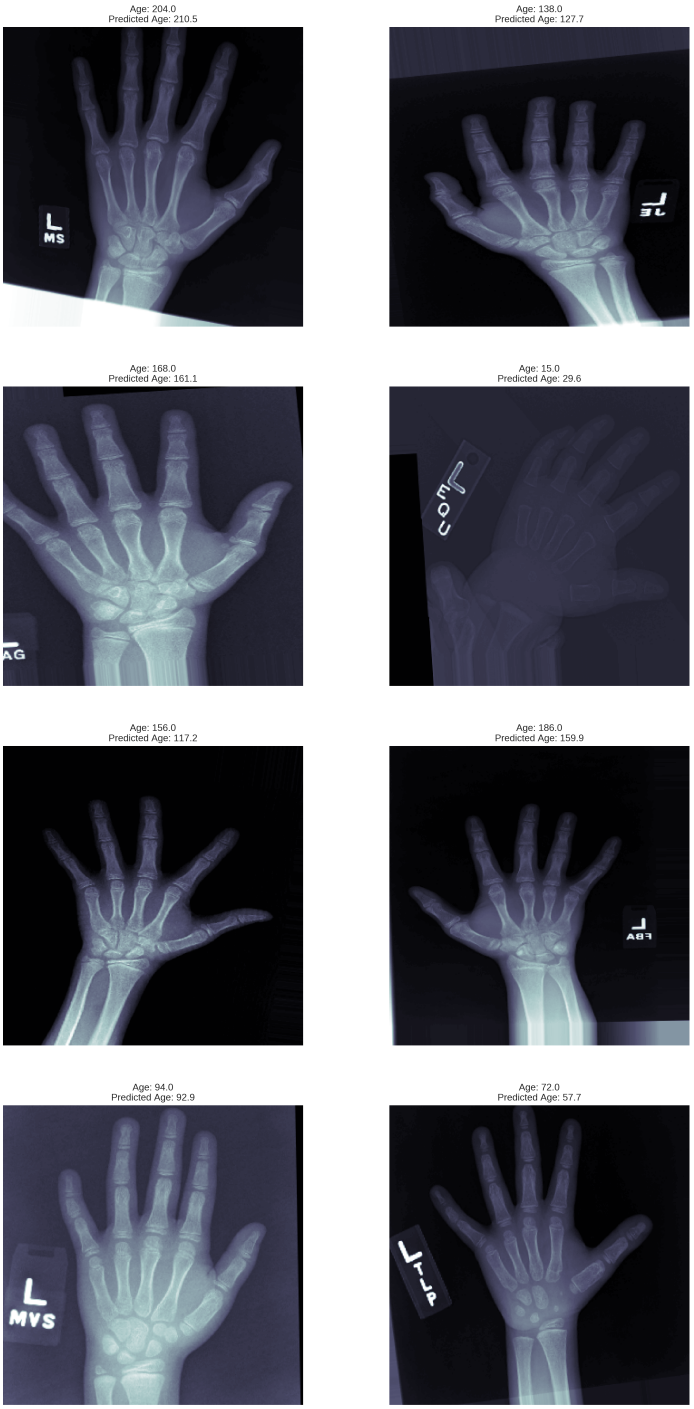


Figure 3.19: Examples of comparison predicted bone ages with its real bone ages of the radiographs by adopting InceptionResNet-V2 weights on validation set.

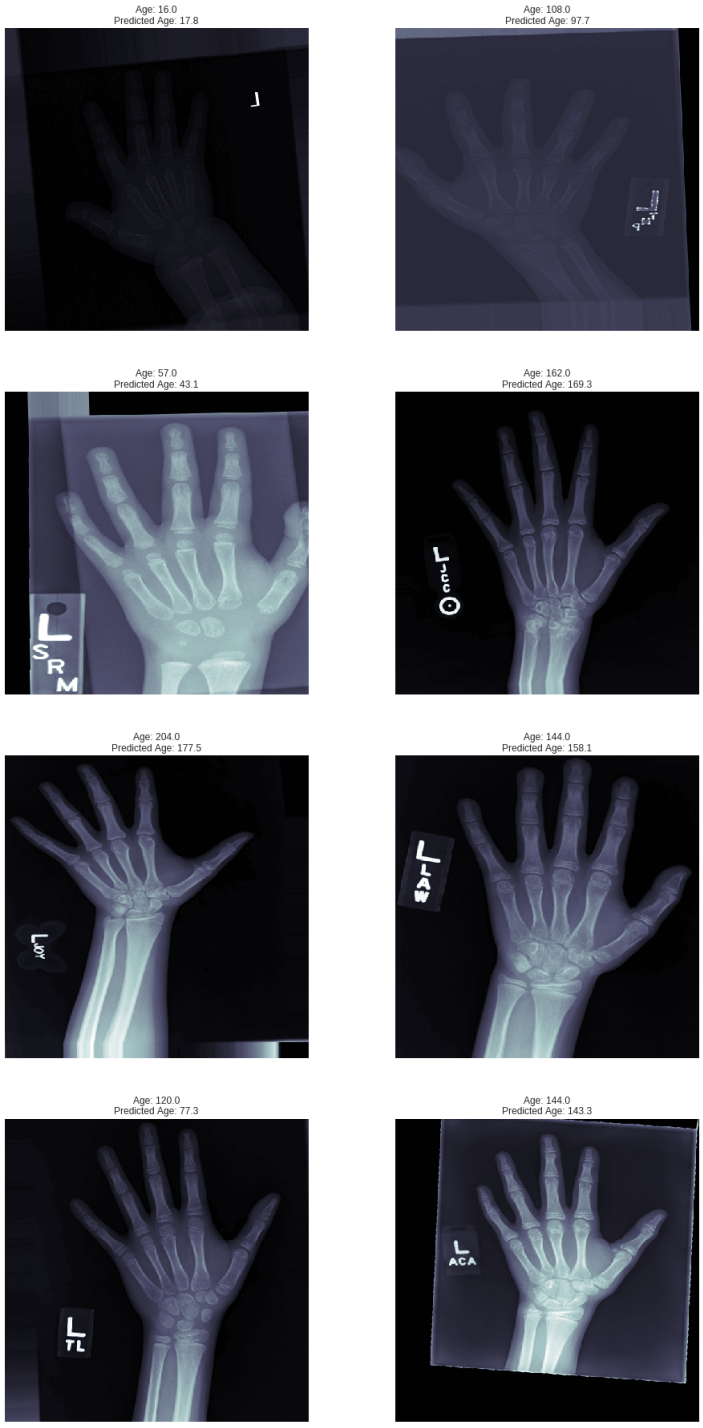


Figure 3.20: Examples of comparison predicted bone ages with its real bone ages of the radiographs by adopting NasNetLarge weights on validation set.

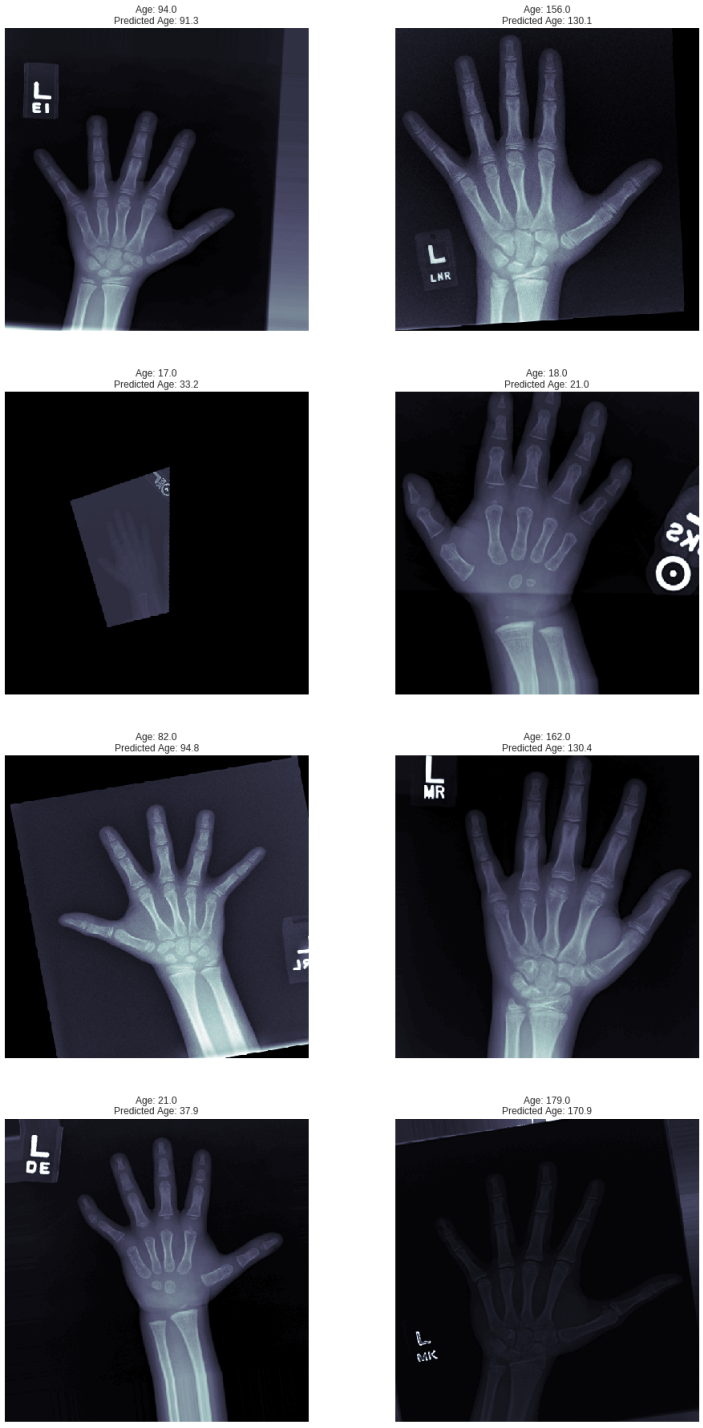


Figure 3.21: Examples of comparison predicted bone ages with its real bone ages of the radiographs by adopting Inception-V3 weights on validation set.

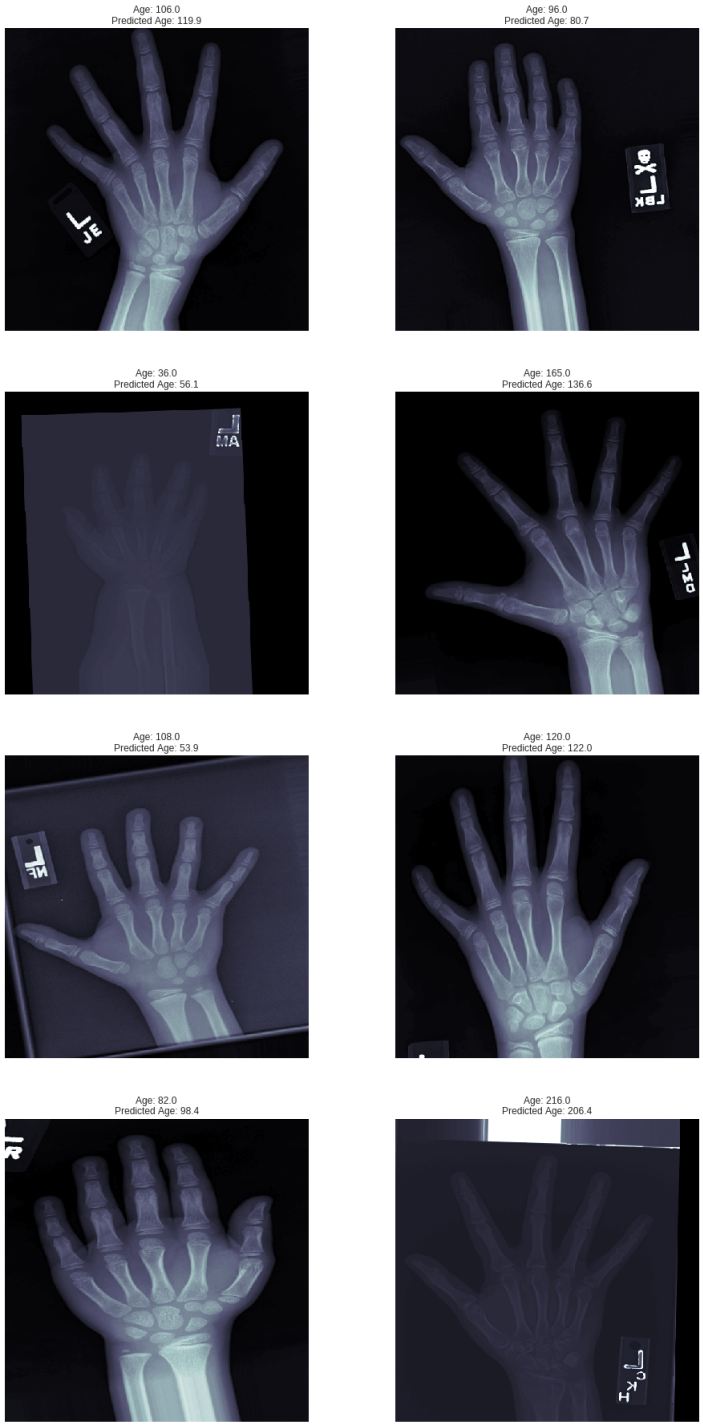


Figure 3.22: Examples of comparison predicted bone ages with its real bone ages of the radiographs by adopting MobileNet weights on validation set.

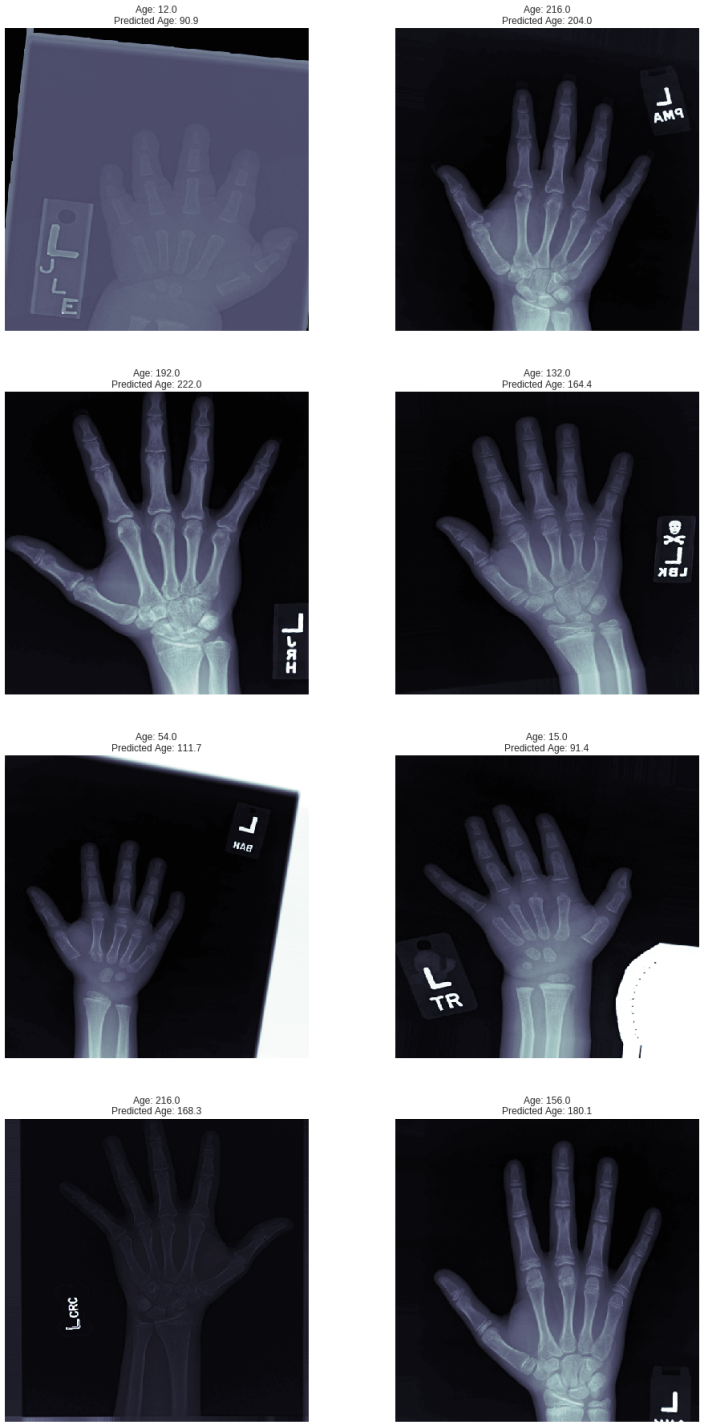


Figure 3.23: Examples of comparison predicted bone ages with its real bone ages of the radiographs by adopting VGG19 weights on validation set.

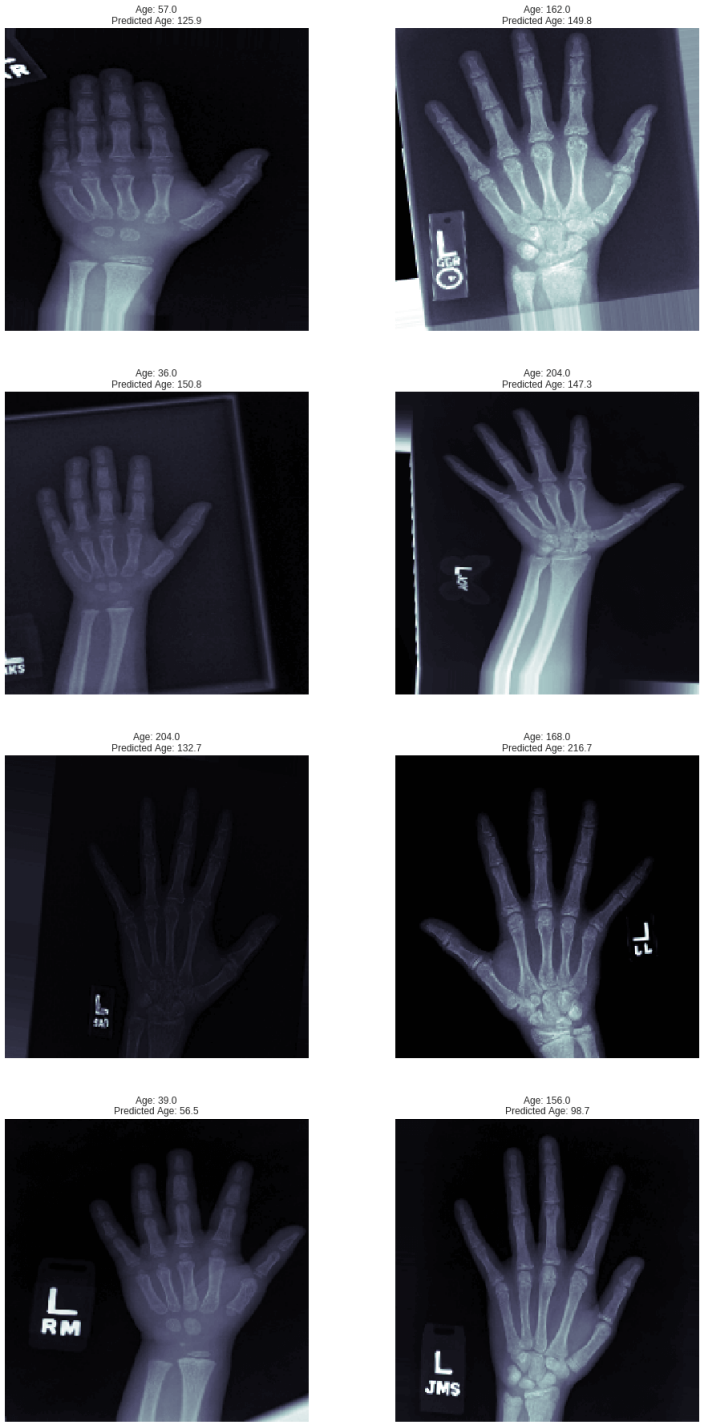


Figure 3.24: Examples of comparison predicted bone ages with its real bone ages of the radiographs by adopting VGG16 weights on validation set.

3.13 Conclusion

In this chapter, we introduced a computerized method for BAA based on human wrist bones X-ray images. The method was developed using state-of-the-art deep learning algorithms. The proposed BAA algorithm allows radiologists to predict the bone age more accurate by eliminating human observation variations, leading to cost reduction in the hospitals or clinics. In addition, the presented approach facilitates the BAA by providing a faster process than manual methods. The algorithm was evaluated through MAE metric for six models, which created from various pre-trained models weights. Among the proposed deep learning models, the result of the InceptionResNet-V2 architecture indicates more accurate result as an automated method for estimating the bone age.

CHAPTER 4

Conclusions and Future Work

This thesis has presented an efficient global spectral graph wavelet to analyze 3D carpal bones shapes. We have demonstrated the superior results of the proposed GSGW through extensive experiments on different 3D shapes of wrist bones compared to the existing method in the literature. In addition, we investigated an automated BAA algorithm to perform a regression task by applying state-of-the-art pre-trained models' weights in our architectures that have a significant impact on the performance of a hand bone age prediction. In Section 4.1, we discuss the contributions of the previous chapters and then state the limitations and future work of this research in Section 4.2.

4.1 Contributions of the Thesis

In chapter 2, using the eigensystem of Laplace-Beltrami operator and spectral graph wavelets, we introduced a global descriptor that is able to decompose a carpal bone surface into its constituent intrinsic components and represents each surface by a global spectral graph wavelet descriptor. This global descriptor enjoys several desirable properties including simplicity, compactness, invariance to isometric deformations, and computational feasibility. Thereby, this helps facilitate the statistical analysis of wrist bones. Moreover, our method needs only a small number of eigenfunctions to identify the variations in the bones' surfaces, and can easily distinguish not only between carpal bones of different groups (female, male) but also bones from the same group. This property tremendously simplifies the problem and speeds-up the computation of the shape descriptor, particularly when dealing with shapes consisting of thousands of vertices, albeit increasing the number of mesh vertices tends to provide slightly better results. Moreover, the experimental results on the

carpal bone dataset demonstrate the effectiveness of the proposed GSGW framework.

In chapter 3, we developed an efficient method for evaluation of BAA to address some disadvantages of manual BAA approach such as being subjective of bone age estimation and time-consuming process. The presented methods are automated bone age techniques by applying the state-of-the-art deep learning architectures (pre-trained models' weights). So, the high-level features are extracted from X-ray images providing fairly accurate prediction of real bone age. Our system enjoys some nice advantages such as being cost-saving as no radiologists and doctors are required to perform the BAA, meaning that there is no chance for observation variability by radiologists. Moreover, the automated BAA is a time-saving process as the predicted age is obtained within a few minutes.

Finally, to give more validity to our results, statistical analysis, i.e. regression analysis, was performed during the evaluation stage of the methods. It should be noted that the presented BAA methods achieved fairly good results for MAE, where the best result was 14.28 months for the model built from IncetionResNet-V2 weights among the other pre-trained models' weights reviewed in this thesis.

4.2 Limitations and Future Work

A number of limitations and interesting future directions are discussed below based on this thesis.

4.2.1 Surface Analysis in Carpal Bones

In chapter 2, while the carpal bone dataset used in this study consists of a small size of individuals, we plan in the not-too-distant future to test the GSGW approach on a larger population. The used carpal bone dataset includes subjects who are healthy, young, and who lacked remarkable pathological findings. So, it would be interesting to investigate surface differences in wrist bones when, for instance, a group is suffering from specific pathologies like osteoarthritis or wrist instability. Moreover, this dataset does not include information related to the dominant hand of the individuals. Hence, we may consider this information to analyze the bone shape differences in an effort to find the probable association. Since men and women in the carpal bone dataset are in the same age range, the factors related to age can also be investigated by considering a different age range in two groups of men and women. In addition, the other factors that may change the carpal bone shapes of individuals include the body size, genetic, metabolic and environmental factors can be regarded, which are not taken into consideration in our experiments because they are not available

in the carpal bone dataset. For future work, we also plan to apply the proposed framework to other shape analysis problems, and in particular segmentation and clustering tasks.

4.2.2 Bone Age Assessment

Apparently, a number of challenges for a computerized assessment of bone age still exist among the available automated systems as discussed in Chapter 3. The main challenge is the lack of research on the role of other bones in the body and their contributions to automated assessment of bone age. In fact, the problem is still ambiguous in a case that a hand image of a person is unavailable for any reason, whether the information given by his/her body bones can be reliable for BAA system or not. Thus, we plan to investigate hybrid evidence, i.e. radiograph of the left hand, physical and teeth examination, to further improve the accuracy and a standard method for BAA system [80].

In addition, lack of sufficient image processing techniques has led to low accuracy for BAA. Regardless of the used pre-processing techniques (resizing and rotating) on the dataset, other techniques such as extracting region of interest (ROI) and image segmentation can be used to minimize the BAA error rate. Due to the lack of definition of segmentation on radiographs and unique standard methods for the implementation, image segmentation is still challenging for the current automated system [81]. We also plan to explore the use of segmentation, which is defined to separate the specific region of an X-ray hand. Various methods use different regions of X-ray images such as the carpal bones, phalangeal, ulna epiphyses, and the radius [82]. We will then perform image segmentation based on attributes of the X-ray image such as bone texture and intensity, and consequently implement it by a bone region or a contour of bone edge.

References

[1] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *arXiv preprint arXiv:1409.1556*, 2014.

[2] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *CVPR*, pp. 2818–2826, 2016.

[3] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetv2: Inverted residuals and linear bottlenecks," in *CVPR*, pp. 4510–4520, 2018.

[4] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Thirty-First AAAI Conf. on Artificial Intelligence*, 2017.

[5] A. M. Zafar, N. Nadeem, Y. Husen, and M. N. Ahmad, "An appraisal of greulich-pyle atlas for skeletal age assessment in pakistan," *The Journal of the Pakistan Medical Association*, vol. 53, no. 7, p. 552, 2010.

[6] C. M. Gaskin, S. L. Kahn, J. C. Bertozzi, and P. M. Bunch, *Skeletal Development of the Hand and Wrist: A Radiographic Atlas and Digital Bone Age Companion: A Radiographic Atlas and Digital Bone Age Companion*. Oxford University Press, 2011.

[7] K. Khan and A. S. Elayappen, "Bone growth estimation using radiology (greulich pyle and tanner whitehouse methods)," *Handbook of Growth and Growth Monitoring in Health and Disease*, pp. 2937–2953, 2012.

[8] M. Mansourvar, *Bone age assessment using hand and clavicle x-ray images*. PhD thesis, University Malaya, 2014.

[9] M. Masoumi, M. Rezaei, and A. B. Hamza, "Global spectral graph wavelet signature for surface analysis of carpal bones," *Physics in medicine and biology*, vol. 63, no. 10, p. 035034, 2018.

[10] M. Masoumi, M. Rezaei, and A. B. Hamza, "Shape analysis of carpal bones using spectral graph wavelets," in *Vertex-Frequency Analysis of Graph Signals*, pp. 419–436, Springer, 2019.

[11] J. Crisco, J. Coburn, D. Moore, and M. Upal, "Carpal bone size and scaling in men versus in women," *The Journal of Hand Surgery*, vol. 30, no. 1, pp. 35–42, 2005.

[12] D. Moore, J. Crisco, T, and E. Leventhal, "A digital database of wrist bone anatomy and carpal kinematics," *Journal of biomechanics*, vol. 40, no. 11, pp. 2537–2542, 2007.

[13] J. Xu, J. Wu, B. Notghi, M. Simon, and J. Brigham, "A feasibility study on kinematic feature extraction from the human interventricular septum toward hypertension classification," in *Proc. Int. Symp. Computational Modeling of Objects Represented in Images*, pp. 36–47, 2014.

[14] J. Wu, Y. Wang, M. Simon, M. Sacks, and J. Brigham, "A new computational framework for anatomically consistent 3D statistical shape analysis with clinical imaging applications," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 1, no. 1, pp. 13–27, 2013.

[15] M. Reuter, F. Wolter, and N. Peinecke, "Laplace-Beltrami spectra as 'Shape-DNA' of surfaces and solids," *Computer-Aided Design*, vol. 38, no. 4, pp. 342–366, 2006.

[16] R. Rustamov, "Laplace-Beltrami eigenfunctions for deformation invariant shape representation," in *Proc. Symp. Geometry Processing*, pp. 225–233, 2007.

[17] A. Bronstein, M. Bronstein, L. Guibas, and M. Ovsjanikov, "Shape Google: Geometric words and expressions for invariant shape retrieval," *ACM Trans. Graphics*, vol. 30, no. 1, 2011.

[18] S. Biasotti, A. Cerri, M. Abdelrahman, M. Aono, A. Ben Hamza, M. El-Melegy, A. Farag, V. Garro, A. Giachetti, D. Giorgi, A. Godil, C. Li, Y.-J. Liu, H. Martono, C. Sanada, A. Tatsuma, S. Velasco-Forero, and C.-X. Xu, "SHREC'14 track: Retrieval and classification on textured 3D models," in *Proc. Eurographics Workshop on 3D Object Retrieval*, pp. 111–120, 2014.

[19] M. Masoumi, C. Li, and A. Ben Hamza, "A spectral graph wavelet approach for nonrigid 3D shape retrieval," *Pattern Recognition Letters*, vol. 83, pp. 339–348, 2016.

[20] E. Rodola, L. Cosmo, O.Litany, M. M. Bronstein, A. M. Bronstein, N. Audebert, A. Ben Hamza, A. Boulch, U. Castellani, M. N. Do, A.-D. Duong, T. Furuya, A. Gasparetto, Y. Hong, J. Kim, B. L. Saux, R. Litman, M. Masoumi, G. Minello, H.-D. Nguyen, V.-T.

Nguyen, R. Ohbuchi, V.-K. Pham, T. V. Phan, M. Rezaei, A. Torsello, M.-T. Tran, Q.-T. Tran, B. Truong, L. Wan, and C. Zou11, "SHREC'17 track: Deformable shape retrieval with missing parts," in *Proc. Eurographics Workshop on 3D Object Retrieval 2017*, pp. 1–9, 2017.

[21] M. Masoumi and A. Ben Hamza, "Spectral shape classification: A deep learning approach," *Journal of Visual Communication and Image Representation*, vol. 43, pp. 198–211, 2017.

[22] E. Elsheh and A. Ben Hamza, "Secret sharing approaches for 3D object encryption," *Expert Systems with Applications*, vol. 38, no. 11, pp. 13906–13911, 2011.

[23] A. Chaudhari, R. Leahy, B. Wise, N. Lane, R. Badawi, and A. Joshi, "Global point signature for shape analysis of carpal bones," *Physics in Medicine and Biology*, vol. 59, pp. 961–973, 2014.

[24] Z. Gao, Z. Yu, and X. Pang, "A compact shape descriptor for triangular surface meshes," *Computer-Aided Design*, vol. 53, pp. 62–69, 2014.

[25] J. Sun, M. Ovsjanikov, and L. Guibas, "A concise and provably informative multi-scale signature based on heat diffusion," *Computer Graphics Forum*, vol. 28, no. 5, pp. 1383–1392, 2009.

[26] M. Aubry, U. Schlickewei, and D. Cremers, "The wave kernel signature: A quantum mechanical approach to shape analysis," in *Proc. Computational Methods for the Innovative Design of Electrical Devices*, pp. 1626–1633, 2011.

[27] S. Bu, Z. Liu, J. Han, J. Wu, and R. Ji, "Learning high-level feature by deep belief networks for 3-D model retrieval and recognition," *IEEE Trans. Multimedia*, vol. 24, no. 16, pp. 2154–2167, 2014.

[28] D. Shuman, B. Ricaud, and P. Vandergheynst, "Vertex-frequency analysis on graphs," *Applied and Computational Harmonic Analysis*, vol. 40, no. 2, pp. 260–291, 2016.

[29] L. Stankovic, E. Sejdic, and M. Dakovic, "Vertex-frequency energy distributions," *IEEE Signal Processing Letters*, 2017.

[30] R. Coifman and S. Lafon, "Diffusion maps," *Applied and Computational Harmonic Analysis*, vol. 21, no. 1, pp. 5–30, 2006.

[31] D. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129–150, 2011.

[32] C. Li and A. Ben Hamza, "A multiresolution descriptor for deformable 3D shape retrieval," *The Visual Computer*, vol. 29, pp. 513–524, 2013.

[33] N. Zaproudina, O. O. P. Hanninen, and O. Airaksinen, "Effectiveness of traditional bone setting in chronic neck pain: randomized clinical trial," *Journal of Manipulative and Physiological Therapeutics*, vol. 30, no. 6, pp. 432–437, 2007.

[34] M. A. Martin-Fernandez, M. Martin-Fernandez, and C. Alberola-Lopez, "Automatic bone age assessment: a registration approach," in *Medical Imaging 2003: Image Processing*, pp. 1765–1776, 2003.

[35] A. M. Zafar, N. Nadeem, Y. H. Y, and M. N. Ahmad, "An appraisal of greulich-pyle atlas for skeletal age assessment in pakistan," *Journal of the Pakistan Medical Association*, vol. 60, no. 7, p. 552, 2010.

[36] C. M. Gaskin, S. L. Kahn, J. C. Bertozzi, and P. M. Bunch, *Skeletal development of the hand and wrist: a radiographic atlas and digital bone age companion*. Oxford University Press, 2011.

[37] P. Kaplowitz, S. Srinivasan, J. He, R. McCarter, M. R. Hayeri, and R. Sze, "Comparison of bone age readings by pediatric endocrinologists and pediatric radiologists using two bone age atlases," *Pediatric radiology*, vol. 41, no. 6, pp. 690–693, 2011.

[38] B. H. Adler, "Vicente gilsanz, osman ratib: Bone age atlas," vol. 35, no. 10, pp. 1035–1050, 2005.

[39] A. Albanese, C. Hall, and R. Stanhope, "The use of a computerized method of bone age assessment in clinical practice," *Hormone Research in Paediatrics*, vol. 44, no. 3, pp. 2–7, 1995.

[40] E. Pietka, S. Pospiech, A. Gertych, F. Cao, H. K. Huang, and V. Gilsanz, "Computer automated approach to the extraction of epiphyseal regions in hand radiographs," *Journal of Digital Imaging*, vol. 14, no. 4, pp. 165–172, 2001.

[41] E. Pietka, A. Gertych, S. P. E. Kurkowska, F. Cao, H. K. Huang, and V. Gilzanz, "Computer-assisted bone age assessment: graphical user interface for image processing and comparison," *Journal of Digital Imaging*, vol. 17, no. 3, pp. 175–188, 2004.

[42] E. D. Ubeyli, "Comparison of different classification algorithms in clinical decision-making," *Expert Systems*, vol. 24, no. 1, pp. 17–31, 2007.

[43] M. Yildiz, A. Guvenis, E. Guven, D. Talat, and M. Haktan, "Implementation and statistical evaluation of a web-based software for bone age assessment," *Journal of Medical Systems*, vol. 35, no. 6, pp. 1485–1489, 2011.

[44] D. J. Michael and A. C. Nelson, "Handx: a model-based system for automatic segmentation of bones from digital hand radiographs," *IEEE Trans. on Medical Imaging*, vol. 8, no. 1, pp. 64–69, 1989.

[45] E. Pietka, M. F. McNitt-Gray, M. L. Kuo, and H. K. Huang, "Computer-assisted phalangeal analysis in skeletal age assessment," *IEEE Trans. on Medical Imaging*, vol. 10, no. 4, pp. 616–620, 1991.

[46] S. M. Garn, K. P. Hertzog, A. K. Poznanski, and J. M. Nagy, "Metacarpophalangeal length in the evaluation of skeletal malformation," *Radiology*, vol. 105, no. 2, pp. 375–381, 1972.

[47] E. Pietka, "Computer-assisted bone age assessment based on features automatically extracted froma hand radiograph," *Computerized Medical Imaging and Graphics*, vol. 19, no. 3, pp. 251–259, 1991.

[48] C. W. Hsieh, T. L. Jong, and C. M. Tiu, "Bone age estimation based on phalanx information with fuzzy constrain of carpals," *Medical and Biological Engineering and Computing*, vol. 45, no. 3, pp. 283–295, 2007.

[49] J. Liu, J. Qi, Z. Liu, Q. Ning, and X. Luo, "Automatic bone age assessment based on intelligent algorithms and comparison with tw3method," *Computerized Medical Imaging and Graphics*, vol. 32, no. 8, pp. 678–684, 2008.

[50] K. D. Brandt, P. Dieppe, and E. L. Radin, "Etiopathogenesis of osteoarthritis," *Rheumatic Disease Clinics of North America*, vol. 34, no. 3, pp. 531–559, 2008.

[51] S. Modgil, T. Hutton, P. Hammond, and J. Davenport, "Combining biometric and symbolic models for customised, automated prosthesis design," *Artificial intelligence in medicine*, vol. 25, no. 3, pp. 227–245, 2002.

[52] V. Gilsanz and O. Ratib, *Hand bone age: a digital atlas of skeletal maturity*. Springer Science & Business Media.

[53] E. Pietka, A. Gertych, S. Pospiech, F. Cao, H. K. Huang, and V. Gilsanz, "Computer-assisted bone age assessment: Image preprocessing and epiphyseal/metaphyseal roi extraction," *IEEE trans. on medical imaging*, vol. 20, no. 8, pp. 715–729, 2001.

[54] M. Craigen and J. K. Stanley, "Wrist kinematics: row, column or both?," *Journal of Hand Surgery*, vol. 20, no. 2, pp. 165–170, 1995.

[55] M. W. Tocheri, C. M. Orr, S. G. Larson, T. Sutikna, E. W. Saptomo, R. A. Due, T. Djubiantono, M. J. Morwood, and W. L. Jungers, "The primitive wrist of homo floresiensis and its implications for hominin evolution," *Science*, vol. 317, no. 5845, pp. 1743–1745, 2007.

[56] S. Biasotti, E. Thompson, M. Aono, A. Ben Hamza, B. stos, S. Dong, B. Du, A. Fehri, H. Li, F. Limberger, M. Masoumi, M. Rezaei, I. Sipiran, L. Sun, A. Tatsuma, S. Forero, R. Wilson, Y. Wu, J. Zhang, T. Zhao, F. Fornasa, and A. Giachetti, "SHREC'17 track: Retrieval of surfaces with similar relief patterns," in *Proc. Eurographics Workshop on 3D Object Retrieval 2017*, pp. 1–10, 2017.

[57] M. Masoumi, M. Toews, and H. Lombaert, "Waveletbrain: Characterization of human brain viaspectral graph wavelets," *arXiv:1906.06158v2*, 2019.

[58] M. Masoumi, M. Marcoux, L. Maignel, and C. Pomar, "Spectralweight: a spectral graph wavelet framework for weight prediction of pork cuts," *arXiv*, pp. arXiv–2005, 2020.

[59] S. Rosenberg, *The Laplacian on a Riemannian Manifold*. Cambridge University Press, 1997.

[60] H. Krim and A. Ben Hamza, *Geometric methods in signal and image analysis*. Cambridge University Press, 2015.

[61] M. Meyer, M. Desbrun, P. Schröder, and A. Barr, "Discrete differential-geometry operators for triangulated 2-manifolds," *Visualization and mathematics III*, vol. 3, no. 7, pp. 35–57, 2003.

[62] M. Wardetzky, S. Mathur, F. Kälberer, and E. Grinspun, "Discrete Laplace operators: no free lunch," in *Proc. Eurographics Symp. Geometry Processing*, pp. 33–37, 2007.

[63] A. Schmeling, G. Geserick, W. Reisinger, and A. Olze, "Age estimation," *Forensic Science International*, vol. 165, no. 2-3, pp. 178–181, 2007.

[64] K. Fleshman, "Bone age determination in a pediatric population as an indicator of nutritional status," *Tropical Doctor*, vol. 30, no. 1, pp. 16–18, 2000.

[65] M. W. Warren, K. R. Smith, P. R. Stubblefield, S. S. Martin, and H. A.Walsh-Haney, "Use of radiographic atlases in a mass fatality," *Journal of Forensic Sciences*, vol. 45, no. 2, pp. 467–470, 2000.

[66] R. P. Spencer, S. Sami, and M. Karimeddini, "Role of bone scans in assessment of skeletal age," *International Journal of Nuclear Medicine and Biology*, vol. 8, no. 1, pp. 33–38, 1981.

[67] F. J. van Lenthe, H. C. G. Kemper, and W. VanMechelen, "Skeletal maturation in adolescence: a comparison between the tanner-whitehouse ii and the fels method," *European Journal of Pediatrics*, vol. 157, no. 10, pp. 798–801, 1998.

[68] A. Dimeglio, Y. P. Charles, J. Daures, V. de Rosa, and B. Kabore, "Accuracy of the sauvegrain method in determining skeletal age during puberty," *Journal of Bone and Joint Surgery A*, vol. 87, no. 8, pp. 1689–1696, 2005.

[69] W. W. Greulich and S. I. Pyle, *Radiographic Atlas of Skeletal Development of Hand and Wrist*. Stanford University Press, 1959.

[70] J. M. Tanner and R. H. Whitehouse, *Assessment of Skeletal Maturity and Prediction of Adult Height (TW2 Method)*. Academic Press, 1975.

[71] D. O'Keeffe, *Denoising of carpal bones for computerised assessment of bone age*. PhD thesis, University of Canterbury Christchurch, 2010.

[72] A. Gertych, A. Zhang, J. Sayre, S. Pospiech-Kurkowska, and H. K. Huang, "Bone age assessment of children using a digital hand atlas," *Computerized Medical Imaging and Graphics*, vol. 31, no. 4-5, pp. 322–331, 2007.

[73] G. R. Milner, R. K. Levick, and R. Kay, "Assessment of bone age: a comparison of the greulich and pyle, and the tanner and white house methods," *Clinical Radiology*, vol. 37, no. 2, pp. 119–121, 1986.

[74] A. Schmeling, P. M. G. PM, J. L. Prieto, and M. L. Landa, "Forensic age estimation in unaccompanied minors and young living adults," *Forensic Medicine-From Old Problems to New Challenges*.

[75] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," in *arXiv preprint arXiv:1704.04861*, 2017.

[76] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *arXiv preprint arXiv:1707.07012*, 2017.

[77] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[78] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *European conference on computer vision*, pp. 630–645, 2016.

[79] P. J. Hagen, I. J. Hartmann, O. S. Hoekstra, M. P. Stokkel, G. J. Teule, and M. H. Prins, "How to use a gestalt interpretation for ventilation-perfusion lung scintigraphy," *Journal of Nuclear Medicine*, vol. 43, no. 10, pp. 1317–1323, 2002.

[80] A. Schmeling, C. Grundmann, A. Fuhrmann, H. J. Kaatsch, B. Knell, F. Ramsthaler, W. Reisinger, T. Riepert, S. Ritz-Timme, F. W.Rosing, and K. Rotzscher, "for age estimation in living individuals," *International Journal of Legal Medicine*, vol. 122, no. 6, pp. 457–460, 2008.

[81] J. M. Sotoca, J. M. Inesta, and M. A. Belmonte, "Hand bone segmentation in radioabsorptiometry images for computerised bone mass assessment," *Computerized Medical Imaging and Graphics*, vol. 27, no. 6, pp. 459–467, 2003.

[82] E. Shane, D. Burr, P. R. Ebeling, B. Abrahamsen, R. A. Adler, T. D. Brown, A. M. Cheung, F. Cosman, J. R. Curtis, R. Dell, and D. Dempster, "Atypical subtrochanteric and diaphyseal femoral fractures: report of a task force of the american society for bone and mineral research," *Journal of Bone and Mineral Research*, vol. 25, no. 11, pp. 2267–2294, 2010.