Deep Landmarking: Reconocimiento automático de estructuras anatómicas por medio de morfometría geométrica

Universidad Nacional del Sur

Lic. Celia Cintas

Director: Dr. Claudio Delrieux

24 Septiembre 2017

Prefacio

Esta Tesis se presenta como parte de los requisitos para optar al grado Académico de Doctor en Ciencias de la Computación, de la Universidad Nacional del Sur y no ha sido presentada previamente para la obtención de otro título en esta Universidad u otra. La misma contiene los resultados obtenidos en investigaciones llevadas a cabo en el ámbito del Laboratorio de Ciencias de las Imágenes de la UNS y el Instituto Patagónico de Cs. Sociales y Humanas CCT-CENPAT durante el período comprendido entre el 01 de Abril de 2013 y el 27 de Septiembre de 2017, bajo la dirección del Dr. en Ciencias de la Computación Claudio Delrieux de la Universidad Nacional del Sur.

A Guillermo Ventura Cintas.

Resumen

La adquisición de información fenotípica es un aspecto clave en diversos contextos, incluyendo análisis biométricos, estudios bioantropológicos, investigaciones biomédicas, y ciencia forense por citar algunos. Para ello se requiere la identificación automática de estructuras anatómicas de interés biométrico, como por ejemplo huellas dactilares, patrones en el iris, o rasgos faciales. Estas estructuras son utilizadas masivamente, pero poseen la desventaja de requerir intrusión para adquirir la información a ser analizada. En esta tesis presentamos un nuevo método, basado en la Morfometría Geométrica, para la detección y extacción automática de datos anatómicos característicos (features) en la forma de hitos (landmarks) en 2D o 3D. Para ello se entrenó una red neuronal con conjuntos de datos obtenidos en forma supervisada por medio de expertos antropólogos y biólogos. El sistema resultante posee la capacidad de realizar landmarking en forma automática en imágenes y video sin preparación previa, obteniéndose parámetros de calidad equivalente o superiores a los adquiridos por expertos humanos. Estos resultados abren la posibilidad de generar en forma automática y confiable vectores de atributos basados en propiedades fenotípicas. Se exploran algunas aplicaciones en diversos contextos incluyendo biometría, videojuegos, interfases naturales y otras aplicaciones.

Abstract

Accurate gathering of phenotypic information is a key aspect in several subject matters, including biometric identification, biomedical analysis, bioanthropology studies, forensics, and many other. Automatic identification of anatomical structures of biometric interest, such as fingerprints, iris patterns, or facial traits, are extensively used in applications like access control, anthropological research, and surveillance, all having in common the drawback of requiring intrusive means for acquiring the required information. In this thesis we present a new method, based on two well established methodologies, Geometric Morphometrics and Deep Learning algorithms, for automatic phenotype detection and feature extraction in the form of 2D and 3D landmarks. A convolutional neural network was trained with a set of manually landmarked examples. The trained network is able to provide morphometric landmarks on images automatically, with a performance that matches human assisted landmarking. The ability to perform in the open (*i.e.*, in images or video taken with no specific acquisition preparation). The feasibility of using landmarks as feature vectors for different classifications tasks is explored in a novel spectrum of biometrics, video games, and natural user interfaces applications.

Reconocimientos

Esta Tesis fue desarrollada gracias a la ayuda económica del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) mediante becas estratégicas Bioinformática (Abril 2013 - Abril 2018), y con el subsidio PGI 24/K061 de la SECyT-UNS.

CANDELA (Consortium for the Analysis of the Diversity and Evolution of Latin Americans) obtuvo el financiamiento de Leverhulme Trust (Titulo del Proyecto: Network for the study of the evolution of Latin American populations, # F/07134/DF), lo que permitió tomar la muestra en los cinco países arriba mencionados. Además, se recibió financiación del Biotechnology and Biological Sciences Research Council, que permitió caracterizar genómicamente a los voluntarios. El Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET-Argentina) otorgó mediante sus mecanismos establecidos y regulares tres veces relacionadas con la temática CANDELA, incluyendo la que sustentó esta presente disertación.

Agradecimientos

Quiero agradecer a mis directores Claudio Delrieux y Rolando González-José ya que sin toda su ayuda, conocimientos y aportes esta tesis no sería posible.

A Andrés y Anita Cintas por siempre acompañarme desde el comienzo de este trabajo, por ser apoyo incondicional y Anita en especial por sus puntillosas correcciones y recomendaciones a lo largo de los años y de este manuscrito.

A mi compañero y pareja Defo, quien me hizo *support* en todas las etapas y metamorfismos de los últimos años.

Al Grupo de Investigación en Biología Evolutiva Humana, por formar parte y generar las investigaciones y experimentos más divertidos e interdisciplinarios que he tenido (Mirshita, Caio, Virginia, Caro, Sol, Pablo, Anahi y Bruno!).

Abreviaturas

AAM Active Apperance Models

ARI Adjusted Rand Index

ASM Active Shape Models

BGD Batch Gradient Descent

CNN Convolutional Neural Nets

DL Deep Learning

ERT Extremely Randomized Tree

EV Explained Variance

GPA Generalized Procrustes Analysis

GPU Graphic Process Unit

IID Independent and identically distributed random variables

MC Momentum Clásico

MSE Mean Squared Error

NAG Nesterov Accelerated Gradient

RI Rand Index

RMSE Root Mean Squared Error

ROC Receiver Operating Characteristic

ROI Region of Interest

SGD Stochastic Gradient Descent

Índice general

Índice de figuras

Todas las figuras que no posean referencias a su fuente han sido generadas para esta Tesis por la autora C.C.

Índice de tablas

Lista de Algoritmos

Parte I

Presentación General

Capítulo 1

Introducción

El landmarking automático aplicado a rostros humanos, definido en visión computacional como la detección y localización de puntos característicos de la cara (y todos sus componentes, orejas, ojos, etc), es un paso intermedio muy importante para muchas operaciones subsecuentes de análisis facial. Este análisis incluye diversos aspectos y aplicaciones, incluyendo biometría (Lanitis et al., 1995, Wiskott et al., 1997, Campadelli et al., 2003), animación facial (Kang Liu et al., 2008), interacción humano-computadora, detección de pose y mirada, comprensión de la expresión facial (Tian et al., 2001, Pantic and Rothkrantz, 2000), reconstrucción 3D de rostros (Ying et al., 2006), video juegos y estudios de antropología física. Más allá de su sencilla definición, este problema presenta varios desafíos, desde la misma variabilidad intrínseca de la especie humana, hasta factores como la pose de las personas, su expresión facial, las condiciones de iluminación, las posibles oclusiones parciales, la existencia de defectos en la toma de la imagen o video, etc. Si bien en esta tesis se resuelve el problema del landmarking automático 2D y 3D orientado para aplicaciones específicas en estudios antropométricos, se deja también como resultado un *framework* generalizado disponible para otras aplicaciones.

Tradicionalmente los estudios antropométricos sobre el rostro humano se fundamentan en medidas de características básicas, lo cual es un trabajo supervisado intensivo, y carece de un acercamiento integral al problema. Por otro lado, el fenotipado a gran escala y de alto rendimiento se vuelve una necesidad en la era pos-genómica (Guo et al., 2013), en que la capacidad de relevar y procesar datos masivos tanto en el espectro genómico como el fenotípico es clave para potenciar la detección de factores genéticos y no genéticos res-

ponsables de la variación normal y patológica en humanos. El desarrollo de nuevas técnicas avanzadas en procesamiento de imágenes puede ser un camino valioso para la recolección minuciosa y abarcativa de datos morfológicos de diferentes organismos. En particular, el tejido blando de la cara humana es una compleja geometría, compuesta por varios órganos, incluyendo, ojos, nariz, orejas, boca, etc. Dadas sus funciones biológicas principales, el rostro humano es un tópico central en varias investigaciones, con un amplio rango de aplicaciones, incluyendo, antropología (Gómez-Valdés et al., 2013, Quinto-Sánchez et al., 2015a, Schlager and Rüdell, 2015, Paschetta et al., 2016), medicina genética (Hammond, 2007, Hammond et al., 2005, Weinberg et al., 2008), ciencias forenses (Alexander et al., 2011, Kurniawan et al., 2014, Liu et al., 2015, Albert et al., 2007), envejecimiento (Ramanathan et al., 2009, Fu et al., 2010) y genómica cuantitativa (Liu et al., 2012, Adhikari et al., 2016). Sin embargo, por un largo período de tiempo, no han podido ser usadas en toda su potencialidad las importantes variables cuantitativas sobre rostros humanos, dado que estos estudios requieren un ingreso de datos manual sobre tediosas mediciones tomadas a partir de un conjunto de coordenadas por parte de especialistas. Este ingreso de datos es determinado subjetivamente, dando lugar a grandes variaciones tanto en un mismo especialista a lo largo del tiempo (debido a distracciones, fatiga, etc.) como a variaciones intersubjetivas entre dos o más especialistas (Segev et al., 2010, Kamoen et al., 2001). Estos errores inherentes hacen aún más crítico el proceso de captura de datos.

1.1. Contexto

La determinación de las coordenadas cartesianas de estos puntos emplazados sobre la estructura a estudiar se realiza a través de múltiples dispositivos como brazos digitalizadores, tomografías computadas, escáner de superficie (en el caso de coordenadas 3D) y fotografías digitales de los especímenes orientados en determinado plano (en el caso de coordenadas en 2D). Por razones presupuestarias y logísticas, normalmente se tiende a localizar puntos bidimensionales sobre imágenes digitales del objeto bajo estudio, sea éste un objeto tridimensional que será estudiado descartando una dimensión (e.g. un cráneo de vertebrado visto en normal lateral), o bien se trate de un objeto de carácter claramente bidimensional (p. ej. el ala de una mosca, una hoja de árbol).

Dada esta diversidad de enfoques, la obtención, gestión y sistematización de landmarks es auxiliada por medio de software específico. Una vez obtenidas las configuraciones de landmarks en toda la muestra, el análisis del conjunto de puntos homólogos se facilita por medio de técnicas de superposición como el método Generalizado de Procrustes (ver Sección 3.2), que permite una clara separación entre el tamaño de los objetos y su forma, entendida como toda la variación que permanece en las localizaciones de los landmarks una vez que se eliminaron los efectos de traslación, rotación y escala Goodall and Mardia (1991).

La obtención de coordenadas de landmarks en 2D y 3D, se encuentra en un estado más incipiente, por las dificultades técnicas, económicas y operativas que plantea: ya que debe contarse con dispositivos caros, o no-portátiles, y capaces de recolectar puntos en 3D sobre los objetos bajo estudio. Sin embargo, su aplicación proporcionaría un significativo impacto potencial. Una de las posibles soluciones a este problema radica en obtener coordenadas de landmarks 3D a partir del uso de varias tomas fotográficas con diferentes ángulos que, mediante un ajuste fotogramétrico, permitan derivar las ubicaciones en 3D de landmarks redundantes (fiduciarios) observados en más de una toma 2D. Dicho procedimiento es laborioso, requiere técnicas fotogramétricas complejas, y está sujeto a errores intra observador e inter toma fotográfica.

En esta tesis se propone investigar e implementar técnicas para la adquisición, reconstrucción y manipulación de estructuras bi y tridimensionales. Las mismas pueden provenir tanto de fotografías como ser construidas a partir de los datos provistos por secuencias de fotografías o videos tomadas con diferentes ángulos. Para ello se utilizó el marco formal de la morfometría geométrica, y se utilizaron datos de calidad provistos por especialistas para entrenar redes neuronales convolucionales.

La implementación de esta investigación se realizó en el marco del proyecto internacional CANDELA (Consortium for the Analysis of the Diversity and Evolution of Latin Americans). Dicho proyecto se basa en un consorcio multidisciplinario internacional que incluye científicos especialistas en la diversidad biológica de los latinoamericanos y su entorno socio-cultural. El trabajo de CANDELA se centra en poblaciones urbanas de cinco países: México, Colombia, Perú, Chile y Brasil. Se caracterizó la apariencia física, el acervo genético y el entorno social de personas pertenecientes a esas poblaciones, así como su

percepción y actitudes en torno a sí mismos y a sus semejantes. El objetivo general de este consorcio es poner a prueba una serie de hipótesis relevantes para la antropología, la investigación biológica y médica, así como la relación entre la auto percepción de la identidad y el aspecto físico externo, la ancestría genética y el ambiente socio cultural. Los resultados obtenidos en esta tesis permitieron un significativo avance en la adquisición y análisis de datos con la calidad requerida para estas investigaciones.

El proyecto realizó un muestreo de aproximadamente 7500 individuos, logrando para su primera publicación un total de 7,342 individuos (Ruiz-Linares et al., 2014) a los cuales se aplicó un protocolo general de toma de muestra sanguínea, fenotipado (mediciones antropométricas generales y fotografías faciales) y una encuesta socioeconómica. El proyecto tuvo las aprobaciones de los comités de ética de la Universidad Nacional Autónoma de México, Escuela Nacional de Antropología e Historia (México), Universidad de Tarapacá (Chile), Universidade Federal do Rio Grande do Sul / Universidade Estadual do Sudoeste da Bahia (Brasil), y el University College of London (Reino Unido). En 2009 CANDELA obtuvo el financiamiento de Leverhulme Trust (Titulo del Proyecto: Network for the study of the evolution of Latin American populations, # F/07134/DF), lo que permitió tomar la muestra en los cinco países arriba mencionados. Además, se recibió financiación del Biotechnology and Biological Sciences Research Council, que permitió caracterizar genómicamente a los voluntarios. El Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET-Argentina) otorgó mediante sus mecanismos establecidos y regulares tres veces relacionadas con la temática CANDELA, incluyendo la que sustentó esta la presente disertación. A la fecha el proyecto cuenta con una serie de publicaciones que la respaldan, entre las que se encuentran un análisis genético del color de la piel en relación al mestizaje (De Cerqueira et al., 2014), la exploración del patrón de ancestría genética de la muestra (Ruiz-Linares et al., 2014), asimetría fluctuante asociado a ancestría genética (Quinto-Sánchez et al., 2015b) y un estudio sobre las asimetrías fluctuantes a lo largo de la escala socioeconómica (Quinto-Sánchez et al., 2017), uno de los capítulos de ésta tesis (Cintas et al., 2016a) sobre landmarking automático del pabellón auditivo y un trabajo sobre asociación de genoma completo identificando siete zonas del genoma altamente relacionadas con la variación de la oreja en humanos y ratones (Adhikari et al., 2015).

1.2. Objetivos

- Investigar y definir técnicas de reconstrucción y mallado en 2D y 3D basadas en landmarks, obtenidos desde secuencias de fotografías, video, o desde escáneres 3D.
- Investigar y desarrollar métodos para la toma de landmarks en forma automatizada de distintos fenotipos faciales, utilizando como fuente de imágenes la base de datos de CANDELA, con exactitud y precisión superiores a las de los operadores humanos.
- Investigar la aplicabilidad de dichas metodologías en otros contextos (orejas, perfiles laterales, cuerpos completos, etc.).

1.3. Contribuciones

El principal aporte de este trabajo de investigación es el desarrollo de un *workflow* de procesamiento sobre grandes volúmenes de imágenes, para la obtención automática de landmarks. Si bien en esta tesis se presentan para estructuras específicas (pabellón auditivo, lateral y 3D corporal), es fácilmente extensible a problemas de landmarking en otro tipo de estructuras biológicas Carvajal-Rodríguez et al. (2005), Shipunov and Bateman (2005), o también para uso de orientación de objetos para varias aplicaciones Lovato et al. (2014), Tompson et al. (2014), de reconocimiento Lanitis et al. (1995), realidad aumentada State et al. (1996), Kitanovski and Izquierdo (2011), análisis de gestos Tian et al. (2001), Pantic and Rothkrantz (2000), etc.

Cómo resultados complementarios podemos mencionar los siguientes puntos relevantes:

1. Estudios comparativos sobre la ubicación de regiones de interés (ROI) utilizando redes neuronales convolucionales (CNN) y el algoritmo de Viola-Jones (ver Sección 6.6).
2. Métodos para el reconocimiento biométrico a partir de una configuración de landmarks emplazados sobre el pabellón auditivo utilizando árboles aleatorios (ERT) (ver Sección 6.5).

3. Un análisis sobre la importancia de ciertos landmarks a la hora del ensamblado del vector de características en problemas de identificación de personas (ver Sección 6.5.2).

4. Estudios comparativos de configuraciones de landmarks sobre vista lateral y su ulterior impacto como vectores de características (ver Sección 7.5).

5. Métodos para la clasificación de género a partir de imágenes de vista lateral (ver Sección 7.5).

6. Redes pre-entrenadas y conjunto de datos disponibles online para su uso por parte de la comunidad científica, tanto usuarios finales de problemas de identificación como investigadores que deseen entrenar sus propios modelos teniendo como base nuestro modelo.

Adicionalmente, algunos resultados fueron publicados en periódicos científicos y congresos, mientras que otros resultados se encuentran en revisión y preparación.

1.3.1. Publicaciones en Revistas Internacionales

- Automatic ear detection and feature extraction using Geometric Morphometrics and Convolutional Neural Networks. **CELIA CINTAS** ; MIRSHA QUINTO-SÁNCHEZ ; VICTOR ACUÑA ; CAROLINA PASCHETTA; SOLEDAD DE AZEVEDO; CAIO CESAR SILVA DE CERQUEIRA; VIRGINIA RAMALLO; CARLA GALLO; GIOVANNI POLETTI; MARIA CATIRA BORTOLINI; SAMUEL CANIZALES-QUINTEROS; FRANCISCO ROTHHAMMER; GABRIEL BEDOYA; ANDRES RUIZ-LINARES; ROLANDO GONZALÉZ-JOSÉ; CLAUDIO DELRIEUX. *IET Biometrics*;Volume 6, Issue 3, May 2017, p. 211 – 223. (http://digital-library.theiet.org/content/journals/10.1049/iet-bmt.2016.0002) (Cintas et al., 2016a)

- Socioeconomic Status Is Not Related With Facial Fluctuating Asymmetry: Evidence From Latin-American Populations. MIRSHA QUINTO-SÁNCHEZ; **CELIA CINTAS**; CAIO CESAR SILVA DE CERQUEIRA; VIRGINIA RAMALLO; VICTOR ACUÑA; KAUSTUBH ADHIKARI; JORGE GOMEX-VALDÉS; PAOLA EVERARDO; FRANCISCO DE ÁVILA; TÁBITA HüNEMEIER; CLAUDIA JARAMILLO; WI-

LLIAM ARIAS; CARLA GALLO; MACARENA FUENTES; GIOVANNI POLETTI; LAVINIA SCHULER-FACCINI; MARIA CATIRA BORTOLINI; SAMUEL CANIZALEZ-QUINTEROS; FRANCISCO ROTHHAMMER; GABRIEL BEDOYA; JAVIER ROSIQUE; ANDRES RUIZ-LINARES; ROLANDO GONZALEZ-JOSE. *PLOS ONE*; Año: 2016. (Quinto-Sánchez et al., 2017)

- Predicting Physical Features and Diseases by DNA Analysis: Current Advances and Future Challenges. CAIO CESAR SILVA DE CERQUEIRA; VIRGINIA RAMALLO; TABITA HÜNEMEIER; DE AZEVEDO, SOLEDAD; MIRSHA QUINTO-SÁNCHEZ; PASCHETTA, CAROLINA ANDREA; **CINTAS, CELIA**; GONZÁLEZ, MARINA FERNANDA; MARIA CÁTIRA BORTOLINI; GONZÁLEZ-JOSÉ, ROLANDO. *Journal of Forensic Research*; Año: 2016. (DOI: 10.4172/2157-7145.1000337) (de Cerqueira et al., 2016).

- Shifts in subsistence type and its impact on the human skull's morphological integration. PASCHETTA, CAROLINA ANDREA; DE AZEVEDO, SOLEDAD; GONZÁLEZ, MARINA FERNANDA; QUINTO-SÁNCHEZ MIRSHA; **CINTAS, CELIA**; VARELA, HÉCTOR HUGO; GÓMEZ-VALDÉS JORGE; SÁNCHEZ-MEJORADA GABRIELA; GONZÁLEZ-JOSÉ, ROLANDO. *American Journal of Human Biology*; Año: 2015 vol. 28 p. 118 - 128. (Paschetta et al., 2016)

- Facial asymmetry and genetic ancestry in Latin-American admixed populations. MIRSHA QUINTO-SÁNCHEZ; KAUSTUBH ADHIKARI; VICTOR ACUÑA-ALONZO; **CINTAS, CELIA**; CAIO CESAR SILVA DE CERQUEIRA; VIRGINIA RAMALLO; LUCIA CASTILLO; ARODI FARRERA; CLAUDIA JARAMILLO; WILLIAMS ARIAS; MACARENA FUENTES; PAOLA EVERARDO; FRANCISCO DE AVILA; JORGE GOMEZ-VALDÉS; TÁBITA HÜNEMEIER; SHARA GIBBON; CARLA GALLO; GIOVANNI POLETTI; JAVIER ROSIQUE; MARIA CÁTIRA BORTOLINI; SAMUEL CANIZALES-QUINTEROS; FRANCISCO ROTHHAMMER; GABRIEL BEDOYA; ANDRES RUIZ-LINARES; ROLANDO GONZÁLEZ-JOSÉ *American Journal of Physical Anthropology*; Año: 2015. (Quinto-Sánchez et al., 2015b)

1.3.2. Actas en Conferencias Nacionales e Internacionales

- **CINTAS, CELIA**; PABLO NAVARRO; VIRGINIA RAMALLO; BRUNO PAZOS; CLAUDIO DELRIEUX; ANAHÍ RUDERMAN; CAROLINA PASCHETTA; SOLEDAD DE AZEVEDO; ROLANDO GONZALEZ-JOSÉ. Posicionamiento Automático de landmarks corporales 3D mediante morfometría geométrica y redes neuronales:Aplicaciones Bioantropológicas. XIII Jornadas Nacionales de Antropología Biológica. Lugar: Necochea, Argentina. Año: 2017. (Cintas et al., 2017)

- **CINTAS, CELIA**; CLAUDIO DELRIEUX; MIRSHA QUINTO-SÁNCHEZ; CAROLINA PASCHETTA; VIRGINIA RAMALLO; CAIO CESAR DE CERQUEIRA; SOLEDAD DE AZEVEDO; ROLANDO GONZALEZ-JOSÉ. Posicionamiento automático de landmarks anatómicos 2D mediante morfometría geométrica y deep learning: aplicaciones bioantropológicas. XIV Congreso Asociación Latinoamericana de Antropología Biológica. Lugar: Tacuarembó, Uruguay. Año: 2016; (Cintas et al., 2016b)

- MIRSHA QUINTO-SANCHEZ; JORGE GOMEZ-VALDÉS; VICTOR ACUÑA-ALONZO; **CELIA CINTAS**; CAIO CESAR SILVA DE CERQUEIRA; RAMALLO, VIRGINIA; KAUSTUBH ADHIKARI; CLAUDIA JARAMILLO; WILLIAM ARIAS; MACARENA FUENTES; PAOLA EVERARDO; FRANCISCO DE AVILA; TÁBITA HÜNEMEIER; SHARA GIBBON; CARLA GALLO; GIOVANNI POLETTI; BORTOLINI, MARIA CATIRA; SAMUEL CANIZALES-QUINTEROS; FRANCISCO ROTHHAMMER; GABRIEL BEDOYA; ANDRÉS RUIZ-LINARES; GONZÁLEZ-JOSÉ, ROLANDO. Asimetría Fluctuante Facial en Poblaciones Latinoamericanas XVIII. Coloquio Internacional de Antropología Física "Juan Comas". Lugar: Durango Durango México; Año: 2015;

- **CINTAS, CELIA**; MIRSHA QUINTO-SÁNCHEZ; CLAUDIO DELRIEUX; ROLANDO GONZALEZ JOSÉ Automatic Landmarking App for Bioanthropology. IV EURO WG Conference on Operational Research in Computational Biology, Bioinformatics and Medicine. Lugar: Biedrusko, Polonia. Año: 2014. (Cintas et al., 2014a)

- **CINTAS, CELIA** ; QUINTO-SÁNCHEZ MIRSHA; GONZÁLEZ-JOSÉ, ROLANDO; CLAUDIO DELRIEUX; Python in the world of Biological Anthropology. Eu-

roPython 2014; Lugar: Berlin, Alemania; Año: 2014. (Cintas et al., 2014b)

- **CINTAS, CELIA**; MIRSHA QUINTO-SÁNCHEZ; GLORIA BIANCHI; NAHUEL DEFOSSÉ; CLAUDIO DELRIEUX; ROLANDO GONZALEZ-JOSÉ Aplicación Bioantropológica para manipulación de landmarks faciales en 2D. 16° Edición del Workshop de Investigadores en Ciencias de la Computación. Lugar: Ushuaia. Año: 2014. (Cintas et al., 2014c)

- QUINTO-SÁNCHEZ MIRSHA; **CINTAS CELIA**; CASTILLO LUCIA,; CAIO CESAR SILVA DE CERQUEIRA; VIRGINIA RAMALLO; ROLANDO GONZALEZ-JOSÉ Asimetría fluctuante facial y su relación con índices socioeconómicos. XIII versión del Congreso de la Asociación Latinoamericana de Antropología Biológica (ALAB). Lugar: Santiago; Año: 2014; (Quinto-Sánchez et al., 2014)

- **CINTAS, CELIA**; GLORIA BIANCHI; NAHUEL DEFOSSÉ; CLAUDIO DELRIEUX; MIRSHA QUINTO-SÁNCHEZ; ROLANDO GONZALEZ-JOSÉ PopEye: Bioanthropologic app for the automatic positioning of anatomical landmarks in 2D and 3D 4to. Congreso Argentino de Bioinformática y Biología Computacional (4CAB2C) y 4ta. Conferencia Internacional de la Sociedad Iberoamericana de Bioinformática (SolBio) Lugar: Rosario. Año: 2013. (Cintas et al., 2013a)

- **CINTAS, CELIA**; CLAUDIO DELRIEUX; MIRSHA QUINTO-SÁNCHEZ; ROLANDO GONZALEZ-JOSÉ Posicionamiento Automático de Landmarks Anatómicos en 2 y 3D: aplicaciones bioantropológicas XI Jornadas Nacionales de Antropología Biológica Lugar: Buenos Aires. Año: 2013; (Cintas et al., 2013b)

- **CINTAS, CELIA**; CLAUDIO DELRIEUX; ROLANDO GONZALEZ JOSÉ. Posicionamiento Automático de Landmarks Anatómicos en Ojos. CACIC Congreso Argentino de Ciencias de la Computación. Lugar: Bahia Blanca. Año: 2012. (Cintas et al., 2012)

1.3.3. Otras publicaciones y presentaciones en Congresos

- *(Taller)* Introduction to Deep Learning with Python: The force awakens. **CELIA CINTAS**. PyCon Ireland 2017, Dublin, Ireland.

- *(Presentación)* Introduction to Deep Learning with Python: The force awakens. **CELIA CINTAS**. PyCon UK 2017, Cardiff, Wales.
- *(Taller)* Introducción a Deep Learning: El despertar de la fuerza. **CELIA CINTAS**. 3er Workshop de Ingenieria en Sistemas de Información - UTN FRD 2017 Campana, Buenos Aires, Argentina.
- *(Presentación)* Introducción a Deep Learning: El despertar de la fuerza. **CELIA CINTAS**. Nerdear.la 2017, Buenos Aires, Argentina.
- *(Taller)* Introducción a Deep Learning: El despertar de la fuerza. **CELIA CINTAS**. PyCon Argentina 2016, Bahia Blanca, Argentina.
- *(Presentación)* Introducción a programación paralela con PyOpenCL. **CELIA CINTAS**. PyCon Argentina 2016, Bahia Blanca, Argentina.
- *(Taller)* Introduction to High-Performance Scientific Computing with PyOpenCL. **CELIA CINTAS**. SciPy Latinoamericana 2016, Florianopolis, Brasil.
- *(Taller)* Fit and predict your data: Introduction to Scikit-learn. **CELIA CINTAS**. Scipy Latinoamericana 2016, Florianopolis, Brasil.
- *(Presentación)* Showing some of the goodies: pandas, scikit-learn and matplotlib. **CELIA CINTAS**. SciPy Latinoamericana 2015, Posadas, Argentina.
- *(Presentación)* 3D sensors and Python: A space odyssey. **CELIA CINTAS**. Europython 2014, Berlin, Alemania.
- *(Poster)* Visualización de datos geográficos-genomicos con Basemap. **CELIA CINTAS**. SciPy Latinoamericana 2015, Posadas, Argentina.
- *(Poster)* Construyendo redes neuronales de forma fácil con Lasagne. **CELIA CINTAS**. SciPy Latinoamericana 2015, Posadas, Argentina.
- *(Poster)* PopEye: Python in the world of Biological Anthropology. **CELIA CINTAS**. Europython 2014, Berlin, Alemania.

- RAMALLO, VIRGINIA; SILVERA, SILENE; ANGARONI, CELIA; **CINTAS, CELIA**; GONZÁLEZ-JOSÉ, ROLANDO; DODELSON DE KREMER, RAQUEL; LARÓVERE, LAURA; DIPIERRI, JOSÉ EDGARDO; Genes y apellidos: estudio de un aislado poblacional del gen ctln1 en Villa Mercedes(San Luis). XIV Congreso de la Asociación Latinoamericana de Antropología Biológica; Lugar: Tacuarembó, Uruguay; Año: 2016;

- PASCHETTA, CAROLINA ANDREA; RAMALLO, VIRGINIA; DE AZEVEDO, SOLEDAD; GONZÁLEZ, MARINA FERNANDA; QUINTO-SÁNCHEZ MIRSHA; **CINTAS, CELIA**; VARELA, HÉCTOR HUGO; GONZÁLEZ-JOSÉ, ROLANDO Cambios en el tipo de subsistencia y si impacto en la integración morfológica del cráneo humano XII Jornadas Nacionales de Antropología Biológica Lugar: Corrientes; Año: 2015;

- PASCHETTA, CAROLINA ANDREA; DE AZEVEDO, SOLEDAD; GONZÁLEZ, MARINA FERNANDA; QUINTO-SÁNCHEZ MIRSHA; **CINTAS, CELIA**; SILVA DE CERQUEIRA, CAIO; RAMALLO, VIRGINIA; GONZÁLEZ-JOSÉ, ROLANDO. La fuerza de mordida y su relación con la forma del cráneo. XIII CONGRESO DE LA ASOCIACIÓN LATINOAMERICANA DE ANTROPOLOGÍA BIOLÓGICA. Lugar: Santiago de Chile; Año: 2014;

- ROLANDO GONZALEZ-JOSÉ; PASCHETTA, CAROLINA; DE AZEVEDO, SOLEDAD; GONZÁLEZ, MARINA; MIRSHA QUINTO SÁNCHEZ; **CELIA CINTAS**. Grupo de Investigación en Biología Evolutiva Humana (GIBEH). Segundo Encuentro de Morfometría. Morfometría y estudios ontogenéticos. Lugar: La Plata; Año: 2013;

1.4. Aplicaciones

El landmarking automático facial puede ser utilizado sobre el tracking en video (para clasificación de gestos Pantic and Rothkrantz (2000)). La animación facial realista es un aspecto clave en aplicaciones de interfaz máquina-humano, como agentes virtuales, asistentes virtuales en aplicaciones *e-commerce*, *e-learn*, *e-care* Kang Liu et al. (2008). Ciertas configuraciones de landmarks faciales, tales como la vista frontal y lateral son

utilizadas para modelar envejecimiento, obteniendo un generador automático de avatars de diferentes edades de forma realista Ramanathan et al. (2009). Diferentes configuraciones de landmarks pueden ser diseñadas para la determinación de edad y género de un individuo Katina et al. (2004) o ser datos auxiliares para la estimación de pose o ubicación de un individuo Lovato et al. (2014), Tompson et al. (2014). El uso de configuraciones de landmarks en vista frontal ha sido usado extensivamente para la identificación y reconocimiento de personas Shi et al. (2006), Jiazheng et al. (2005), Beumer et al. (2006). Por otro lado, el uso de landmarking sobre el pabellón auditivo también ha sido probado de gran uso para identificación de individuos Cintas et al. (2016a), Pflug and Busch (2012). En el ámbito médico el landmarking facial de vista lateral es utilizado para aportar información en procesos de cirugía plástica Freitas et al. (2015). Dado el gran rango de aplicabilidad que posee el uso de landmarks en variados campos, como se vio anteriormente, es crucial que la ubicación de los mismos se realice de forma automática, fiable y eficiente a gran escala.

1.5. Estructura de la Tesis

Esta tesis se encuentra desarrollada en cuatro partes. La Parte I detalla aspectos clave de la detección de automática de landmarks, la presentación del problema, metas y contribuciones realizadas. En la Parte II se presenta estado del arte, el marco teórico relacionado con la morfometría geométrica, y una introducción a los aspectos requeridos de *deep learning*, particularmente las CNN que fueron utilizadas extensivamente para resolver los problemas asociados a este trabajo. Dentro de la Parte III se hace referencia a los modelos utilizados para la realización del ladmarking automático sobre los distintos fenotipos evaluados, su diseño, implementación del pipeline detallado y evaluación. También se hace referencia, para cada fenotipo, de un posible uso de la salida del landmarking como vector de características para diferentes tareas. Por último, en la parte IV se presentan las Conclusiones, trabajos en curso y futuros. Además, para que este documento sea autocontenido, se incluyen capítulos anexos con una introducción concisa a las tecnologías utilizadas.

Capítulo 2

Trabajos previos

2.1. Modelos de formas Activos (ASM)

Una de las aproximaciones más populares para realizar landmarking 2D automático ha sido el uso de modelos estadísticos de forma (statistical shape models), en algoritmos como por ejemplo el *Active Shape Model* (ASM) Cootes and Taylor (1992), Cootes et al. (1994), Hill et al. (1996), Cristinacce and Cootes (2007). Para aplicar ASM, se alinea un conjunto de imágenes de entrenamiento a un mismo sistema de coordenadas mediante un análisis de Procrustes (ver Sección 3.2). Los ASM clásicos están caracterizados por el uso de la *distancia de Mahalanobis* sobre perfiles unidimensionales asignados a cada landmark, y un modelo lineal de distribución de puntos. El ASM es entrenado con un set de imágenes sobre las que se realizó el landmarking manualmente. Luego del entrenamiento se utiliza el ASM para buscar catacterísticas (features) sobre el objeto. La idea general es tratar de encontrar cada landmark de forma independiente y luego corregir la ubicación de cada uno con respecto de los otros. Un ASM está compuesto por los siguientes elementos:

Modelo de perfiles: Para cada *landmark*, se describen las características de la imagen alrededor de ese punto. Para generar este vector g del landmark i, se extrae un vector ortogonal al borde de la forma global, el cual esta formado por valores en escala de grises (0.,255) como se observa en la Figura 2.1. El modelo nos indica la forma que se espera de la imagen cerca de ese landmark. Durante el entrenamiento, se toman muestras del área que rodea cada landmark a lo largo de todo el set de imágenes para crear un modelo de perfiles para cada landmark. Este modelo se

construye obteniendo un perfil promedio $\bar{g}$ y una matriz de covariancia S_g de todos los perfiles de entrenamiento, uno por cada imagen de ese landmark. Se supone que todos los perfiles extraídos se ubican aproximadamente en una distribución Gaussiana Multivariada, por lo cual pueden ser descriptas con dichos parámetros estadísticos. Si se cuenta con n landmarks en el modelo, se tendrán $\bar{g}_1 \cdots \bar{g}_n$ y $S_g^1, \cdots S_g^n$ parámetros. Durante la búsqueda, se toman muestras del área cercana a cada landmark tentativo, y se ubica el mismo en la posición que mejor concuerda con el modelo de perfiles de ese landmark. La distancia entre el perfil de búsqueda g y el promedio de su modelo $\bar{g}$ es calculado usando la distancia de Mahalanobis $(g - \bar{g})^T S_g^{-1}(g - \bar{g})$. El perfil que cuente con la menor distancia será la nueva posición tentativa del landmark i (normalmente llamada *forma sugerida*). Este proceso se repite para cada landmark antes de entregar el control al modelo de forma.

Modelo de forma: Este modelo define las posibles posiciones relativas de los landmarks. Durante la búsqueda, el modelo de forma se ajusta a la forma sugerida por el modelo de perfiles, para conformar una forma legítima del objeto cuyo landmarking fue realizado. Esto es necesario dado que el perfil concuerda con cada landmark y no con el consenso global. El propósito del modelo de forma es convertir las formas sugeridas por el modelo de perfiles a una forma del objeto válida. Antes de construir el modelo de forma, las formas de entrenamiento son alineadas. El modelo se compone de la forma promedio de todos los elementos provenientes del entrenamiento y un conjunto de distorsiones posibles.

Modelos Activos de Apariencia

Como continuación a estas ideas, Cootes, Edwards y sus colegas desarrollaron un modelo basado en apariencia (Active Appearance Model, Active Apperance Models (AAM)) Cootes et al. (1998), Edwards et al. (1998). Los AAM unifican el modelo de forma y textura. La textura es medida sobre todo el objeto, (por ejemplo, un rostro). Se utiliza una técnica de optimización que pre-calcula valores residuales durante el entrenamiento para la asignación de parámetros a partir de la textura de la imagen al modelo. Los AAM extienden los ASM, lo que nos da un punto de comparación entre ambos métodos. Los

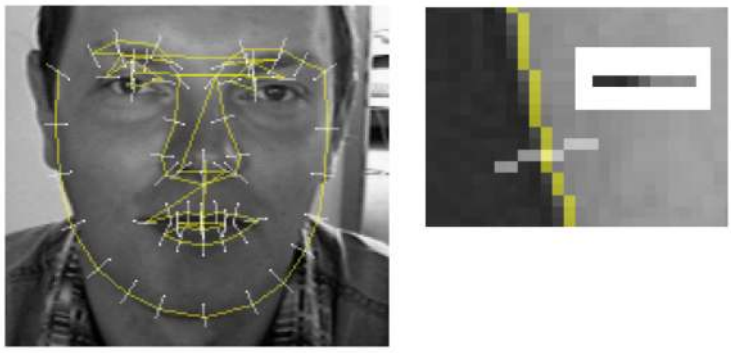


Figura 2.1: En amarillo la "cara promedio" al comienzo de la búsqueda. Las líneas blancas son los vectores de características de cada landmark que forman parte del modelo de perfiles (Milborrow (2007), Milborrow et al. (2013), Milborrow and Nicolls (2014)).

ASM sólo toman información de textura sobre el landmark, mientras que los AAM almacenan la textura de todo el objeto. Los AAM utilizan información sobre la textura a lo largo de todo el objeto, en comparación con ASM que sólo toman información en la vecindad de una coordenada dada. Es debatible, si este punto presenta una ventaja o no, dado que los landmarks son posiciones ubicadas estratégicamente en la ROI, y contienen la información más importante para un determinado problema. Esto convierte a los ASM de forma natural en modelos que no contienen datos irrelevantes. Visto desde otro punto de vista, los AAM necesitan menor cantidad de landmarks dado que contienen información extra en el modelo de textura. Los ASM para caras pueden ser entrenados en pocos minutos, mientras que los AAM tardan más en entrenar ya que cuentan con mayor cantidad de datos. En Cootes et al. (1999) se reportó que la exactitud de la ubicación de landmarks es mejor en modelos del tipo ASM.

2.2. Características SIFT

Para detección de puntos faciales se pueden encontrar varias investigaciones Milborrow et al. (2013), Li et al. (2009), Rattani et al. (2007) que utilizan descriptores locales y globales para la determinación de puntos. Estos trabajos utilizan en general el método SIFT (*Scale-Invariant Feature Transformation*) Lowe (1999, 2004) con distintos tamaños de ventanas. Este tipo de características son invariantes frente a cambios de escala y rotación y otras transformaciones afines. El costo computacional del método se reduce

basándose en una aproximación de filtrado en cascada, realizado a través de los siguientes pasos:

- Detección de escalas y espacios en el conjunto de datos. En este punto se encuentran las distintas escalas y ubicaciones en que se puede encontrar un mismo objeto en todas las imágenes del conjunto de datos. La localización de áreas que son invariantes a la escala puede ser llevada a cabo mediante la búsqueda de características invariantes a la misma. Se mostró en Koenderink (1989) y Lindeberg (1999) que el kernel Gaussiano es adecuado como modelo para la detección de puntos invariantes frente a escala espacial. Esta escala está definida por una función $L(x, y, \sigma)$, la cual es la convolución de una variable de escala Gaussiana $G(x, y, \sigma)$ con la imagen $I(x, y)$ bajo estudio:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y), \tag{2.1}$$

  donde $*$ es el operador de convolución sobre (x, y), y,

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2, y^2)/2\sigma^2}. \tag{2.2}$$

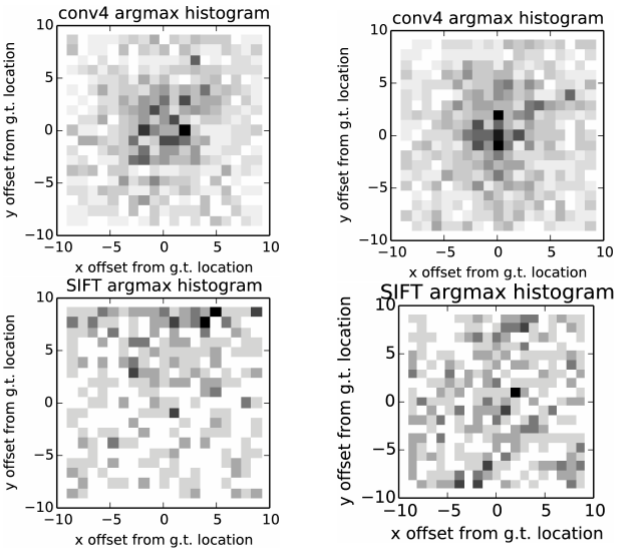
- Ubicación de coordenadas. Se localizan puntos candidatos a ser puntos característicos (*features*). A cada uno de ellos se les determina la ubicación y la escala. Luego se realiza una selección de los mismos en base a su estabilidad a lo largo del conjunto de datos.

- Determinación de orientación. Una o varias orientaciones son asignadas al punto encontrado basándose en la dirección de los gradientes locales en esa porción de la imagen.

- Construcción de descriptores por cada punto encontrado. Se calculan los gradientes de la imagen alrededor de cada punto. Estos se almancenan como un vector de distorsión de forma e iluminación local.

2.2.1. SIFT vs. CNN

El uso de SIFT para la determinación de puntos anatómicos sobre rostros ha mostrado resultados satisfactorios Milborrow et al. (2013), Li et al. (2009), Rattani et al.



(a) cat left eye (b) cat nose

Figura 2.2: Se muestra la capacidad de localización en granularidad fina que tienen las CNN contra características SIFT Long et al. (2014).

(2007). Sin embargo, son más eficientes en tareas de reconocimiento de objetos, ya que la determinación correcta de las características debe ser establecida para cada aplicación en particular. En Long et al. (2014) se presenta un estudio comparativo entre SIFT y CNN. Si bien los clasificadores basados en SIFT no parecen ser sensibles a la ubicación precisa de los puntos, en varios casos las CNN son capaces de una ubicación más precisa. En la Figura 2.2, por ejemplo, se reproducen los resultados de un estudio comparativo de precisión en la ubicación de puntos mediante SIFT y CNN.

2.3. Método Viola-Jones

Este algoritmo consiste en un *framework* para detección de objetos que ha sido ampliamente utilizado en la bibliografía Viola and Jones (2001). El mismo está basado en la representación por medio de imágenes integrales (lo cual facilita el cómputo de luminancias dentro de rectángulos), un algoritmo de aprendizaje basado en *AdaBoost* Freund Robert Schapire (1999) (el cual selecciona la cantidad mínima de características visuales críticas para su posterior clasificación), y una combinación de clasificadores en cascada (lo cual permite una rápida eliminación de características pertenecientes al fondo, dejando más

tiempo de procesamiento para las partes de la imagen potencialmente pertenecientes a objetos). Las características utilizadas en este método son similares a las funciones de la base de *Haar* previamente utilizadas en Papageorgiou et al. (1998). Particularmente se utilizan tres características (ver Figura 2.3):

1. El valor de una característica basada en dos rectángulos, la diferencia entre la suma de los píxeles de dos regiones rectangulares adyacentes (verticales o horizontales, A o B en la Figura mencionada).

2. El valor de una característica basada en tres rectángulos, que calcula la suma entre los dos rectángulos exteriores menos la suma de los píxeles del rectángulo del medio (C).

3. El valor de una característica basada en cuatro rectángulos, que calcula la diferencia entre los rectángulos de la diagonal (D).

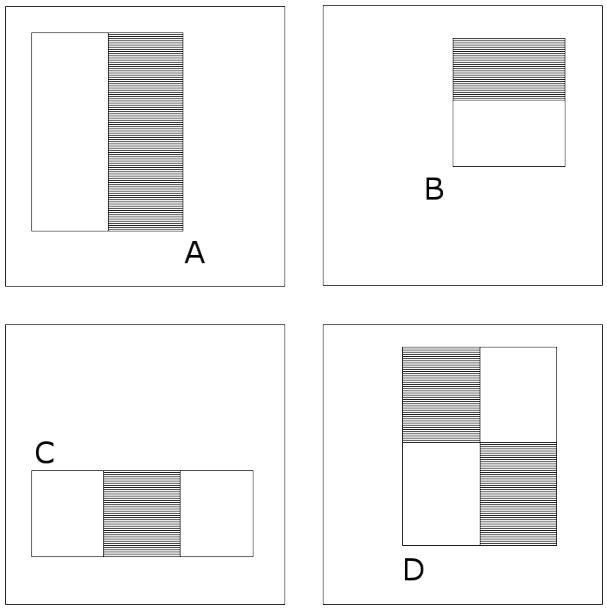


Figura 2.3: Ejemplo de características rectangulares utilizadas en Viola and Jones (2001). La suma de los píxeles que se encuentran en el área blanca se restan de la suma de los píxeles pertenecientes al área gris.

2.3.1. Viola-Jones vs. CNN

En los trabajos preliminares de esta tesis, inicialmente se utilizó el método de Viola and Jones (2001) para identificar la ROI del pabellón auditivo dentro de la imagen. Se puede ver en detalle en el Capítulo 6 el entrenamiento del mismo. Sin embargo, a partir de trabajos posteriores (Cintas et al. (2016a)) se realizaron pruebas de entrenamiento de CNN sobre imágenes completas y se analizó el uso de CNN para la ubicación de la ROI, obteniéndose una mejor performance. Para más detalles ver Sección 6.6.

Parte II

Tecnologías

Capítulo 3

Morfometría Geométrica

En este Capítulo se presenta el marco conceptual subyacente a la Morfometría Geométrica y el método conocido como *Generalized Procrustes Analysis* (GPA). Se definen asimismo los elementos matemáticos que han sido implementados y utilizados extensivamente en esta tesis. Los resultados se pueden ver en los Capítulos 6, 7 y 8. La implementación del método se encuentra disponible en el repositorio (`https://github.com/celiacintas/deeplandmarking`).

3.1. Introducción

La Morfometría Geométrica plantea una metodología para el análisis cualitativo de la forma de los objetos, específicamente en los organismos biológicos en general Zelditch et al. (2004), Mitteroecker and Gunz (2009) y en humanos en particular Slice (2005). Por dicha razón sus aplicaciones principales se encuentran en el análisis de registros fósiles, el estudio de cambios alométricos van der Linde and Houle (2009), ontogenéticos Vidarsdottir et al. (2002), heterocrónicos Lieberman et al. (2007), la biomecánica Lieberman et al. (2004), Paschetta et al. (2010), la evaluación de alteraciones en el desarrollo Klingenberg et al. (2002), los estudios filogenéticos González-José et al. (2008b), Perez et al. (2011) (ver Figura 3.1), y las comparaciones de variación intra e interpoblacional en el contexto histórico González-José et al. (2008a), Perez et al. (2011), entre otras.

La morfometría geométrica se basa en el análisis del tamaño y la forma de los especímenes bajo estudio a partir del desplazamiento en el espacio bi o tridimensional de una

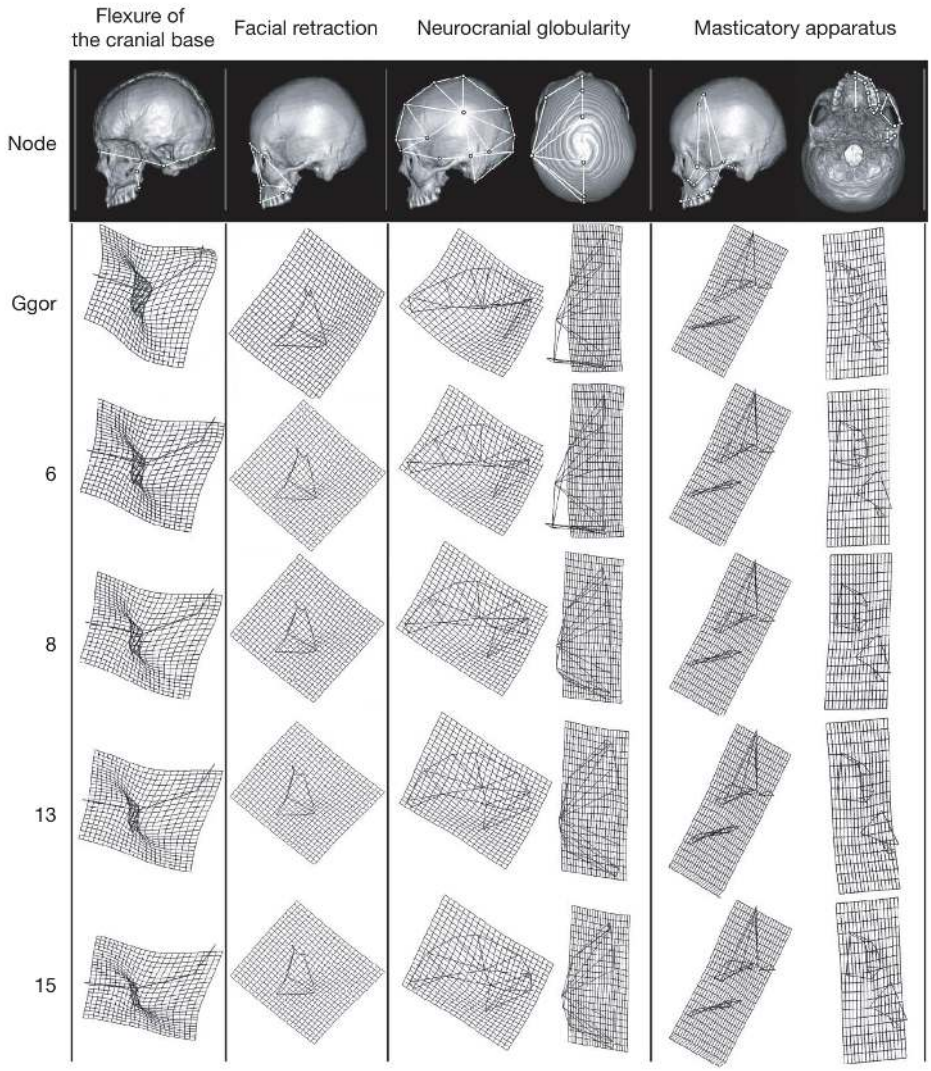


Figura 3.1: Estudios filogenéticos basados en morfometría geométrica (González-José et al. (2008a)).

configuración de puntos anatómicos (landmarks). Dichos landmarks se establecen siguiendo criterios que determinen que son homólogos desde el punto de vista anatómico y/o geométrico (Zelditch et al., 2004) a través de la muestra a estudiar (otros organismos de una misma especie, el mismo organismo en diferentes estados de desarrollo, etc.).

Su principal innovación teórica radica en un cambio rotundo en la aproximación al estudio del tamaño y la forma de las estructuras investigadas. En lugar de enfocarse en el análisis multivariante de un conjunto de medidas lineales entre puntos morfométricos, la Morfometría Geométrica propone estudiar los cambios en el tamaño y la forma a partir del desplazamiento en el plano (2D) o en el espacio (3D) de un conjunto de puntos mor-

fométricos o *landmarks*. La relación espacial en dos o tres dimensiones de estos landmarks siempre se conserva a lo largo de todo el análisis, lo que permite reconstruir con tanta precisión como se desee la forma y el tamaño del espécimen estudiado. Desde comienzos de los años '90 se ha desarrollado un conjunto de métodos analíticos y gráficos que permiten observar estos cambios espaciales desde una óptica estadística, un paso fundamental en el desarrollo de un método biológico.

Se debe resaltar que en Morfometría se hace referencia tanto a la forma como a datos que se refieren sólo a tamaño y contorno. Por forma (*shape*) nos referimos a las propiedades geométricas de un objeto que son invariantes respecto de la ubicación, escala u orientación. Al estar enfocada únicamente en la geometría, se puede dejar de lado propiedades tales como el color y la textura que no aportan elementos esenciales para el estudio. En cambio, las cualidades de invariancia respecto de la posición, escala u orientación pasan a ser centrales. Esto se puede lograr mediante la utilización de medidas invariantes como lo son las distancias de radios, ángulos o mediante la utilización de métodos que permiten transformar todos los datos en un sistema de coordenadas común.

Los métodos de Morfometría Geométrica cuantifican la forma de cada espécimen de acuerdo a la ubicación en el espacio de un conjunto de landmarks o puntos que son homólogos entre individuos. Luego, el tamaño y la forma son separados a través del análisis de Procustes (ver Sección 3.2), que traslada dichos puntos a un origen común de coordenadas, los escala a un tamaño común, y los rota hasta minimizar la suma de cuadrados de las distancias entre sí. El método de GPA Goodall and Mardia (1991) (ver Sección 3.2) permite cuantificar la forma como una desviación multidimensional de los landmarks de un espécimen respecto de una configuración de referencia (usualmente, el promedio de todas las configuraciones de la muestra).

3.1.1. Antecendentes Históricos

A pesar de que los naturalistas del mundo antiguo y renacentista se interesaron tempranamente por la variación morfológica de los organismos, el pensamiento biométrico se sustenta principalmente en los avances introducidos a partir del siglo XIX por Adolphe Quetelet, Francis Galton, Karl Pearson y sir Ronald Fisher. La biometría nace con Quetelet (1796-1874), quien fue quizá la primera figura relevante en el pensamiento biométrico.

Este astrónomo y matemático belga introdujo el concepto de "hombre medio"(una idea del promedio de una serie) y fue el primero en entender que este parámetro se hacía constante al considerar grandes muestras.

La teoría estadística se vio impulsada por los matemáticos del siglo XIX como Galton (1822-1911) quien introdujo la aplicación del método estadístico al análisis de la variación biológica Sokal and Rohlf (1995). A Galton se debe la introducción de conceptos centrales en la Biometría actual, como los cálculos de regresión y correlación. Su metodología ha llegado a ser el fundamento de la aplicación de la estadística a la Biología. Luego, Karl Pearson (1857-1936) continuó la tradición de Galton y extendió el estudio hasta la estadística descriptiva y los métodos de correlación.

Más tarde, la figura dominante en Biometría fue Ronald Fisher. La contribución de Fisher (1890-1962) abarca el desarrollo de métodos para ser aplicados a muestras pequeñas, el descubrimiento de funciones de distribución precisas para muchas muestras estadísticas, la invención del concepto de máxima verosimilitud y prueba de hipótesis, así como el análisis de la varianza, todos ellos vigentes en la actualidad. Si bien los aportes de estos matemáticos y biólogos forman el núcleo duro de la biometría clásica, también es cierto que todos sus aportes son incorporados a la teoría y práctica de la morfometría geométrica.

Con el advenimiento conjunto de la Teoría Sintética de la Evolución y la genética de poblaciones, en la década de 1930, el estudio de los rasgos morfológicos es sintetizado en una nueva estructura de pensamiento, donde el objetivo principal es el estudio de la variabilidad y sus causas desde una óptica evolutiva. Los avances en genética poblacional vienen acompañados de la necesidad de una mayor comprensión del determinismo en estos caracteres, cuya expresión es el resultado de complejas interacciones entre genes y ambiente, necesidad que da origen a la llamada genética de los caracteres cuantitativos, y fundamentada principalmente en los aportes de Sewall Wrifht Wright (1984), Falconer and Mackay (1996).

3.1.2. Ventajas y desventajas con respecto a la morfometría clásica

Algunos de los problemas de la morfometría clásica están relacionados con el alto grado de correlación que existe entre las medidas de distancia lineales y el tamaño, sesgando por

lo tanto los patrones de variación en la forma Bookstein (1984). Otra de las dificultades es establecer homologías entre distancias lineales, debido a que muchas distancias (por ejemplo, ancho máximo) no se definen por puntos homólogos. También se plantea el problema de que el mismo conjunto de medidas puede ser obtenido a partir de dos formas diferentes.

Por último, en la morfometría clásica usualmente no es posible generar representaciones gráficas de los cambios de la forma porque las relaciones geométricas entre las variables no se conservan, de manera que algunos aspectos concernientes a la forma del objeto de estudio se pierden Zelditch et al. (2004).

3.1.3. Landmarks, semilandmarks y contornos

Los *Landmarks* son puntos en un espacio bi o tridimensional que corresponden a la posición de un rasgo en particular en un objeto de interés. En la Morfometría Geométrica, los landmarks se corresponden con ubicaciones discretas anatómicas que pueden ser reconocidas en la misma ubicación en todos los especímenes del estudio. Los landmarks ideales cumplen los siguientes requisitos:

1. Proveen una adecuada cobertura morfológica.
2. Pueden ser identificados repetitivamente y con precisión.
3. Deben ubicarse en el mismo plano.

Podemos clasificar a los landmarks en tres clases:

1. **Tipo I**, son puntos cuya supuesta homología de un individuo a otro es respaldada por una significación biológica.
2. **Tipo II**, son puntos matemáticos cuya supuesta homología de un individuo a otro es respaldada únicamente por la geometría y no por evidencia anatómica.
3. **Tipo III**, se refieren a puntos localizados en cualquier lugar a lo largo de un contorno o entre dos landmarks de tipo I o II. Comúnmente denominados semilandmarks.

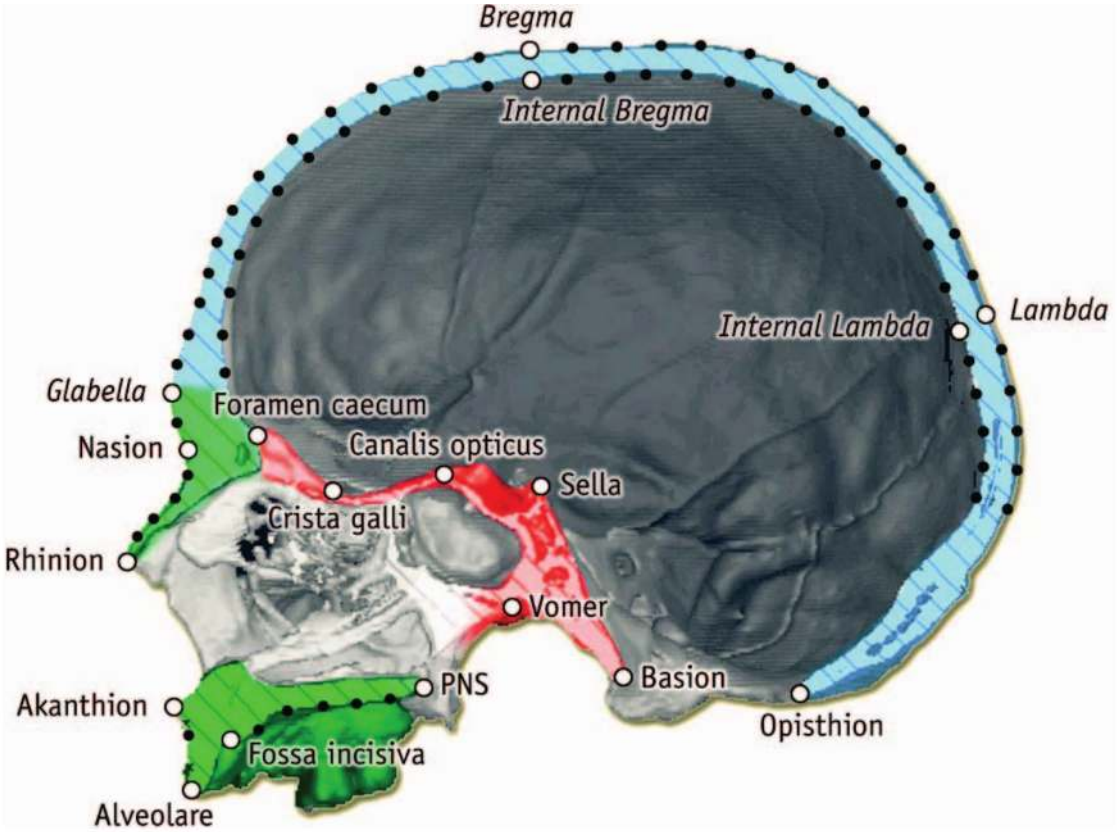


Figura 3.2: Ejemplo de configuración de landmarks (puntos blancos) y semilandmarks (puntos negros) (Bookstein et al. (2003)).

3.1.4. Tamaño y Forma en Morfometría Geométrica

La *forma* está definida como toda información geométrica que se mantiene invariante luego de eliminar efectos de traslación, rotación y escala sobre un objeto Kendall (1977). La forma se define como un conjunto de coordenadas, también llamado configuración de landmarks (ver la definición 3.1.1). Una implicación crucial de la definición de forma de Kendall es que, eliminando diferencias entre configuraciones que pueden ser adjudicadas a ubicación, escala y orientación, deja únicamente la información correspondiente a la forma Zelditch et al. (2004). Un ejemplo de los efectos de este proceso se puede ver en la Figura 3.3.

Definición 3.1.1. La configuración es un conjunto de landmarks sobre un objeto particular. La matriz de configuración X es una matriz de dimensión $k \times m$ de coordenadas cartesianas de k landmarks en m dimensiones. El espacio de configuraciones es el espacio de todas

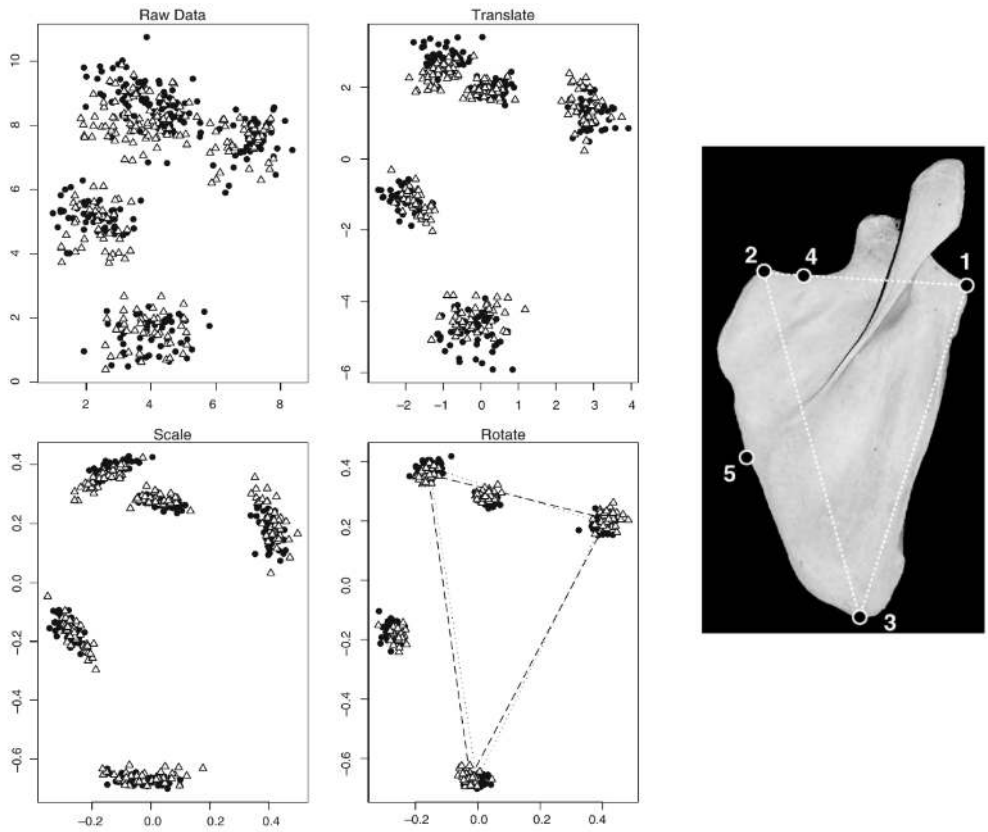


Figura 3.3: Superposición de datos de escápulas. El proceso incluye los datos tomados del digitalizador, trasladados al mismo origen, escalados a la misma unidad y orientados en la misma dirección Slice (2005).

las coordenadas posibles.

La definición de *forma* de Kendall menciona la escala como uno de los efectos que deben eliminarse para analizar diferencias de forma entre dos configuraciones. El concepto subyacente de la escala geométrica es simple e intuitivo si lo consideramos desde un punto de vista gráfico (los landmarks se encontrarán más distanciados entre sí en una configuración que en otra). Ahora bien, debemos definir a qué nos referimos con *tamaño*, para poder ser eliminado de la configuración. Consideremos que X es una matriz $k \times m$ de coordenadas cartesianas de k landmarks en m dimensiones reales (la matriz de configuración del objeto). Podemos entonces dar la siguiente definición:

Definición 3.1.2. Una medida de tamaño $g(X)$ es una función que devuelve un valor real

positivo tomando como entrada la matriz de configuración tal que:

$$g(aX) = ag(X) \tag{3.1}$$

para cualquier valor escalar positivo a.

Antes de calcular la escala geométrica, necesitamos determinar la ubicación del centro de la forma, y calcular las distancias desde el mismo a cada landmark de la configuración. A continuación, se calcula la escala geométrica, también llamada "tamaño del centroide".

Definición 3.1.3. El tamaño del centroide $S(X)$ está definido de la siguiente manera:

$$S(X) = \|CX\| = \sqrt{\sum_{i=1}^{k}\sum_{j=1}^{m}(X_{ij} - \bar{X}_j)^2}, X \in \mathbb{R}^{km}, \tag{3.2}$$

donde X_{ij} es la entrada (i, j) de X y el promedio aritmético de la dimensión jth es $\bar{X}_j = \frac{1}{k}\sum_{i=1}^{k}$. C corresponde a la matriz de *centring*

$$C = I_k - \frac{1}{k}1_k 1_k^T \tag{3.3}$$

$\|X\| = \sqrt{trace(X^T X)}$ es la norma Euclideana, I_k es la matriz identidad de $k \times k$ y 1_k es el vector de 1s $k \times 1$.

La Ecuación 3.1 satisface la propiedad $S(aX) = aS(X)$. El tamaño del centroide $S(X)$ es la suma de la raíz de los cuadrados de las distancias Euclideanas para cada landmark hacia el centroide.

$$S(X) = \sqrt{\sum_{j=1}^{k}\left\|(X)_j - \bar{X}\right\|^2}, \tag{3.4}$$

donde $(X)_j$ es la $j - esima$ fila de $X(j = 1, \cdots, k)$ y $\bar{X} = (\bar{X}_1, \cdots, \bar{X}_m)$ es el centroide.

3.2. Análisis de Procrustes

El método de Procrustes es útil para estimar una *forma promedio* y explorar la variación de la forma (conceptos desarrollados en la Sección 3.1.4) a lo largo del conjunto de datos. El análisis de Procrustes está formulado como una serie de superposiciones, colocando cada configuración encima de la otra en posiciones óptimas de acuerdo a la distancia Euclidiana mediante rotación, traslación y escalado. Comenzaremos describiendo un análisis de Procrustes Simple, utilizado para la "alineación" de sólo dos configuraciones y posteriormente extenderemos la definición para n configuraciones.

3.2.1. Análisis de Procrustes Simplificado

Dadas dos configuraciones X_1 y X_2 de dimensiones $k \times m$, queremos alinear ambas configuraciones lo más cerca posible, en este caso asumiremos que ambas configuraciones se encuentran centradas.

Definición 3.2.1. El método Simple de Procrustes se encarga de aproximar dos configuraciones mediante el uso del ajuste por mínimos cuadrados utilizando transformaciones afines. La estimación de los parámetros afines γ, Γ y β mediante la minimización de la distancia euclidiana al cuadrado.

$$D^2(X_1, X_2) = \left\| X_2 - \beta X_1 \Gamma - 1_k \gamma^T \right\|^2 \tag{3.5}$$

Donde $\|X\| = \{trace(X^T X)\}^{\frac{1}{2}}$ es la norma euclidiana, Γ es la matriz de rotación $(m \times m)$, $\beta > 0$ es el factor de escala y γ es un vector $(m \times 1)$de ubicación. La minimización de la Ecuación 3.5 nos dará la alineación de las configuraciones X_1 y X_2 con la menor distancia posible.

Definición 3.2.2. El ajuste de Procrustes de X_1 en X_2 es:

$$X_1^P = \hat{\beta} X_1 \hat{\Gamma} + 1_k \hat{\gamma}^T \tag{3.6}$$

Donde usamos P como denominación de superposición de Procrustes. La matriz residual luego del ajuste de Procrustes se define como:

$$R = X_2 - X_1^P \tag{3.7}$$

La matriz residual nos permite obtener información respecto a la diferencia de las formas, por ejemplo, considerar si existe un área en particular del objeto que muestra la variación.

3.2.2. Análisis Generalizado de Procrustes (GPA)

El Análisis Generalizado de Procrustes (GPA) fue desarrollado por Gower (1975) y perfeccionado en trabajos ulteriores Goodall and Mardia (1991), Ten Berge (1977), Langron and Collins (1985), Rohlf and Slice (1990).

Ahora veamos el mismo proceso visto en la Sección anterior aplicado a n configuraciones. Consideremos un caso genérico donde contamos con $n \geq 2$ configuraciones de matrices $X_1, \cdots X_n$ (definidas en 3.1.1). Estos n elementos forman parte de una muestra aleatoria tomada de una población con media μ y deseamos estimar la *forma promedio* de la población $[\mu]$. Tomemos tres modelos de perturbación para la configuración $k \times m$ de la matriz X_i.

$$X_i = \mu + E_i, \tag{3.8}$$

$$X_i = (\mu + E_i)\Gamma_i + 1_k\gamma_i^T, \tag{3.9}$$

$$X_i = \beta_i(\mu + E_i)\Gamma_i + 1_k\gamma_i^T, \tag{3.10}$$

donde E_i son matrices de error $k \times m$ con media igual a cero, μ es la matriz de la configuración promedio (la forma de $k \times m$), y β_i, Γ_i y γ_i son parámetros con ruido para las transformaciones de escala, rotación y traslación.

Se puede estimar de forma directa a μ desde la Ecuación 3.8, la cual se refiere a la medida del error del modelo, aunque este modelo es rara vez aplicable en la práctica. En cambio, no se pueden estimar todos los valores de μ con las Ecuaciones 3.9 y 3.10, pero se puede estimar $[\mu]$ o el tamaño y forma de $[\mu]_S$ siguiendo la Ecuación 3.9 y podemos estimar la forma $[\mu]$ siguiendo la Ecuación 3.10. Para realizar el análisis de forma no es necesario que nuestros objetos estén ajustados a escala, por lo que pueden ser utilizadas las Ecuaciones 3.9 y 3.10. En cambio, si los análisis que se desean realizar son de tamaño y forma es necesario utilizar el modelo de la Ecuación 3.9.

Definición 3.2.3. El método completo de GPA implica trasladar, re-escalar y rotar todas las configuraciones de forma relativa entre ellas, minimizando la suma total de las diferencias al cuadrado:

$$G(X_1, \cdots, X_n) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=i+1}^{n}\left\|(\beta_i X_i\Gamma_i 1_k\gamma_i^T) - (\beta_j X_j\Gamma_j 1_k\gamma_j^T)\right\|^2, \tag{3.11}$$

sujeto a restricciones del tamaño promedio dado por,

$$S((\bar{X})) = 1, \tag{3.12}$$

donde $\Gamma_i \in SO(m), \beta_i > 0, \|X\| = \sqrt{trace(X^T X)}$ y $S(X)$ es el tamaño del centroide y la configuración promedio es

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n}(\beta_i X_i \Gamma_i + 1_k \gamma_i^T). \tag{3.13}$$

Llamamos a $G(X_1, \cdots, X_n)$ para el mínimo de la Ecuación 3.11, sujeto a restricciones dadas por la Ecuación 3.12. $G(X_1, \cdots, X_n)$ es llamado *suma cuadrada generalizada de Procrustes* y $SO(m)$ es un conjunto de rotaciones especiales ortogonales. La restricción de la Ecuación 3.12 previene que $\hat{\beta}_i$ se aproxime a cero.

Las coordenadas completas de Procrustes de cada configuración X_i están dadas por:

$$X_i^P = \hat{\beta} X_i \hat{\Gamma}_i + 1_k \hat{\gamma}_i^T, i = 1, \cdots, n, \tag{3.14}$$

donde $\hat{\Gamma}_i \in SO(n)$ (matriz de rotación), $\hat{\beta}_i > 0$ (parámetro de escala), $\hat{\gamma}_i^T$ (parámetro de ubicación), con $i = 1, \cdots, n$ son los parámetros que minimizan.

3.2.3. Algoritmo para aplicar GPA

1. **Traslación**: Se centran las configuraciones para eliminar la ubicación individual de cada una. Inicialmente:

$$X_i^P = X_i, i = 1, \cdots, n. \tag{3.15}$$

2. **Rotación**: Para la configuración *i-esima* será:

$$\bar{X}_{(i)} = \frac{1}{n-1}\sum_{j \neq i} X_j^P, \tag{3.16}$$

por lo que el nuevo X_i^P es tomado como la superposición de Procrustes básica, teniendo sólo en cuenta rotación del viejo X_i^P en $\bar{X}_{(i)}$. Las n figuras son rotadas en este paso. Este paso se repite hasta que la Ecuación 3.11 converja, es decir, que no se pueda reducir más.

3. **Escalado**: Sea Φ la matriz de correlación $n \times n$ del vector $vec(X_i^P)$ con el autovector $\phi = (\phi_1, \cdots, \phi_n)^T$ correspondiente al autovalor más grande. Luego a partir de Ten Berge (1977)):

$$\hat{\beta}_i = \left(\frac{\sum_{k=1}^{n} \left\| X_k^P \right\|^2}{\left\| X_i^P \right\|^2} \right)^{\frac{1}{2}} \phi_i \tag{3.17}$$

La cual es repetida para todos los i.

3.3. Aplicaciones en Morfometría Geométrica

En esta Sección presentaremos varias investigaciones llevadas a cabo con técnicas de morfometría geométrica para un diverso campo de aplicaciones.

Análisis de cráneos de gorilas: O'Higgins and Dryden (1993) desarrolló un análisis sobre cráneos de gorilas para estudiar dimorfismo sexual. En este trabajo se diseño una configuración de 8 landmarks anatómicos y una muestra de cráneos de gorilas correspondientes a 29 machos y 30 hembras.

Análisis de Escaneos MR cerebrales sobre pacientes esquizofrénicos: Bookstein (1996) consideró 13 landmarks tomados en 2D de varias capas de una Resonancia magnética de 14 pacientes esquizofrénicos y 14 pacientes control. En este trabajo se buscaban discriminantes de promedios de forma o variabilidad de la misma.

Análisis de Imágenes: Anderson (1997) trabajó con reconocimiento de código postal. En esta investigación se recolectaron códigos postales escritos a mano, con base en landmarks matemáticos y pseudo-landmarks sobre imágenes colocados manualmente. Se analizó el promedio de la forma y su variabilidad, las cuales fueron utilizadas como un modelo previo para el reconocimiento de dígitos.

Reconocimiento de patrones en imágenes: Cootes et al. (1995) describe un método para la creación de modelos que aprenden a partir de la variabilidad de elementos de un conjunto de imágenes anotadas, el cual puede ser utilizado para búsquedas de patrones genéricas en varios problemas, desde formas de manos hasta transistores. Este método es llamado *Active shape models* ASM, el cual fue detallado en la Sección 2.1

3.4. Conclusiones

Como se puede observar, la Morfometría Geométrica tiene gran poder descriptivo de forma, el cual puede ser aplicado a innumerables problemas de reconocimiento de estructuras Anderson (1997), Cootes et al. (1995), Cooper et al. (1991), Lindner et al. (2013), y dentro de un mismo tipo de elemento para definir variabilidad dentro de la misma

población Valladares et al. (2010), Bookstein (1997), Valentin et al. (2002). En este trabajo utilizamos esta metodología para identificar personas dentro de un grupo conocido (ver Capítulo 6), determinar el género de personas en imágenes laterales (ver Capítulo 7) y para determinar y caracterizar la forma 3D de diferentes individuos, obteniendo, por ejemplo, el esqueleto a partir de una malla poligonal (ver Capítulo 8). Por otro lado, una de las restricciones más importantes para poder utilizar Morfometría Geométrica en problemas abiertos está en el ingreso masivo de datos de calidad, lo cual podría realizarse en forma automática utilizando aprendizaje de máquina, contando con un conjunto de entrenamiento supervisado confiable. Sobre esa temática nos enfocaremos en el siguiente capítulo.

Capítulo 4

Deep Learning

4.1. Introducción

El aprendizaje profundo o *Deep learning* permite que modelos computacionales compuestos por varias capas de procesamiento puedan aprender representaciones sobre datos con multiples niveles de abstracción y, mediante ese concepto, descubrir representaciones precisas de forma autónoma en grandes volúmenes de datos. El Deep Learning (Deep Learning (DL)) ha logrando recientemente grandes avances en el ámbito de la ciencia, incluyendo el reconocimiento de imágenes Tompson et al. (2014), Krizhevsky et al. (2012a), Farabet et al. (2013), Szegedy et al. (2015), reconocimiento del habla Mikolov et al. (2011), Hinton et al. (2012), el análisis de datos del acelerador de partículas Sadowski et al. (2014), la reconstrucción de circuitos del cerebro Helmstaedter et al. (2013), predicción de efectos en mutaciones *npn-coding* DNA en expresión genética y enfermedades Leung et al. (2014), Xiong et al. (2015) y muchos otros contextos más.

En DL, cada capa subsecuente extrae progresivamente representaciones más abstractas a partir de los datos de entrada y crea una nueva representación para ser utilizada por la siguiente capa. Un ejemplo de esto se puede ver en la Figura 4.2. Las capas más altas en la jerarquía amplifican características importantes para su discriminación, que quizás fueron pasadas por alto en el análisis supervisado. Por ejemplo, en una imagen, podrían ser la ocurrencia de bordes en cierta orientación, la siguiente capa podría detectar bordes en una disposición específica y la siguiente capa, probablemente identificaría estos bordes como partes de un objeto en particular.

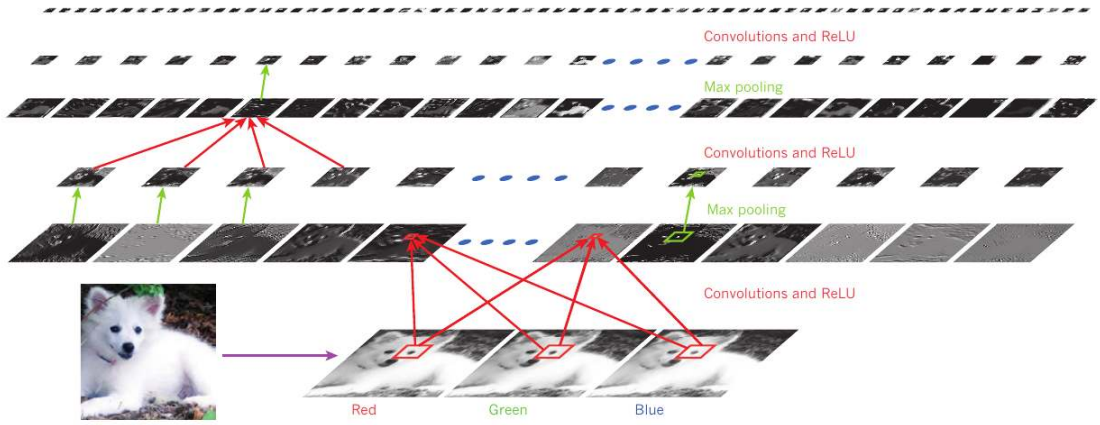


Figura 4.1: CNN para clasificación de raza de perros LeCun et al. (2015).

Un caso particular de DL son las redes convolucionales o *Convolutional neural networks* (CNN) Fukushima (1980), LeCun et al. (1998), las cuales constituyen actualmente el estado del arte de varios problemas de visión computacional, dado su buen desempeño de clasificación en grandes volúmenes de imágenes Toshev and Szegedy (2014), Krizhevsky et al. (2012b), Dieleman et al. (2015a). Su gran rendimiento está basado en el uso de cuatro conceptos fundamentales: conexiones locales, pesos compartidos, *polling* y el uso de varias capas LeCun et al. (2015). Por otro lado, dada la gran cantidad de grados de libertad de un modelo de aprendizaje en ConvNets, en ciertos problemas como los que nos ocupan en esta tesis la red debe aprender millones de parámetros, por lo que se deben tomar precauciones para evitar *overfitting* (la red puede simplemente memorizar todos los ejemplos y tener una pobre generalización).

Otra limitación, de las técnicas clásicas de *machine learning* es su escasa capacidad para procesar datos sin una preparación previa (*raw form*). Para construir sistemas de *pattern matching* o *machine learning* siempre ha sido necesario crear un extractor de características que transforme los valores crudos en un vector de características para representar el problema en un espacio de características que sea adecuado (*feature space*). Por lo tanto, se requiere tradicionalmente realizar una ingeniería previa que transforme los datos de su formato original al espacio de características. En ese sentido, una ventaja adicional de las CNN es su capacidad de trabajar directamente con los datos sin preparación previa, en caso de una imagen, los píxeles LeCun et al. (2015).

Una arquitectura de red basada en DL es una pila de múltiples capas simples, en

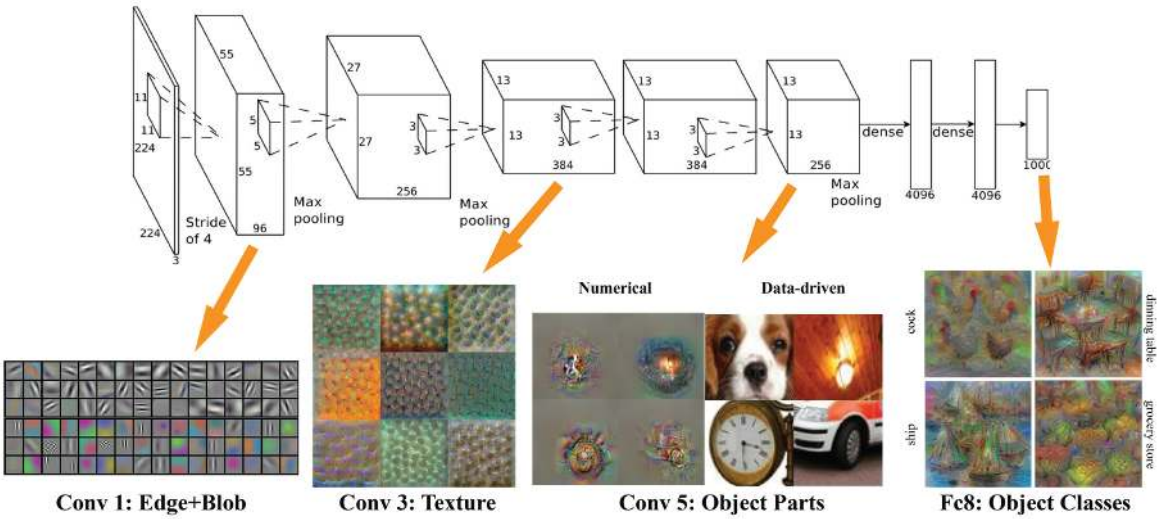


Figura 4.2: Diferentes niveles de abstracción de la representación a medida que se avanza en la pila de capas Mahendran and Vedaldi (2015).

las cuales gran parte de ellas calculan un mapeo no lineal entre la entrada y la salida. Cada capa en la pila transforma su entrada para incrementar la selectividad y al mismo tiempo su invariancia sobre su representación. Con una arquitectura de entre 5 y 15 capas de profundidad, una red es capaz de implementar funciones muy complejas sobre sus entradas, que son simultáneamente sensibles a detalles minuciosos e insensibles a grandes variaciones irrelevantes. En el caso de imágenes, la red puede aprender a ser indiferente a elementos irrelevantes como el fondo, la postura de las personas, variaciones en la iluminación y parámetros de la cámara, y la presencia de objetos aledaños.

El *Aprendizaje de representación* es un conjunto de métodos que permite descubrir automáticamente representaciones a partir de los datos sin preparación previa, las cuales son útiles para la detección y clasificación. Los métodos de DL son aprendizajes de representación con varios niveles o capas de representación, que se construyen mediante varios módulos no lineales, cada uno de estos transforman la representación de un nivel inicial (*raw data*) en una representación cada vez más abstracta. Un aspecto clave en DL es que las distintas abstracciones sobre los datos son inferidas desde los datos originales usando procedimientos genéricos de aprendizaje.

La forma más utilizada de aprendizaje de máquina, sea DL o no, es el aprendizaje supervisado. En el caso de visión por computadora, durante el entrenamiento, la computadora

recibe imágenes y genera un vector de puntajes como salida, uno por cada categoría que contiene. Luego se calcula una *función objetivo* que mide el error entre la salida del vector de puntajes calculados por el método y el real. El algoritmo modificará sus parámetros internos para reducir este error. Estos parámetros son comúnmente llamados *pesos*, que se pueden pensar como controles que nos permiten modelar la función de entrada/salida de nuestra máquina. Normalmente, en DL, se cuenta con miles de millones de estos pesos, y entre miles y millones de ejemplos etiquetados para su entrenamiento. Para ajustar debidamente estos pesos, el algoritmo de aprendizaje calcula un vector de gradiente, que por cada peso, indica la cantidad por la que el error se incrementará o decrementará si el valor de este peso se incrementara en una pequeña cantidad. Luego, el vector de pesos se ajusta en la dirección opuesta al vector de gradiente.

Este método básico de *descenso por el gradiente* ha sido aplicado exitosamente en muchos contextos de aprendizaje, y existen modelos matemáticos que garantizan su convergencia bajo supuestos razonables. La función objetivo, promediada sobre todos los elementos de entrenamiento, puede verse como una superficie con picos en el espacio de valores correspondientes a los pesos. El vector de gradiente negativo indica la dirección de la máxima pendiente en la superficie, de esta manera lo lleva más cerca del mínimo en donde el error de salida es más bajo en promedio (sobre toda la muestra de ejemplos).

En la práctica, el método más utilizado es *Stochastic Gradient Descent (SDG)*, que será explicado en detalle en la Sección 4.3.1, pero que consiste en mostrar el vector de entrada a solo un subconjunto de ejemplos de entrenamiento, se calculan las salidas, errores y conforme a estos se modifican los pesos de la red para estos valores. Este proceso se repite para varios subconjuntos de ejemplos hasta que el promedio de la función objetivo deja de decrecer. Su origen estocástico se debe a que cada pequeño subconjunto de ejemplos da un estimado ruidoso sobre el promedio general sobre todos los ejemplos. Luego que la red fue entrenada, se evalúa su performance sobre un grupo de ejemplos llamado subconjunto de prueba. Esta evaluación nos permite analizar qué tan bien generaliza la red, y cómo genera respuestas sobre ejemplos previamente desconocidos durante el entrenamiento.

4.2. Aprendizaje Supervisado

El aprendizaje es útil cuando deseamos que un algoritmo aprenda a ejecutar una función tan compleja que no puede ser programada por las vías convencionales. Por ejemplo, no está claro como realizar un programa que reconozca e identifique el sentido semántico de una escena a partir de una imagen. Pero lo que si se puede obtener es una gran cantidad de muestras de imágenes etiquetadas con objetos y sus relaciones para entrenar de forma supervisada un algoritmo que aproxime la relación de los datos entrada-salida implicados por los datos recolectados.

Ahora definamos formalmente un problema de aprendizaje supervisado. Sea X el espacio de entrada e Y el de salida, y D la distribución de valores sobre $X \times Y$ que describe los datos que se observan. Por cada tupla (x, y) perteneciente a D, la variable x es la entrada, e y el valor de salida deseado. El objetivo del aprendizaje supervisado es usar los datos de entrenamiento con n muestras de elementos Independent and identically distributed random variables (IID), sea S = $(x_i, y_i)_{i=1}^{n} \sim D^n$, para encontrar una función $f : X \to Y$ cuyo error de testeo sea el menor posible Sutskever (2013):

$$Test_D(f) = E_{(x,y)\sim D}[L(f((x); y)]. \tag{4.1}$$

Tomemos $L(z; y)$ como una función de pérdida, que mide el costo en el que se incurre cada vez que se predice y como z. Cuando encontremos una función cuyo error de testeo es lo suficientemente pequeño, el problema de aprendizaje se considera resuelto. Lo ideal sería encontrar un minimizador global que encuentre el menor error de testeo, pero dado que esto en general no es posible, podemos aproximar el error de testeo con el error de entrenamiento.

$$Train_S(f) \equiv E_{(x,y)\sim S}[L(f((x); y)] \approx Test_D(f), \tag{4.2}$$

donde S es una distribución uniforme sobre los elementos del conjunto de entrenamiento incluyendo repeticiones de los mismos. Se pretende encontrar el f que tenga menor error de entrenamiento posible. Este problema no se resuelve con la idea trivial de memorizar los ejemplos, dado que no garantiza que la red trabaje bien sobre elementos que todavía no observó. El problema de la *generalización* consiste específicamente en lograr que el modelo pueda predecir adecuadamente casos o ejemplos desconocidos. Conceptualmente

se puede resolver este problema reduciendo la cantidad de f permitidas, definiendo un grupo de funciones $\mathcal{F}$:

$$f^* = \underset{f \in \mathcal{F}}{\operatorname{argmin}} \, Train_S(f). \tag{4.3}$$

Lamentablemente, no se cuenta con una guía de como elegir un buen $\mathcal{F}$ para un determinado problema de aprendizaje. En la práctica es mejor experimentar con clases de funciones que hayan sido exitosas en problemas similares.

4.3. Redes Neuronales

En la Sección 4.2 se introdujo al problema general de aprendizaje supervisado. En esta sección nos enfocaremos particularmente a este sobre redes neuronales. Dado un problema en el cual tenemos acceso a datos etiquetados del tipo (x_i, y_i). Las redes neuronales nos permiten definir hipótesis no lineales y complejas de la forma $h_{W,b}(x)$, donde W, b son los parámetros que podemos adaptar a nuestros datos (x_i, y_i).

Consideremos una red neuronal con N capas, las capas de entrada y salida están representadas respectivamente como X_0 y X_N, donde el vector X_{n-1} es la capa de entrada de n (con $n = 1, \cdots, N$). Si W_n es la matriz de pesos y b_n el vector de bias, la salida de la capa X_n, puede ser representada de la siguiente manera:

$$X_n = f(W_n X_{n-1} + b_n), \tag{4.4}$$

donde f es la función de activación (en nuestro caso en esta tesis se utilizó la *función lineal rectificada* $f(x) = \text{máx}(x, 0)$). Dado un problema de reconocimiento, la tarea del entrenamiento consiste en encontrar el conjunto de parámetros óptimos $\{W_n, b_n\}$ que minimizan el error de clasificación. Para determinar cómo estos parámetros se deberían modificar para reducir el error, se utiliza comúnmente un algoritmo de *Gradient Descent* ya mencionado y que será presentado en detalle en la Sección 4.3.1.

4.3.1. Optimización: Gradient Descent

Siguiendo con la nomenclatura utilizada en la sección anterior (4.3), definamos X_{true} a los valores de salida esperados correspondientes a los valores de entrada X_0. Durante la etapa de entrenamiento, todos los parámetros de la red son optimizados para lograr

aproximar lo más posible los valores de salida de X_n a X_{true}. La predicción del error la denotamos como $e(X_N, X_{true})$. El gradiente de $e(X_N, X_{true})$ se calcula teniendo en cuenta los parámetros del modelo $\{W_n, b_n\}$. Los valores de cada capa son actualizados mediante la toma de pequeños pasos en la dirección opuesta del gradiente:

$$W_n \leftarrow W_n - \eta \frac{\partial e(X_N, X_{true})}{\partial W_n}, \tag{4.5}$$

y

$$b_n \leftarrow b_n - \eta \frac{\partial e(X_N, X_{true})}{\partial b_n}, \tag{4.6}$$

donde, η es la *tasa de aprendizaje*, un hiperparámetro que controla el tamaño del paso hacia la convergencia.

Stochastic Gradient Descent

Los métodos batch, como el *limited memory BFGS* Liu and Nocedal (1989), utilizan la totalidad de los ejemplos de entrenamiento para calcular la siguiente actualización de parámetros de cada iteración, normalmente convergen muy bien en *local optima*. Por otro lado, también son fáciles de implementar (minFunc) ya que cuentan con muy pocos hiperparámetros que ajustar.

En realidad, calcular el costo y gradiente del conjunto completo de entrenamiento puede ser muy lento y hasta inviable si nos encontramos trabajando sobre una única computadora y todo el conjunto de datos requiere más espacio que el disponible en su memoria principal. Además, los métodos de optimización *batch* no son amenos cuando se trata de una modalidad *online*. Esto quiere decir que si nuestro modelo desea agregar nuevos datos dentro del entrenamiento es necesario reiniciar todo el proceso sin poder reutilizar o aprovechar nada de lo realizado anteriormente.

El método de optimización *Stochastic Gradient Descent* cubre los dos problemas. Se sigue el vector del gradiente negativo de la función objetivo, tal como se mostró en las Ecuaciones 4.5 y 4.6, pero en base a un pequeño subconjunto aleatorio del conjunto de datos de entrenamiento. Por lo tanto no se calcula este valor sobre el total de la muestra. El uso de SGD en redes neuronales es recomendable dado el alto costo del cálculo exhaustivo de la *back propagation* sobre todo el conjunto de datos de entrenamiento. El SGD puede reducir este costo y al mismo tiempo obtener una convergencia rápida Bottou (2010).

Sea η_k la taza de aprendizaje en la iteración k.
Sea θ El parámetro inicial.
while *Criterio de parada no encontrado* **do**
    Tomar un *minibatch* de m ejemplos del conjunto de entrenamiento $x_1, x_2, \cdots, x_m$ con sus target correspondientes y_j.
    Computar el estimado del gradiente: $\hat{g} \leftarrow +\frac{1}{m}\nabla_\theta \sum_i L(f(x_j, \theta), y_j)$;
    Aplicar la actualización: $\theta \leftarrow \theta - \eta\hat{g}$;
end

Algorítmo 1: Actualización de SGD en la iteración de entrenamiento k.

En SGD, la tasa de aprendizaje η suele ser mucho mas pequeña que en Batch Gradient Descent (BGD) dado que la variancia es mayor en la actualización. Elegir el valor de la tasa de aprendizaje y su comportamiento (cómo debe modificarse la tasa de aprendizaje a medida que se avanza en el entrenamiento) no es una tarea trivial. Una práctica clásica es utilizar una η pequeña y constante que otorgue una convergencia estable en el *epoch*[1] inicial. Particularmente, en esta tesis se decidió trabajar con un η que decrece en forma lineal a medida que los *epochs* avanzaban, mientras que el *momentum* (explicado en la siguiente Sección) se incrementa. Esto se puede definir como un vector linealmente espaciado entre $[start\_value, stop\_value]$ con n_epochs elementos. Los valores precisos varían dentro del fenotipo a trabajar, ya que está sujeto a los datos de entrada y la estructura de la red, por lo que los mismos serán definidos en la Parte III (aplicaciones).

Un último punto importante con respecto a las SGD es el orden en que los ejemplos del conjunto de datos de entrenamiento son procesados. Si los datos están ordenados de una forma significativa, esto puede inducir un sesgo que lleven a una convergencia deficiente. Por lo tanto, es una buena practica distribuir aleatoriamente los datos antes de cada *epoch*.

[1] Llamaremos *epoch* a la cantidad de iteraciones en etapa de entrenamiento en las cuales la red ha observado el conjunto de datos.

El Momentum y el Gradiente Acelerado de Nesterov

El método clásico del *Momentum* (MC) Polyak (1964), es una técnica para acelerar el descenso del gradiente acumulando un vector de velocidad en las direcciones que muestran una reducción constante en la función objetivo a través de varias iteraciones. Sea una función objetivo $f(\theta)$ a ser minimizada, el MC esta determinado por:

$$v_{t+1} = \mu v_t - \eta \nabla f(\theta_t) \tag{4.7}$$

$$\theta_{t+1} = \theta_t + v_{t+1}, \tag{4.8}$$

donde $\eta > 0$ es la tasa de aprendizaje, $\mu \in [0, 1]$, es el coeficiente de *momentum* y $\nabla f(\theta_t)$ es el gradiente sobre θt. Dado que las direcciones d de poca curvatura, son por definición, más lentas a la hora de cambiar su tasa de reducción, tenderán a persistir entre iteraciones y ser amplificadas por el MC.

El gradiente acelerado de Nesterov (1983) (NAG), al igual que MC, es un método de optimización de primer orden que garantiza una tasa de convergencia mejor al descenso del gradiente en ciertas situaciones. Cuando contamos con funciones de convexidad suave, el NAG logra una tasa de convergencia global del orden de $O(1/T^2)$ versus $O(1/T)$ que obtiene el descenso del gradiente clásico, con una constante proporcional al coeficiente de Lipschitz del derivado y una distancia Euclidiana al cuadrado de la solución. Usualmente no se considera al NAG como un tipo diferente de *momentum*, dado que su única diferencia con el MC está en la actualización del vector de velocidad v. La actualización NAG se puede escribir como:

$$v_{t+1} = \mu v_t - \eta \nabla f(\theta_t + \mu v_t) \tag{4.9}$$

$$\theta_{t+1} = \theta_t + v_{t+1}. \tag{4.10}$$

La comparación de MC y NAG a la hora de la actualización de valores se puede observar en la Figura 4.3.

Desde un aspecto teórico, las convergencias para ambos métodos dependen en que los estimados del gradiente no contengan ruido. En la práctica, tomando algunos recuados, es posible aplicarlos en entornos estocásticos. A partir de las Ecuaciones 4.7-4.10, podemos

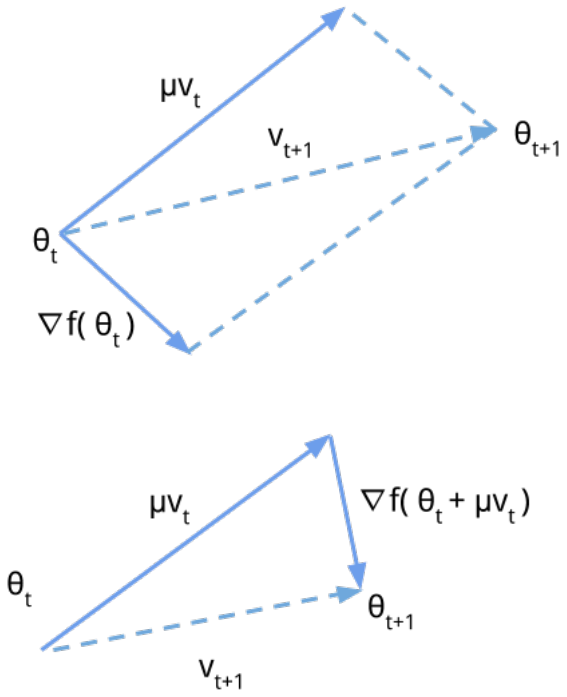


Figura 4.3: Actualización de valores por MC (arriba) y NAG (abajo).

observar que el cálculo de la nueva velocidad se hace aplicando una corrección basada en gradientes del vector de velocidad previo (el cual es disminuido) y luego se le suma la velocidad a θ_t. La diferencia entre estos métodos esta en que MC calculo el gradiente desde su posición actual, NAG calcula una actualización parcial de θ_t, calculando $\theta_t + \mu v_t$, el cual es similar a θ_{t+1} pero sin la corrección aplicada. De esta forma se le permite a NAG cambiar el vector v de forma más rápida y eficiente, siendo más estable que MC en varias situaciones, específicamente cuando μ toma valores altos. La implementación de ambos métodos se puede ver detallada en los Algoritmos 2 para MC y 3 para NAG.

4.3.2. Back-propagation

Cuando utilizamos redes neuronales *feedforward* que toman como entrada x y producen una salida $\hat{y}$, la información fluye hacia adelante sobre toda la red. La entrada x proporciona la infomación inicial, que se propaga sobre las unidades en cada capa oculta, y luego produce $\hat{y}$. Esto se denomina propagación hacia adelante (*forward propagation*).

Sea η la taza de aprendizaje y α el parámetro de momentum.
Sea θ El parámetro inicial y v velocidad inicial.
while *Criterio de parada no encontrado* **do**
    Tomar un *minibatch* de m ejemplos del conjunto de entrenamiento $x_1, x_2, \cdots, x_m$ con sus target correspondientes y_j.
    Computar el estimado del gradiente: $g \leftarrow +\frac{1}{m}\nabla_\theta \sum_i L(f(x_j, \theta), y_j)$;
    Computar la velocidad de actualización: $v \leftarrow \alpha v - \eta g$;
    Aplicar la actualización: $\theta \leftarrow \theta + v$;
end

Algorítmo 2: SGD con *momentum* clásico.

Sea η la taza de aprendizaje y α el parámetro de momentum.
Sea θ El parámetro inicial y v velocidad inicial.
while *Criterio de parada no encontrado* **do**
    Tomar un *minibatch* de m ejemplos del conjunto de entrenamiento $x_1, x_2, \cdots, x_m$ con sus target correspondientes y_j.;
    Aplicar actualización temporal(?): $\tilde{\theta} \leftarrow \theta + \alpha v$;
    Computar el gradiente sobre el punto temporal: $g \leftarrow +\frac{1}{m}\nabla_{\tilde{\theta}} \sum_i L(f(x_j, \tilde{\theta}), y_j)$;
    Computar la velocidad de actualización: $v \leftarrow \alpha v - \eta g$;
    Aplicar la actualización: $\theta \leftarrow \theta + v$;
end

Algorítmo 3: SGD con *Nesterov momentum*.

Durante el entrenamiento, este tipo de propagación puede continuar hasta que produce un costo escalar $J(\theta)$. El algoritmo de propagación hacia atrás o *back-propagation* Rumelhart et al. (1986), Le Cun (1986) permite que la información sobre el costo/error pueda fluir hacia atrás en la estructura de la red, para poder calcular el gradiente adecuado para entrenamiento. El termino *back-propagation* es normalmente interpretado incorrectamente como el responsable de todo el algoritmo de aprendizaje para una red de varias capas, tipo perceptron. En realidad *back-propagation* sólo se refiere al método utilizado para calcular el gradiente, mientras otro tipo de algoritmo, como por ejemplo el *Stochastic gradient*

descent (ver Sección 4.3.1), hace uso del gradiente calculado para el aprendizaje. Es más, el *back-propagation* suele ser asociado a un método específico de redes neuronales, pero en un principio puede computar las derivadas de cualquier otra función.

Específicamente, describiremos cómo calcular el gradiente $\nabla_x f(x, y)$ para una función arbitraria f, donde x es un conjunto de variables que se desea calcular las derivadas e y es un conjunto de variables adicionales, de las cuales no se desea saber sus derivadas. En algoritmos de aprendizaje, el gradiente que se desea calcular usualmente es el gradiente de la función de costo respecto a sus parámetros, $\nabla_\theta J(\theta)$. Otros métodos de aprendizaje requieren el cálculo de otras derivadas, ya sea para el proceso de aprendizaje como para el análisis del modelo entrenado.

Regla de la cadena

El algoritmo *Back-propagation* computa la regla de la cadena de la derivada, dadas las operaciones con un orden específico. Por dicha razón es un método muy eficiente. Sea x un número real, y f y g funciones que asignan un número real a otro número real $y = g(x)$ y $z = f(g(x)) = f(y)$. Luego, la regla de la cadena nos dice que:

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y}\frac{\partial y}{\partial x} \tag{4.11}$$

Esto puede ser generalizado más allá del caso escalar. Supongamos que $x \in \mathbb{R}^m$, $y \in \mathbb{R}^n$, g mapea de $\mathbb{R}^m$ a $\mathbb{R}^n$, y f mapea $\mathbb{R}^n$ a $\mathbb{R}$. Si, $y = g(x)$ y $z = f(y)$, entonces:

$$\frac{\partial z}{\partial x_i} = \sum_j \frac{\partial z}{\partial y_j}\frac{\partial y_j}{\partial x_i} \tag{4.12}$$

4.3.3. Sobreajuste (Overfitting)

Se denomina sobreajuste u *Overfitting* a la diferencia entre el error de entrenamiento $Train_S(f)$ y el error de testeo $Test_D(f)$, los cuales fueron definidos en la Sección 4.2. En esa misma Sección se mencionó el problema y una posible solución a la falta de generalización. Al restringir las funciones f que pertenecerían a la clase $\mathcal{F}$, controlamos el *overfitting* y tratamos el problema de generalización.

Teorema 1. *Si $\mathcal{F}$ es finito y la pérdida está acotada a $L(z; y) \in [0, 1]$, entonces el* overfitting *se encuentra uniformemente acotado con altas probabilidades sobre S del conjunto*

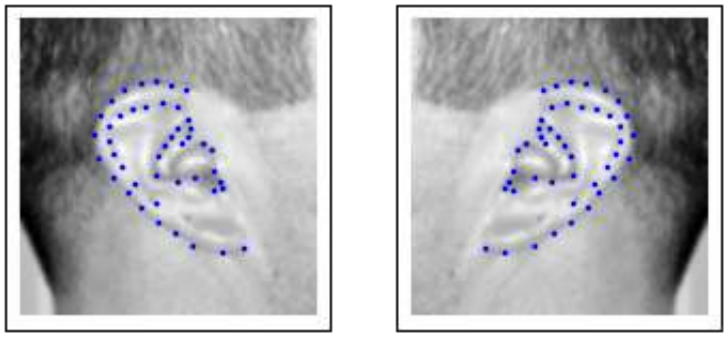


Figura 4.4: Una imagen de ejemplo y sus landmarks asociados, espejada sobre el eje x Cintas et al. (2016a).

de entrenamiento Kearns and Vazirani (1994), Valiant (1984).

$$Pr_{S\sim D^{|S|}}\left[Test_D(f) - Train_S(f) \leq \sqrt{\frac{\log|\mathcal{F}| + \log\frac{1}{\gamma}}{|S|}} \; for\ all\ f \in \mathcal{F}\right] > 1-\gamma$$

El resultado sugiere que cuando el conjunto de entrenamiento es más grande que $log|\mathcal{F}|$, cada error de entrenamiento estará cerca de cada error de testeo. El término $log|\mathcal{F}|$ justifica formalmente la intuición de que cada caso de entrenamiento lleva un número constante de bits sobre la mejor función de $\mathcal{F}$ Sutskever (2013).

4.3.4. Métodos para Optimización y regularización

Un problema crucial en aprendizaje automático es cómo entrenar un algoritmo para que funcione correctamente no sólo en los datos de entrenamiento sino con nuevas entradas. Varias estrategias han sido diseñadas para disminuir el error sobre los datos de prueba, posiblemente a expensas de incrementar el error de entrenamiento. Estas estrategias son conocidas como *regularización*. Existen varios tipos disponibles de regularización en *Deep Learning*, de hecho, la investigación de métodos eficientes de regularización ha sido una de las áreas más activas en este tópico. En esta Sección veremos una breve presentación de los métodos utilizados en esta tesis. Particularmente en nuestra aplicación, las CNNs cuentan con un gran número de parámetros para ser aprendidos, por ejemplo, en el caso del landmarking de orejas se utilizaron 8.622.970 de parámetros. Debido al limitado tamaño de nuestra muestra de entrenamiento, es alta la probabilidad de que ocurra el *overfitting*. A continuación detallaremos los métodos utilizados para tratar de reducir lo más posible el *overfitting* en nuestras redes destinadas a landmarking.

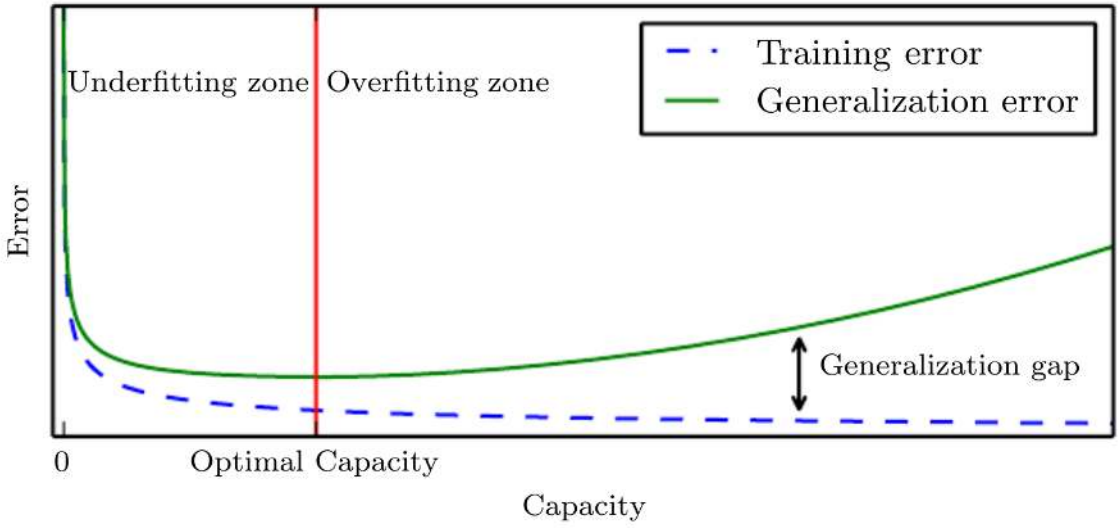


Figura 4.5: Análisis de curvas de error: *Underfitting*, *Overfitting*, Generalización y Capacidad Goodfellow et al. (2016).

Incremento de datos: Este método consiste en aumentar de forma artificial los datos de entrenamiento usando transformaciones que preserven el valor de *target*. De forma aleatoria, algunas de las imágenes y sus landmarks asociados son espejados horizontalmente y agregados al set de entrenamiento (un ejemplo se puede observar en Figura 6.5).

Regularización: La complejidad del modelo encontrado se penaliza mediante el uso de *dropouts* (Hinton et al., 2012), lo que consiste en llevar a cero la salida de cada nodo (de las capas ocultas) con una cierta probabilidad. Esta técnica reduce adaptaciones complejas entre nodos, dado que durante el entrenamiento un nodo no puede depender de la activación de nodos específicos (Krizhevsky et al., 2012b).

Dropout

El uso de Dropout es una técnica introducida por Hinton et al. (2012), Srivastava et al. (2014), la cual fue rápidamente adoptada dado que no sólo era una solución de buen desempeño sino que además es de muy fácil implementación. Con el uso de *Dropout* se previene que una red caiga en *overfitting*, y se provee una manera de combinar de forma eficiente varias arquitecturas de redes neuronales. El termino *dropout* se refiere al

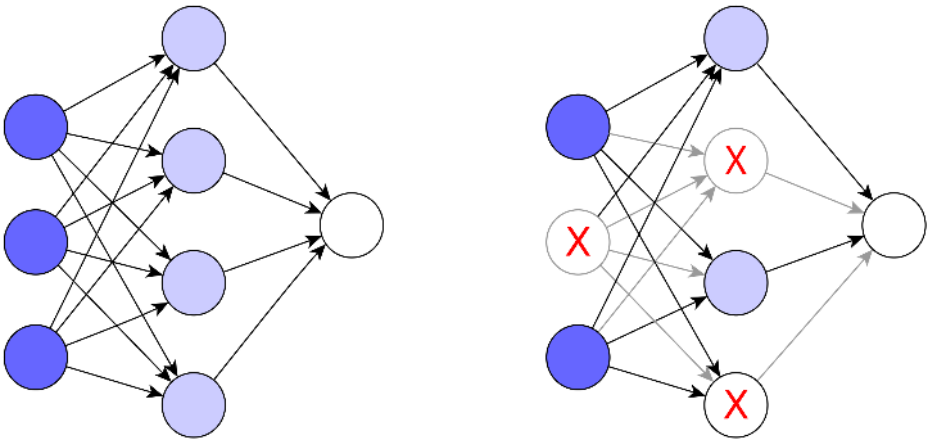


Figura 4.6: Comparación de una red neuronal convencional y con *Dropout*.

descarte de unidades y sus conexiones (ya sea en capas ocultas o no) en una red neuronal. Este descarte es sólo temporario. Una representación gráfica del *dropout* se puede ver en la Figura 4.6. La forma de elección de descarte es aleatoria. A cada unidad se le asocia un valor de probabilidad p independiente del resto entre $[0, 1]$. Algo importante a tener en cuenta es que si se le asignan probabilidades muy altas a todas las unidades comenzaremos a perder información valiosa para su clasificación.

En esta Sección describiremos brevemente el modelo de una red neuronal con *Dropout*. Dada una red neuronal de N capas como las definidas en la Sección 4.3, bajo la forma vista en la Ecuación 4.4. Al agregar dropout a esta red, las operaciones *feed-forward* se ven modificadas de la siguiente manera:

$$\begin{aligned} r_n^j &\sim Bernoulli(p), \\ \tilde{X}_n &= r_n * X_n, \\ X_{n+1} &= f(W_{n+1}\tilde{X}_n + b_{n+1}), \end{aligned} \tag{4.13}$$

donde $*$ es el producto de matriz, para cualquier capa n, r_n es el vector que cuenta con variables independientes y aleatorias de Bernoulli, cada una de estas posee una probabilidad p de ser 1. Este vector es multiplicado por los elementos de esa capa, X_n, para generar una nueva capa con una cantidad disminuida de elementos, aquí denominada, $\tilde{X}_n$. Esta capa es utilizada como entrada de la capa X_{n+1}, y así sucesivamente se aplica este proceso a todas las capas de la red Hinton et al. (2012), Srivastava et al. (2014). Un ejemplo gráfico de este proceso aplicado a una capa se puede observar en la Figura 4.7.

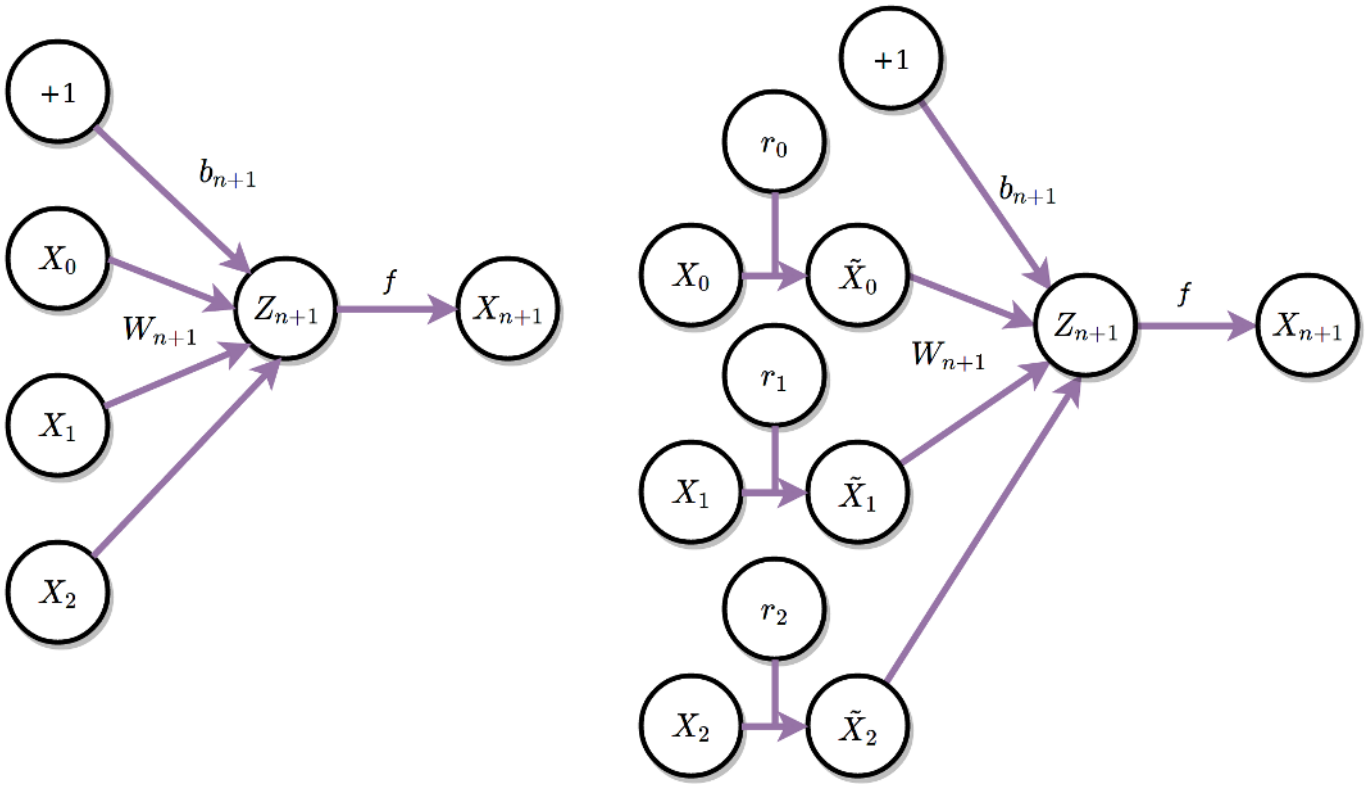


Figura 4.7: Comparación de operaciones en redes neuronales clásicas (izquierda) y con *dropout* (derecha).

Incremento de datos de entrenamiento

La manera fundamental para conseguir que los métodos de aprendizaje automático generalicen adecuadamente es entrenarlos con la mayor cantidad de datos posible. En la práctica la cantidad de datos es limitada. Esto es aún más dificultoso en nuestro problema particular de landmarking, dadas las restricciones de adquisición que enumeramos en la Sección 1.1. Una forma de ampliar nuestra cantidad de datos, es crear datos ficticios y agregarlos al conjunto de entrenamiento.

Para nuestro caso, podemos generar $n-uplas$ $(X_{train}, (x_0, y_0), (x_1, y_1), \cdots (x_n, y_n))$, donde X_{train} es un arreglo de píxeles correspondientes a nuestras imágenes y landmarks originales, luego de aplicar transformaciones geométricas, los pares (x_i, y_i) son las coordenadas de landmarks. La misma transformación que se aplica a la figura se debe aplicar a sus correspondientes landmarks, en general transformaciones afines. Se puede ver un ejemplo de las transformaciones implementadas en este trabajo en la Figura 4.8.

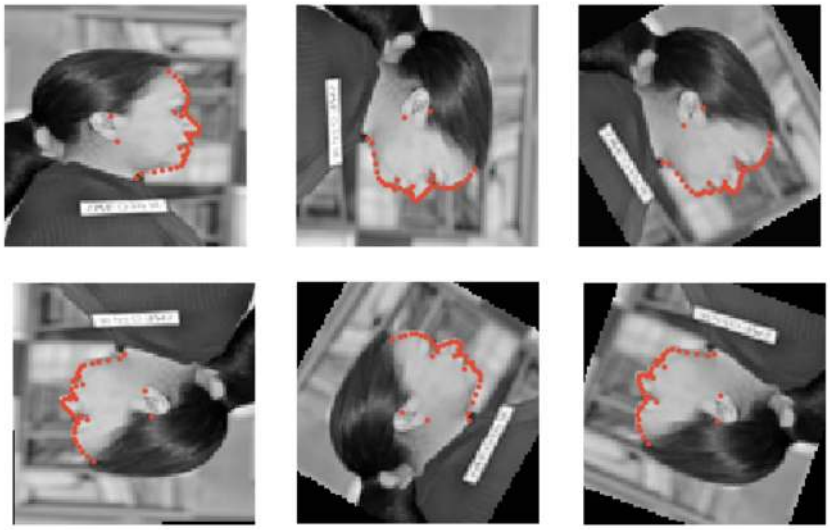


Figura 4.8: Diferentes ángulos de rotación utilizados para el aumento artificial de datos durante etapa de entrenamiento.

El aumento de datos de forma artificial ha sido una técnica efectiva sobre datos como imágenes, dado que las imágenes son datos de varias dimensiones e incluyen una gran cantidad de factores de variación, que en muchos casos pueden ser simulados fácilmente. Operaciones como la traslación y la rotación en diferentes direcciones, puede mejorar mucho la generalización. Además, se ha probado que la utilización de transformaciones como la rotación y escala son efectivas para el fin de generalización. En nuestra implementación, las imágenes transformadas son generadas con Python en el CPU mientras la GPU se encuentra entrenando el *batch* anterior de imágenes, lo cual nos permite la posibilidad de no tener que almacenar en disco estas nuevas imágenes, por lo que este tipo de métodos son de costo cero computacionalmente hablando Krizhevsky et al. (2012b).

Detención Temprana del Entrenamiento

Cuando se entrenan modelos con suficiente capacidad representativa para evitar el overfitting, se suele observar que el error de entrenamiento decrece de forma monotónica a medida que avanza el entrenamiento, pero el error de validación comienza a incrementarse. En la Figura 4.5 se muestra este fenómeno en las curvas de error de entrenamiento y de validación. Esto significa que podemos obtener un modelo con un mejor error de validación

utilizando el conjunto de parámetros en el punto de entrenamiento donde se encontró el menor error de validación. Cada vez que el error de validación mejora, se almacena una copia de los parámetros en ese momento. Cuando el entrenamiento finaliza, retornamos estos parámetros y no los últimos generados. El algoritmo finaliza cuando ningún parámetro ha mejorado sobre el mejor error de validación almacenado sobre una cantidad pre-acordada de iteraciones Prechelt (1998a,b). La implementación de este método se explica formalmente en el Algoritmo 4.

4.3.5. Selección y puesta en funcionamiento de un red neuronal

Funciones de activación

Un principio común que atraviesa el área de las Ciencias de la Computación es que se pueden construir sistemas complejos a partir de componentes mínimos y simples. De forma similar a la memoria de una máquina de Turing sólo necesita poder almacenar estados del tipo 0 o 1, es posible crear un aproximador de funciones que sea universal basándonos en funciones lineales rectificadas. La función de activación Lineal rectificada se describe como:

$$g(z) = Max\{0, z\} \tag{4.14}$$

La función de activación rectilínea es la más frecuentemente utilizada en las redes neuronales de tipo *feed forward* LeCun et al. (2015). Si bien al aplicar esta función a la salida de una transformación lineal, se genera una función no-lineal, ésta es lineal a trozos, y gracias a ello preserva varias propiedades de los modelos lineales que la hace sencilla de optimizar con métodos basados en gradientes (para ver en más detalle, dirigirse a la Sección 4.3.1).

Pre-procesamiento de los datos: Normalización de las entradas

La convergencia suele ser más rápida si el promedio de cada variable de entrada sobre el conjunto de entrenamiento es cercano a cero. Para entender esto, consideremos el caso extremo, en el que todas las entradas son positivas. Los pesos de un nodo de la primera capa son actualizados por una cantidad proporcional a $x\sigma$, donde σ es el error (escalar) en ese nodo y x es el vector de entrada. Cuando todos los elementos del vector x son positivos, todas las actualizaciones de pesos que alimentan a ese nodo van a tener

```
Sea n la cantidad de pasos entre evaluaciones.
Sea p la "paciencia", la cantidad de veces que se tolera errores de validación peores
  al almacenado y no se cancela la ejecución.
Sea θ_0 los parámetros iniciales.
θ ← θ_0;
i ← 0;
j ← 0;
v ← ∞;
θ* ← θ;
i* ← i;
while j < p do
    Actualizar θ entrenando sobre n pasos;
    i ← i + n;
    v' ← validation_error(θ);
    if v' <v then
        j ← 0;
        θ* ← θ;
        i* ← i;
        v ← v';
    else
        j ← j + 1;
    end
end
donde θ* son los mejores parametros y i* es la mejor cantidad de pasos de
  entrenamiento.
```

Algorítmo 4: Meta-algoritmo de *Early Stopping* para determinar mejor cantidad de tiempo de entrenamiento.

el mismo signo. Como resultado, estos pesos sólo pueden crecer o decrecer juntos para algún patrón de entrada determinado. Por lo tanto si un peso debe cambiar de dirección, sólo puede hacerlo mediante una trayectoria en zig-zag en el espacio de los pesos, lo cual

es ineficiente y muy lento Yann (1998). La convergencia es aún más rápida si además de trasladar nuestro datos a un centroide próximo a cero, éstos son escalados de manera que todos tengan el mismo valor de covariancia, C_i, donde:

$$C_i = \frac{1}{P}\sum_{p=1}^{P}(z_i^p)^2 \tag{4.15}$$

P es la cantidad de elementos en el conjunto de entrenamiento, C_i es la covariancia de la i-esima variable del vector y z_i^p es el componente i-esimo del p-esimo elemento del conjunto de entrenamiento. El escalado de datos acelera el aprendizaje dado que ayuda a balancear el valor de la tasa de aprendizaje. Por ello, se aplica una transformación sobre las entradas para que el promedio de cada variable de entrada en conjunto de entrenamiento sea cercano a cero, y que los valores de cada variable de entrada tengan similares covariancias.

Inicialización de pesos

El valor inicial de los pesos puede tener efectos significativos en el proceso de entrenamiento. Si los pesos iniciales de una red son muy pequeños, el vector de salidas de la primera capa tiene una norma pequeña, la cual se achica a medida que pasa por cada capa, hasta que es muy pequeña para ser útil. De forma inversa sucede si la inicialización de pesos cuenta con valores muy grandes. Particularmente se utilizó la inicialización propuesta en Glorot and Bengio (2010), la cual muestrea los valores de los pesos a partir de la distribución uniforme de la siguiente manera:

$$\begin{aligned} a &= \sqrt{\frac{2}{fan_{in} + fan_{out}}} \\ W &\sim U[-a, a], \end{aligned} \tag{4.16}$$

donde W es la distribución de inicialización, fan_{in} la cantidad de conexiones alimentando la entrada de un nodo y fan_{out} la de salida.

Funciones de pérdida

Un aspecto de diseño importante en una red neuronal (sea jerárquica o no) es la elección de la función de costo o pérdida. En general las funciones de pérdida para las redes neuronales son símiles a las de otros modelos paramétricos, como los modelos lineales.

4.4. Convolutional Neural Nets (CNNs)

Desde la introducción de ConvNets o CNN propuestas por LeCun et al. (1989), han demostrado tener un excelente desempeño en tareas como reconocimiento de dígitos y detección de rostros. En los últimos años, varios trabajos han demostrado que pueden tener un excelente desempeño sobre problemas más complejos en visión computacional Farabet et al. (2013), Szegedy et al. (2015), Krizhevsky et al. (2012b), Dieleman et al. (2015a), Donahue et al. (2017). Esto se debe a la disponibilidad de grandes cantidades de datos etiquetados para su uso durante entrenamiento y evaluación, a la implementación de librerías sobre tecnología GPU, permitiendo que el entrenamiento de grandes modelos sea viable en tiempo/capacidad computacional y al surgimiento de nuevas estrategias de regularización como *Dropout* Hinton et al. (2012) como se vio en detalle en la Sección 4.3.4.

En la siguiente sección daremos una introducción a los elementos que componen este tipo de redes: las capas de convolución y de *pooling*. Luego ensamblaremos estos componentes para detallar su interacción, utilizando como ejemplo la red precursora *LeNet* desarrollada en LeCun et al. (1998), basada en el *Neocogitron*.

4.4.1. Introducción

Las *ConvNets* están diseñadas para procesar datos que tienen la forma de arreglos multidimensionales, en nuestro caso imágenes. Como se mencionó en la Sección 4.1 hay cuatro conceptos básicos a tener en cuenta: conexiones locales, pesos compartidos, *pooling* y el uso de varias capas.

Una arquitectura clásica de *ConvNet* está dada por dos etapas. La primer etapa se enfoca en extraer características discriminantes a distintos niveles de abstracción, y la segunda se enfoca en la clasificación a partir de las características obtenidas previamente. La primera instancia está compuesta por dos tipos de capas: de convolución y *pooling*. En las capas de convolución, las unidades están organizadas en *feature maps* (mapas de atributos), en las cuales cada unidad esta conectada a parches locales de los mapas pertenecientes a la capa anterior mediante un conjunto de pesos, llamados banco de filtros. Todas las unidades dentro de un mapa comparten el mismo banco de filtros.

Distintos mapas dentro de la misma capa usan diferentes filtros. Esta disposición tiene dos justificaciones. Por un lado, los datos en forma de arreglos (imágenes en nuestro caso) tienden a estar altamente correlacionados localmente, y por otro lado la estadística local de las imágenes es invariante respecto de su ubicación.

La tarea de las capas de convolución es detectar conjunciones locales de características provenientes de la capa anterior, mientras que el rol de la capa de *pooling* es unificar semánticamente todas las características similares en una única característica. Comúnmente, una unidad perteneciente a una capa de *pooling* calcula el máximo de un parche local de unidades en uno o varios mapas. En la segunda etapa se cuenta con capas completamente conectadas (estas capas son similares a las que se encuentran en las redes neuronales convencionales) para establecer un ranking en la clase a la que pertenecerían todas las características encontradas en la etapa anterior.

4.4.2. Convolución

La convolución es una operación matemática sobre dos funciones f y g, que obtiene una tercera función h a partir de la integral del solapamiento de la función f sobre distintas traslaciones de g. Formalmente se define como:

$$h(t) = \int_{\text{ínf}}^{-\text{ínf}} f(\tau)g(t-\tau)d\tau, \tag{4.17}$$

la cual se denota simplificadamente como $h = f * g$. Además de ser una operación lineal, la convolucióne es el dual del producto bajo la transformada de Fourier, y por lo tanto tiene propiedades algebraicas similares al producto (asociatividad, conmutatividad, distributividad sobre la suma, etc.). En el procesamiento de imágenes la convolución se realiza entre toda la imagen (representada por la función f arriba) y un *kernel* g, usualmente de tamaño pequeño (por ejemplo 3×3 o 5×5. En la Figura 4.9 se pueden ver algunos ejemplos.

La convolución posee propiedades de invariancia frente a traslaciones y rotaciones, lo cual es crítico en problemas de visión computacional, dado que en un caso ideal, el resultado del reconocimiento no se debería ver afectado por este tipo de transformaciones en los objetos presentes en la imagen. Esta propiedad de invariancia fue determinante para el éxito de los algoritmos SIFT y SURF Lowe (1999), Bay et al. (2008) (explicados

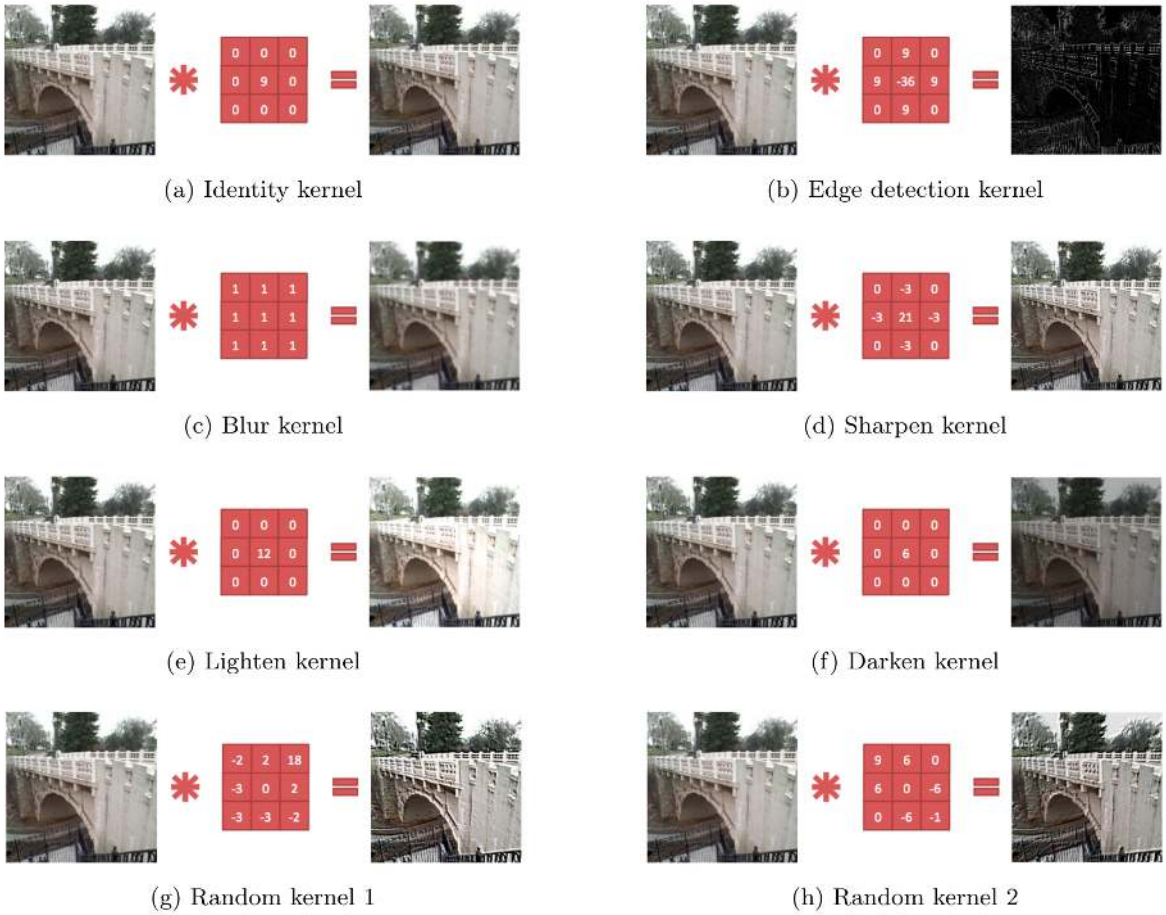


Figura 4.9: Ejemplos de distintos *kernels* de convolución aplicados a una misma imagen (en todos los casos, los coeficientes de las matrices se dividen por 9) Wang and Raj (2017).

en la Sección 2.2). El uso de convoluciones en redes neuronales permite aprovechar esta invariancia, obteniéndose una performance superadora del estado del arte.

4.4.3. Capa de Convolución

En las CNNs, una capa de convolución aplica operaciones de convolución sobre sus elementos de entrada. Este tipo de capas está parametrizada por un conjunto de filtros. Los mapas de características son tomados como entrada y se les aplica una convolución con un conjunto de filtros para producir una pila de mapas de características como salida. Esto puede ser implementado de forma eficiente sustituyendo el producto entre matriz-vector $W_n x_{n-1}$ en la Ecuación 4.4 con una suma de convoluciones Dieleman et al. (2015a).

La entrada de la capa n puede ser desdoblada como un conjunto de K matrices $X_{n-1}^{(k)}$, con $k = 1, \cdots, K$. Cada una de estas matrices representa diferentes entradas en forma de mapas de características. Los mapas de características de salida $X_n^{(l)}$, con $l = 1, \cdots, L$ son representados de la siguiente manera:

$$X_n^{(l)} = f\left(\sum_{k=1}^{K} W_n^{(k,l)} * X_{n-1}^{(k)} + b_n^{(l)}\right), \tag{4.18}$$

donde $*$ representa el operador de convolución en dos dimensiones. Las matrices $W_n^{(k,l)}$ representan los filtros de la capa n y $b_n^{(l)}$ las bias para el mapa de características l. El mapa de características $X_n^{(l)}$ se calcula como la suma de K convoluciones pertenecientes a los mapas de características de la capa anterior.

El término de sesgo o bias $b_n^{(l)}$ puede ser reemplazado por una matriz $B_n^{(l)}$. Así, cada posición espacial en el mapa de características tiene asociado su propio bias independiente. Al reemplazar el producto de matrices con una suma de convoluciones, la conexión de las capas se restringe de forma eficiente, permitiendo aprovechar la naturaleza de los datos de entrada y reducir el número de parámetros a entrenar. Cada unidad esta conectada a un subconjunto de unidades en la capa anterior, y cada una de estas unidades es replicada a lo largo de toda la entrada Dieleman et al. (2015a). Gracias a esta reducción de parámetros las CNN logran una mejor performance de generalización.

Kernels aleatorios no supervisados

El costo computacional más alto del entrenamiento de una CNN es sin duda, el aprendizaje de las características. Cuando se realiza un entrenamiento supervisado con SGD, cada paso del gradiente requiere una propagación hacia adelante y hacia atrás sobre toda la red. Una manera de reducir el costo del entrenamiento es usar características (configuraciones de kernels) no entrenadas de forma supervisada. Hay varias técnicas para esto. Una de ellas es inicializar de forma aleatoria, otra es diseñar los kernels de forma manual, o bien colocar kernels para detectar bordes en una cierta orientación o escala. Finalmente, la red también puede aprender en forma no supervisada qué kernels utilizar. Coates et al. (2011) aplica un clusterizado *k-means* sobre pequeñas partes de la imagen, y luego usa cada centroide aprendido como kernel de convolución. Los kernels aleatorios han generado resultados exitosos recientemente Jarrett et al. (2009), Saxe et al. (2011), Pinto et al.

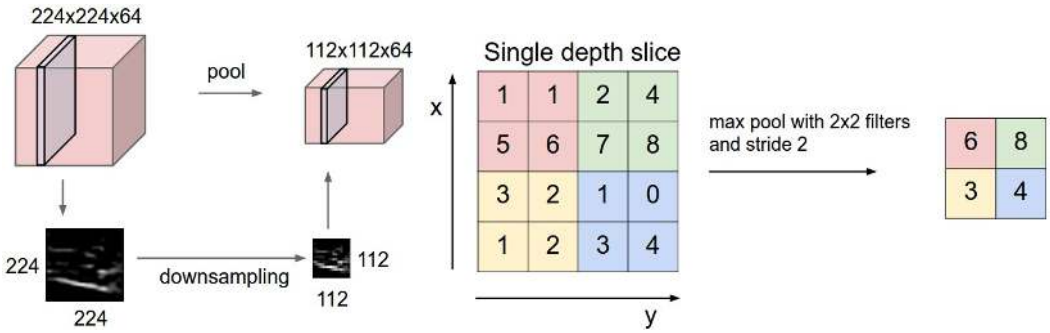


Figura 4.10: Ejemplo de resultados al agregar capas de *pooling* (http://cs231n.github.io/convolutional-networks/).

(2011), Cox and Pinto (2011). Saxe et al. (2011) mostró que una capa de convoluciones seguida por una de *pooling* se vuelven selectivos a frecuencias espaciales (pasabanda) e invariantes a la traslación cuando se les asignan pesos aleatorios.

4.4.4. Capa de pooling

Este tipo de capas realiza un submuestreo, es decir, obtiene una reducción (a veces drástica) de las dimensiones de la entrada generando una salida de menor tamaño. Si la entrada es de tamaño $n \times n$ y el *pool* es de $p \times p$, entonces se subdivide la entrada en $p \times p$ regiones de tamaño igual, y para cada celda del *pool* se elige un valor en función de su región correlativa de la entrada. Existen diferentes estrategias para este submuestreo. Una de ellas, conocida como *max-pooling*, toma el valor máximo de los valores de la región correlativa de entrada. Otra denominada *average-pooling* toma el promedio de dichos valores, de manera idéntica al submestreo de primer orden en procesamiento de imágenes. El método de *probabilistic pooling* simplemente toma un valor al azar de entre los valores en la región de entrada correlativa Lee et al. (2009).

Esta operación se utiliza para reducir la dimensionalidad de los mapas de características. Las capas de *pooling* se ubican entre las capas de convolución Boureau et al. (2010). La Figura 4.10 muestra un ejemplo de este proceso. El propósito principal de las capas de *pooling* es reducir el costo computacional en las capas ulteriores, reduciendo el tamaño de los futuros mapas de características y otorgando una forma de invariancia traslacional.

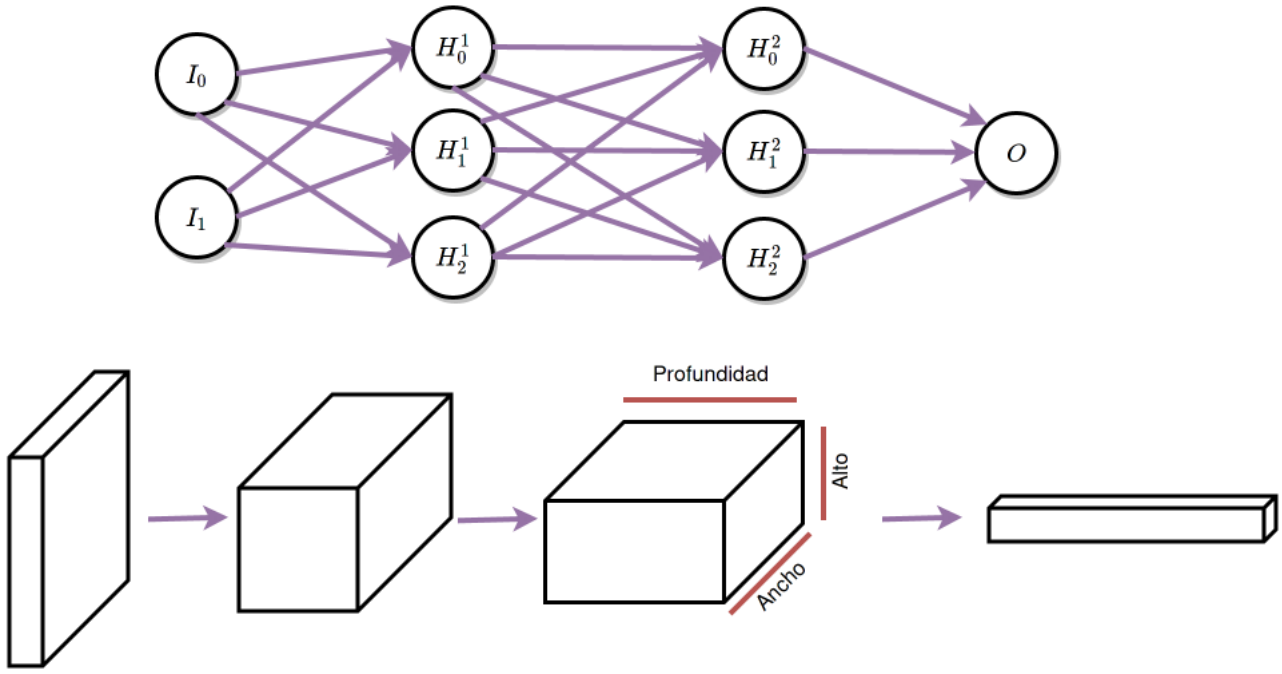


Figura 4.11: Comparación de Arquitecturas de redes del estilo CNN (abajo) y redes neuronales convencionales (arriba).

4.4.5. Arquitectura de CNN

Las arquitecturas de CNN supone que los datos con los que se trabajará son arreglos multidimensionales. Gracias a esta suposición varias cualidades de los datos se pueden embeber en la arquitectura de la red. A diferencia de las redes neuronales convencionales, las CNN cuentan con unidades organizadas en tres dimensiones: ancho, alto y profundidad. Por ejemplo, en nuestro caso tenemos imágenes de $96x96$ píxeles en escala de grises, por lo que las dimensiones de nuestro volumen de entrada serían $96x96x1$. Un ejemplo comparativo entre redes neuronales convencionales y CNN se puede ver en la Figura 4.11.

4.4.6. Visualización e introspección de CNN

Más allá de las capacidades de las ConvNets, hay muy poco desarrollo sobre cómo funcionan internamente estos complejos modelos y cómo es que logran su excelente desempeño. Sin tener una clara idea de cómo y por qué funcionan estos modelos, el desarrollo y mejoras están reducidos a ensayos por prueba y error. En los Capítulos 6 y 7 mostra-

remos técnicas de visualización para evaluar el estímulo de cada mapa de características, y analizar cómo trabajan en distintas capas. En el Anexo A observaremos un análisis de sensitividad a oclusiones parciales Long et al. (2014), Simonyan et al. (2013).

4.4.7. Consideraciones Computacionales

El mayor cuello de botella a la hora de implementar una CNN es la memoria. Varias Graphic Process Unit (GPU)s tienen el límite de 3,4 o 6 GB. Hay tres aspectos principales que determinan el consumo de memoria y que se deben chequear antes de construir una red:

1. Volúmenes de memoria intermediarios. Son los números en crudo de las funciones de activación en cada capa. Estos valores son necesarios para el algoritmo de *backpropagation*.

2. Tamaño de parámetros. Estos valores representan los pesos de la red, sus gradientes durante el *backpropagation*, y normalmente un *cache* de los valores en la iteración anterior si se utiliza algún tipo de *momentum* como en nuestro caso.

3. Otros usos de memoria, como subconjuntos de imágenes, sus versiones aumentadas artificialmente, etc.

Parte III

Aplicaciones

Capítulo 5

Casos de estudio: problemas, modelos y materiales

5.1. Introducción

En los siguientes Capítulos se mostrarán distintos fenotipos obtenidos de forma automática mediante el uso de redes jerárquicas diseñadas y entrenadas en esta tesis. Además de la descripción, implementación y evaluación del modelo en cada caso para landmarking automático de diferentes estructuras biológicas, se muestra una posible aplicación del vector de características obtenido a partir de estas redes. En el Capítulo 6 se utiliza el pabellón auditivo para reconocimiento de personas. En el Capítulo 7 se emplea la configuración de landmarks de la vista lateral para determinación de género del individuo. Finalmente, en el Capítulo 8 se determina el conjunto de landmarks corporales 3D para estimación de partes del cuerpo. En cada caso, además, se realiza un análisis de los vectores de características, su informatividad y robustez. Se propone también presentar un panorama de la diversidad de aplicaciones que pueden derivarse del uso de *deep learning* para predicción de ubicación de coordenadas de landmarks en una gran variedad de campos de aplicación.

5.2. Conjunto de datos utilizados

Para cada fenotipo se contó con diferentes subconjuntos del conjunto de datos, por lo que se dará una introducción de cómo fueron tomados en general, y luego cada fenotipo

explicará su N y variaciones en la configuración de landmarks.

5.2.1. Conjunto de datos CANDELA

El consorcio CANDELA (Consortium for the Analysis of the Diversity and Evolution of Latin Americans), es un Consorcio Internacional multidisciplinario que incluye genetistas, antropólogos, estadistas, bioinformáticos y antropólogos sociales interesados en estudiar la biodiversidad y el entorno sociocultural del poblamiento Latinoamericano Ruiz-Linares et al. (2014))[1]. CANDELA cuenta con una base de datos de 7500 individuos fotografiados bajo un protocolo para tomas de fotos estandarizadas.

Las imágenes consisten en 5 vistas del rostro (lado izquierdo 0°, 45°, frontal 90°, angulo derecho 135° y lado derecho 180°), como se ven en la Figura 5.1. Durante la toma, el objetivo principal era la obtención manual de landmarks, por lo que no se tuvo el cuidado de utilizar un fondo uniforme. En su mayoría la resolución de las fotografías es 2136×3216 aunque existían otros tamaños. Las fotografías fueron tomadas de forma manual a una distancia de $\approx 1{,}5$ metros, al nivel de los ojos, con una cámara Nikon D90 con lentes 50mm AF Nikkor a una apertura de f/11. No se realizó remoción del fondo, lo cual dificulta mucho el uso de algoritmos de procesamiento de imágenes.

Los landmarks y semi-landmarks utilizados para el entrenamiento y evaluación fueron digitalizados y procesados manualmente utilizando TPSDig y TPSUtil[2]. Tanto las fotografías como los landmarks son propiedad del Consorcio.

5.2.2. Conjunto de datos de dominio público

Para las pruebas sobre la evaluación de calidad de landmarking y su posible uso en aplicaciones de clasificación (Ver Sección 6.4 y 7.5) se trabajaron con conjuntos de datos de dominio público. Los cuales son listados a continuación.

CVL Face Database [3] Esta base de datos cuenta con fotografías de 114 personas, una resolución de 640×480 píxeles en formato jpeg, tomadas con una cámara Sony

[1] https://www.ucl.ac.uk/candela

[2] http://life.bio.sunysb.edu/morph/

[3] http://www.lrv.fri.uni-lj.si/facedb.html

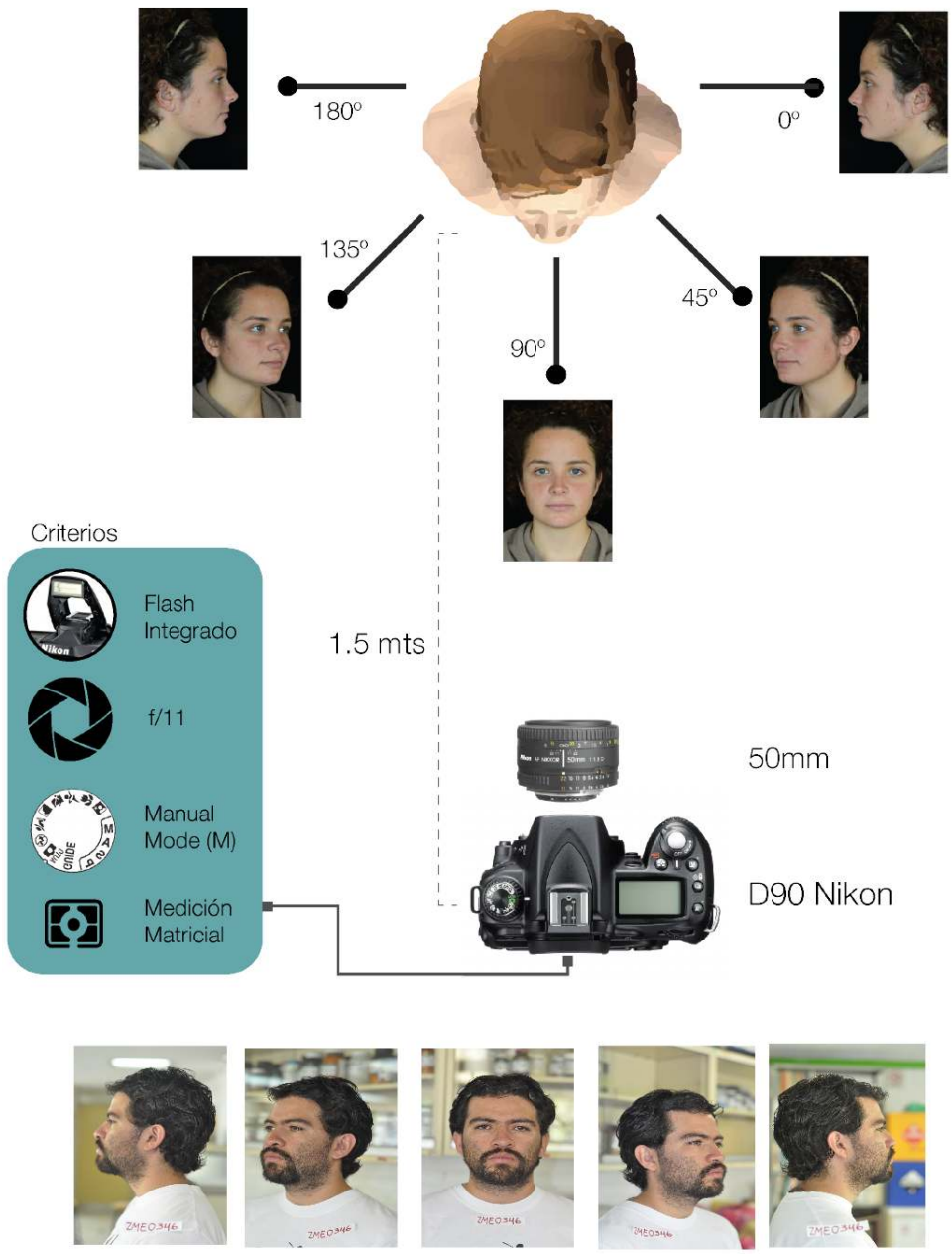


Figura 5.1: Ejemplo de serie de imágenes tomadas bajo el protocolo CANDELA Quinto Sanchez (2015).

Digital Mavica bajo iluminación y fondo uniforme, sin flash. Una muestra de las imágenes se puede ver en la Figura 5.2.

AMI Ear Database [4] Esta base de datos cuenta con fotografías de 100 individuos diferentes. Para cada uno de ellos se cuenta con 7 tomas, con una resolución de 492×702

[4] http://www.ctim.es/research_works/ami_ear_database/

Figura 5.2: Ejemplo de serie de imágenes de CVL Face Database.

Figura 5.3: Ejemplo de serie de imágenes de AMI Ear Database.

píxeles, formato jpeg, tomadas con una cámara Nikon D100. Una muestra de las imágenes se puede ver en la Figura 5.3.

IIT Delhi Ear Database [5] Esta base de datos cuenta con fotografías de 121 individuos, cada uno de ellos con al menos 3 tomas, con una resolución de 272 × 204 píxeles en formato jpeg. Una muestra de las imágenes se puede ver en la Figura 5.4.

En estos conjuntos de datos el fondo es plano pero no exactamente uniforme. Al igual que con el conjunto de datos CANDELA, el fondo no uniforme representa un problema potencial para el procesamiento de imágenes.

[5] http://www4.comp.polyu.edu.hk/~csajaykr/IITD/Database_Ear.htm

Figura 5.4: Ejemplo de serie de imágenes de IIT Delhi Ear Database.

5.3. Métricas

Para evaluar la calidad de las predicciones de nuestros modelos es necesario determinar métricas que nos permitan valorar el error. Dado que se tomó el landmarking automático como un problema de regresión, se detallan a continuación las métricas utilizadas para este tipo de problemas. En los Capítulos 6 y 7 se mostrarán los resultados calculados en base a estas definiciones. Además, como se proponen diferentes aplicaciones de clasificación para nuestros vectores de características basados en landmarks, se detallan métricas de clasificación utilizadas en los experimentos de las Secciones 6.5 y 7.5.

Evaluaciones de la calidad del Landmarking

Variancia explicada (EV por *explained variance*) mide la proporción en la que un modelo tiene en cuenta la variación o dispersión de un determinado conjunto de datos. Sea $\hat{y}$ el valor de salida estimado, y el valor real y Var la variancia (raíz del desvío estándar) de una muestra. La variancia explicada se define como:

$$EV(y, \hat{y}) = 1 - \frac{Var\{y - \hat{y}\}}{Var\{y\}} \tag{5.1}$$

Error cuadrático medio Mean Squared Error (MSE) mide el promedio de los cuadrados de las desviaciones o errores, entendidos como la diferencia entre lo predicho y el valor real. El MSE otorga una métrica de calidad del estimador. Siempre supone un valor no negativos, cuanto más cercano a 0 mejor es el estimador. Si $\hat{y}_i$ es el valor predicho de la muestra $i - esima$ e y es su valor real, el MSE estimado sobre n muestras puede ser definido como:

$$MSE(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2 \tag{5.2}$$

Coeficiente de determinación (R^2) es un número que indica la proporción de variancia en la variable dependiente que el modelo explica o predice a partir de las variables independientes. Este valor nos provee un indicador de la capacidad predictiva del modelo de aprendizaje frente a futuras observaciones. Utilizamos el coeficiente de determinación de Pearson dado que no se realizan estadísticas de rango.

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=0}^{n_{\text{samples}}-1} (y_i - \hat{y}_i)^2}{\sum_{i=0}^{n_{\text{samples}}-1} (y_i - \bar{y})^2} \tag{5.3}$$

Donde $\bar{y} = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} y_i$

Evaluación de clasificaciones basadas en landmarking automático

Además de las métricas para evaluar la performance del landmarking automático, se utilizaron varias métricas de clasificación para los ejemplos de aplicación dados en las Secciones 6.5 y 7.5. Todas las evaluaciones fueron realizadas utilizando Validación cruzada con el uso de *KFolds* estratificados.

Precisión La precisión es la relación $tp/(tp + fp)$, donde tp es el número de verdaderos positivos y fp es el número de falsos positivos. De forma intuitiva, se puede definir como la proporción de casos en los que el modelo acertó al predecir la condición positiva.

Exactitud Es la proporción de casos correctamente predichos (verdaderos positivos y negativos sobre la cantidad total de casos). En un problema de regresión, si $\hat{y}_i$ es el valor predicho del i-th elemento de la muestra e y_i es su valor real, la cantidad de

predicciones correctas sobre el total de la muestra $n_{samples}$ se define como:

$$\texttt{accuracy}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} 1(\hat{y}_i = y_i) \tag{5.4}$$

Recall Es la relación $\frac{tp}{(tp+fn)}$, donde tp es la cantidad de verdaderos positivos y fn el número de falsos negativos. De forma intuitiva podemos decir que el recall o recupero es la capacidad del modelo para predecir casos positivos, lo cual tiene sentido en particular cuando los verdaderos negativos son muy abundantes.

f1-score Es la media geométrica entre precisión y recall.

$$F1 = \frac{2 * (precision * recall)}{(precision + recall)} \tag{5.5}$$

Cuando se trata de clasificaciones multi-clase, este es el promedio de los F1 de cada clase. Se puede generalizar dándole mayor importancia a la precisión o a la inversa, dependiendo del caso.

Adjusted Rand Index Adjusted Rand Index (ARI) Este índice proporciona una medida de similitud entre dos conjuntos de datos. A diferencia del *Rand Index* que devuelve valores entre $[0, 1]$, su versión ajustada proporciona valores negativos si el valor es aún menor al esperado. Si C es el conjunto que cuenta con la asignación de clases correctas (*ground truth*) y K los valores agrupados automáticamente. Definimos a y b como:

a el número de pares de elementos que están en el mismo conjunto en C y en el mismo conjunto en K.

b el número de pares de elementos que están en conjuntos diferentes en C y en conjuntos diferentes en K.

El Rand Index (RI) (Rand Index) se define como:

$$RI = \frac{a + b}{C_2^{n_{samples}}} \tag{5.6}$$

Donde $C_2^{n_{samples}}$ es el número total de posibles pares en todo el conjunto de datos.

$$ARI = \frac{(RI - Expected_{RI})}{(max(RI) - Expected_{RI})} \tag{5.7}$$

Curva ROC Cuando el clasificador se basa en una función umbral, la variación de este último permite obtener diferentes tasas de verdaderos positivos y verdaderos negativos. La curva ROC muestra estos dos valores en función del valor implícito del umbral.

Capítulo 6

Landmarking del Pabellón Auditivo

En este Capítulo se presenta el problema general del landmarking del pabellón auditivo, se detalla la configuración de landmarks utilizada junto con su definición bioantropológica bajo la estructura detallada en el Capítulo 3. Luego continuaremos con la descripción del *pipeline* utilizado para el landmarking automático, incluyendo su diseño, implementación y pruebas realizadas para medir su desempeño. Además analizaremos la importancia de ciertos landmarks y su impacto sobre el factor discriminante de posibles vectores de características basados en landmarks. Finalmente se presenta una aplicación directa del landmarking automático del pabellón auditivo, enfocada en la identificación de individuos a partir de landmarking, describiendo el modelo y los resultados obtenidos.

6.1. Configuración de landmarks: Pabellón Auditivo

El pabellón auditivo humano se compone de una pieza de cartílago cubierto con la piel y unido al cráneo por ligamentos, músculos y tejido fibroso. Este cartílago no se extiende al lóbulo de la oreja, el cual en cambio consiste principalmente de tejido areolar y adiposo. Existe una gran variación no patológica entre humanos en la forma y tamaño del pabellón auditivo, influenciada por edad, género y etnicidad (Alexander et al., 2011, Adhikari et al., 2015, Azaria et al., 2003, Sforza et al., 2009). La variación en el pabellón auditivo fue analizada utilizando 7 landmarks y 38 semi-landmarks. La configuración específica utilizada se puede ver en la Figura 6.1 y su definición anatómica en la Tabla 6.1. Además de la configuración de landmarks, se pueden derivar otro tipo de medidas, por ejemplo, distancias

Tabla 6.1: Configuración y definición anatómica de landmarks y semi-landmarks en el pabellón auditivo humano.

Número	Nombre
1	Otobasion superiorious
2	Concha superiorious
3	Tragus superiorious
4	Intertragic incisure
5	Otobasion inferiorous
6	Helix basal border
7	Crus Helix
8 a 45	Semi-landmarks

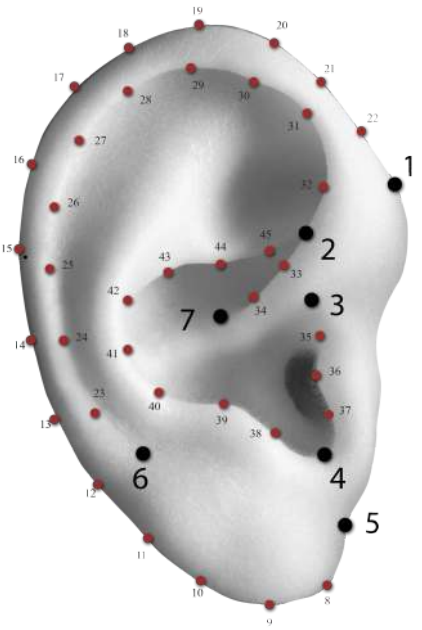


Figura 6.1: Configuración de Landmarks y semi-landmarks junto con su descripción anatómica Purkait and Singh (2008), Ercan et al. (2008).

Euclideanas entre landmarks y semi-landmarks (Purkait and Singh, 2008), o ángulos de interés anatómico. Éstas son algunas medidas que pueden tomarse para su utilización en un vector robusto de características para otras aplicaciones, como por ejemplo la biometría.

6.2. Conjunto de datos: Pabellón Auditivo

Para el procesamiento de las imágenes de orejas se trabajó con 2735 pares de datos de forma $(image, landmarks)$. Las imágenes sin coordenadas de landmarks manuales asociados fueron excluidas. Cada tupla estaba compuesta por una imagen de 2136×3216 píxeles y por una lista de 45 landmarks y semi-landmarks, especificados en el Capítulo 3. Del conjunto de datos utilizado, se designaron 2051 imágenes para el entrenamiento (75 %), y 684 imágenes (25 % de todo el conjunto de datos disponible) para la etapa de validación, seleccionados bajo *cross-validation*. En algunas imágenes se cuenta con orejas parcialmente cubiertas por cabello, aretes o en que el fondo no es uniforme, estas imágenes fueron utilizadas para probar la robustez de nuestros métodos sobre esta clase de ruido semántico.

6.3. Pipeline Desarrollado

La idea principal consiste en diseñar y entrenar una CNN con un conjunto de ejemplos de orejas asociadas con las coordenadas en dos dimensiones (x, y) de sus landmarks provistas por expertos humanos, y una ROI remuestreada y normalizada donde se encuentre el pabellón auditivo. Se entrenó una CNN con un conjunto de 2735 imágenes asociadas con 45 landmarks y semilandmarks ubicados manualmente. Se utilizaron técnicas específicas de entrenamiento para lograr tasas de generalización altas para evitar *overfitting*. En esta sección detallamos todos los pasos del proceso, y por cada paso, describimos los formalismos utilizados.

6.3.1. Pre-procesamiento de las imágenes

Para reducir la carga de procesamiento innecesario en la ConvNet, se extrae una región de interés (ROI) alrededor de la oreja. Para la ubicación de esta región se utilizó el *framework* de detección general de objetos de Viola-Jones (Viola and Jones, 2001), para lo que fue entrenado un filtro *Haar cascade* con 133 imágenes positivas (regiones donde

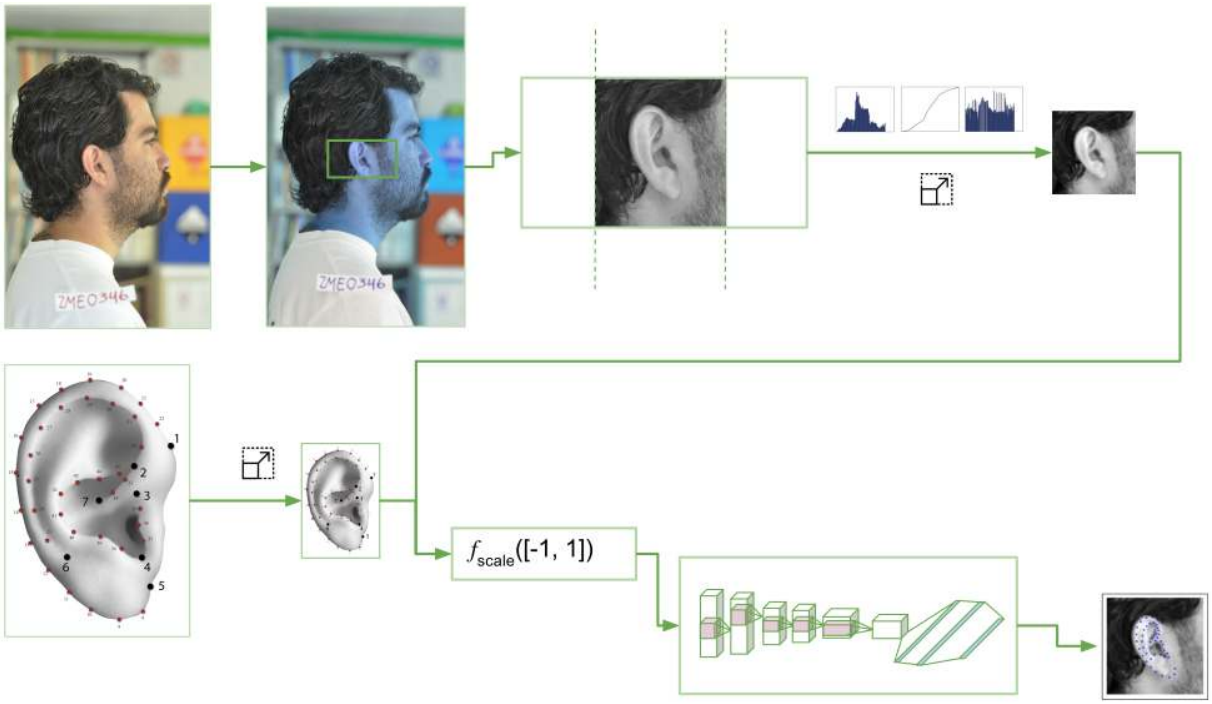


Figura 6.2: Visión general del *Pipeline* desarrollado para landmarking automático sobre la estructura del pabellón auditivo

aparece la oreja) y 667 negativas[1]. Para validar este paso de pre-procesamiento, se tomaron de forma aleatoria 185 imágenes del conjunto de datos de CANDELA. En el 92,43 % de los casos la ROI fue encontrada correctamente, en el 1,62 % la ROI no fue encontrada y en el 5,95 % de los casos fue ubicada incorrectamente. En la Sección 6.6 se realizará un análisis comparativo entre el uso del framework de Viola-Jones y el uso de CNNs para esta etapa.

Luego de que la ROI fue encontrada, se recorta en forma cuadrada teniendo en cuenta los tamaños de alto y ancho de cada ROI, se aplica una ecualización del histograma por estiramiento (*histogram stretch*), mediante el cual los valores de luminancia en la ROI se alteran para cubrir la mayor parte del rango dinámico disponible. Los parámetros de estiramiento del histograma fueron programados para convertir en negro el 2 % de los píxeles y a blanco el 1 % de los píxeles como máximo en ambos casos. Como último paso, la ROI se remuestrea al tamaño final utilizado por la CNN, el cual es de 96×96 pixels, mediante el uso de submuestreo bilineal. En la Figura 6.2 se pueden observar el efecto de

[1]El Haarcascade entrenado puede ser descargado y utilizado desde https://github.com/celiacintas/tests_landmarks/blob/master/files/cascade_lateral_ears.xml

estos pasos.

6.3.2. Pre-procesamiento de landmarks

Los landmarks supervisados fueron generados manualmente por expertos utilizando el software TPSdig. Como se mencionó en la Sección 5.2, este programa ubica el origen en el extremo superior izquierdo, por lo que por comodidad a la hora de implementación, antes de realizar cualquier modificación se invirtieron en el eje y todas las coordenadas para contar con el origen en la parte inferior izquierda. Las coordenadas se toman de un archivo con el siguiente formato:

```
id x00 y00 x01 y01 ... xnn ynn
```

Estas coordenadas se encuentran en el espacio de la imagen completa original de 2136×3216 píxeles. Estas coordenadas se tuvieron que trasladar dentro del espacio de la ROI y ser escaladas por el mismo factor utilizado en la etapa de pre-procesamiento de la imagen visto en la sección 6.3.1. En términos generales se aplicaron las siguientes funciones:

$$(x_{new}, y_{new}) = ((x_{old} - x_{sup\_izq})/dx, (y_{old} - y_{inf\_der})/dy)), \tag{6.1}$$

donde x_{old} e y_{old} son las coordenadas originales respecto a la imagen completa, x_{sup_izq} es la posición x de la esquina superior izquierda del rectángulo correspondiente a la ROI e y_inf_der es la coordenada y de la esquina inferior derecha del mismo. Los factores de escala se mantuvieron como variables diferentes, dado que si se utiliza la imagen completa al no ser cuadrada, los factores serán distintos.

6.3.3. Elección de Arquitectura

Fueron implementadas y puestas a prueba varias arquitecturas de CNN. En esta Sección mostraremos las tres de ellas que obtuvieron mejores resultados. Estas arquitecturas fueron diseñadas y entrenadas con el propósito de realizar landmarking automático, principalmente para detectar e identificar partes anatómicas del pabellón auditivo sobre imágenes. Como datos de entrada se tomaron imágenes de orejas, desde la vista lateral en escala de grises (un único canal) de 96×96 píxeles, estos valores fueron escalados a $[0, 1]$.

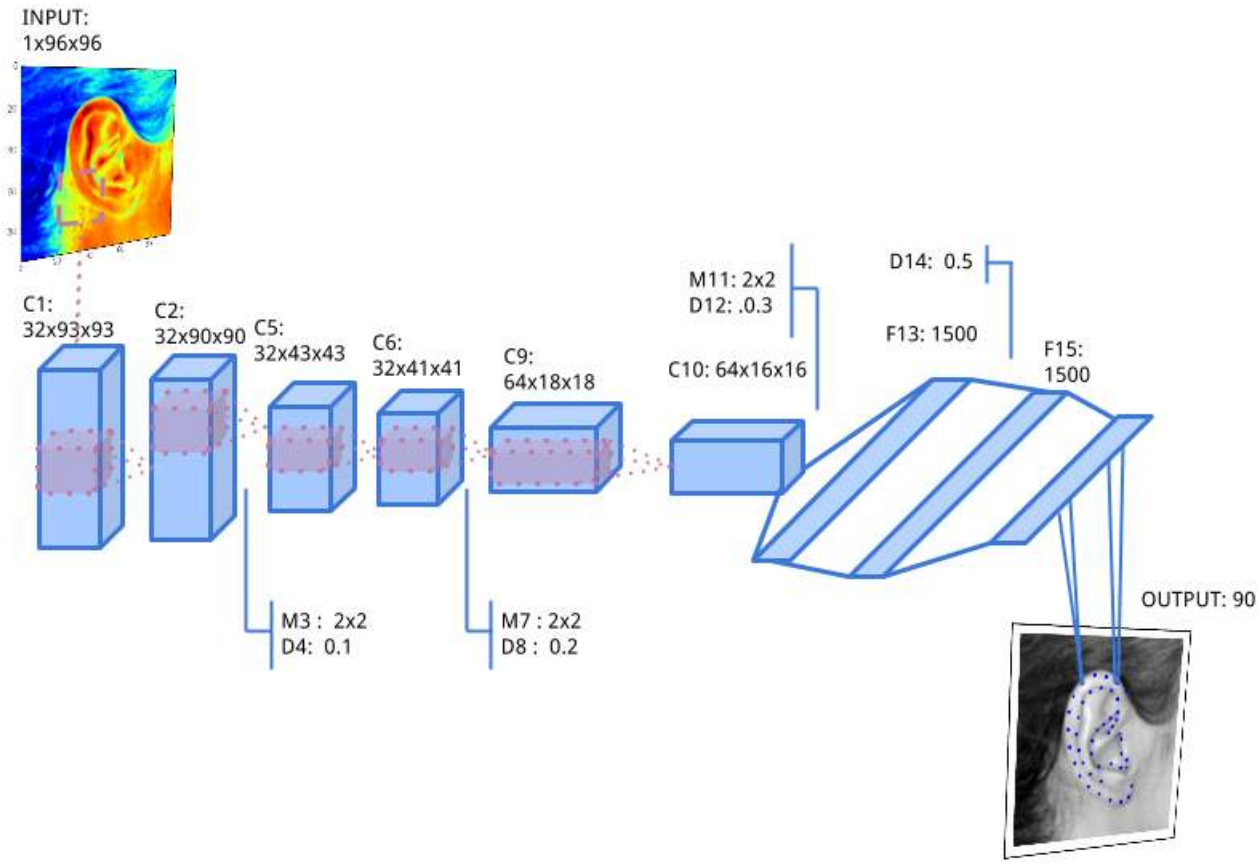


Figura 6.3: Arquitectura general de la CNN con mejor performance.

En la Figura 6.3 se puede observar la arquitectura de la red con mejor desempeño (denominada *Arch0*).

La estructura básica de la red consiste en dos capas de convolución con filtros cuadrados, seguidos de capas *max-pooling* y *dropout*. Esta estructura se repite tres veces para obtener características en distintos niveles de abstracción, con diferentes tamaños de filtros, cantidad de mapas de características y valores de probabilidad para las capas de *dropout*. Siguiendo la Figura 6.3, las capas de convolución C1, C2, C5 Y C6 tienen 32 filtros de 4×4 y 3×3. Se puede ver un ejemplo de los *kernels* generados por C1 y sus respectivos mapas de características ante una determinada imagen en la Figura 6.4. Las capas C9 y C10 cuentan con 64 filtros de 3×3. Todas las capas de *max-pooling* cuentan con un tamaño de 2×2, y los valores de probabilidad utilizados para D4, D8, D12 y D14 son 0,1, 0,2, 0,3 y 0,5 respectivamente. Luego de que se finaliza con la extracción de características, la arquitectura contiene dos capas lineales completamente conectadas, con 1500 unidades cada una (F13 y F15 en la figura), y una capa *dropout* en el medio

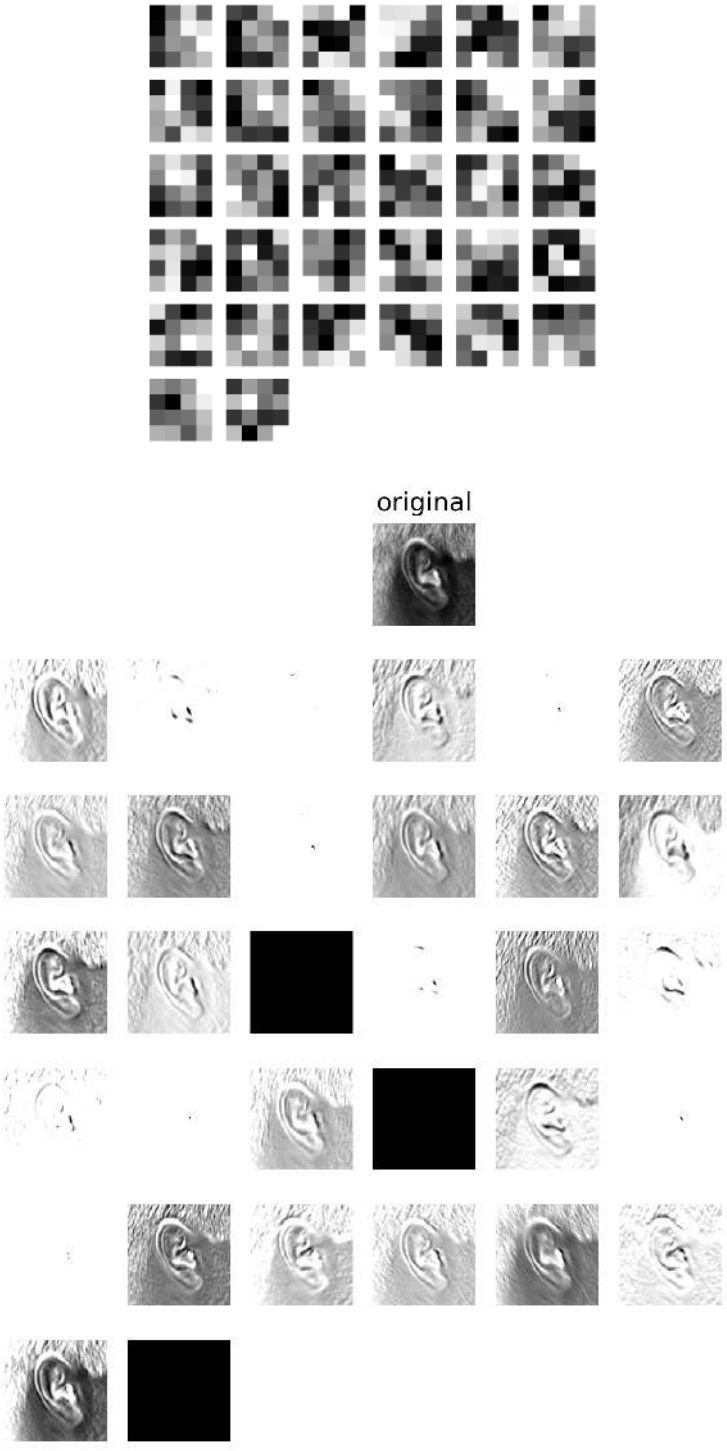


Figura 6.4: *Kernels* y mapas de características de la capa C1 sobre una imagen de entrada X.

(D14). La capa de salida cuenta con 90 unidades (45 pares $[x, y]$) uno para cada posición predicha de landmarks y semi-landmarks.

La implementación fue realizada en Python (Oliphant (2007), van der Walt et al.

(2011)) mediante la biblioteca Lasagne Dieleman et al. (2015b)[2]. A continuación podemos ver una estructura de CNN programada en Python con Lasagne Dieleman et al. (2015b) como ejemplo. Más detalles sobre la implementación de CNN en Python puede observarse en el Anexo A.

```
layers_0 = [
        (InputLayer, {'shape': (None, 1, 96, 96)}),
        (Conv2DLayer, {'num_filters': 32, 'filter_size': (4, 4)}),
        (Conv2DLayer, {'num_filters': 32, 'filter_size': (4, 4)}),
        (MaxPool2DLayer, {'pool_size': 2}),
        (DropoutLayer, {'p': 0.1}),
        (Conv2DLayer, {'num_filters': 32, 'filter_size': (3, 3)}),
        (Conv2DLayer, {'num_filters': 32, 'filter_size': (3, 3)}),
        (MaxPool2DLayer, {'pool_size': 2}),
        (DropoutLayer, {'p': 0.2}),
        (Conv2DLayer, {'num_filters': 64, 'filter_size': (3, 3)}),
        (Conv2DLayer, {'num_filters': 64, 'filter_size': (3, 3)}),
        (MaxPool2DLayer, {'pool_size': 2}),
        (DropoutLayer, {'p': 0.3}),
        (DenseLayer, {'num_units': 1500}),
        (DropoutLayer, {}),
        (DenseLayer, {'num_units': 1500}),
        (DenseLayer, {'num_units': 90, 'nonlinearity': None}),
        ]


def create_network(npochs=1000, batch_s=178, layers_0):
    return NeuralNet(
        layers=layers_0,
        update=nesterov_momentum,
        update_learning_rate=theano.shared(np.float32(0.08)),
```

[2]El código se encuentra disponible en https://github.com/celiacintas/tests_landmarks/blob/master/testing_output_ears.ipynb.

```
        update_momentum=theano.shared(np.float32(0.9)),

        regression=True,
        batch_iterator_train=FlipBatchIterator(batch_size=batch_s),
        on_epoch_finished=[
            AdjustVariable('update_learning_rate', start=0.08, stop=0.001),
            AdjustVariable('update_momentum', start=0.9, stop=0.9999)
        ],
        max_epochs=npochs,
        verbose=1)
net0 = create_network()
```

Esto nos permite el uso de aceleración mediante GPU de una forma sencilla. El entrenamiento de esta red tomó 25 horas aproximadamente sobre una placa NVIDIA GeForce GTX 590. Una vez entrenada, la red puede ser utilizada en hardware convencional, incluso en sistemas embebidos.

Problemas de overfitting

Normalmente las redes neuronales convolucionales cuentan con un número muy grande de parámetros a aprender 8.662.970 en el pabellón auditivo. Debido al limitado tamaño de nuestro set de entrenamiento, contamos con grandes probabilidades de experimentar *overfitting* (Definido en la Sección 4.3.3). La red tenderá a memorizar los ejemplos dados en la etapa de entrenamiento, ya que tiene memoria suficiente para hacerlo. Esto, obviamente, no generalizará de forma deseada ante nuevos datos. Se tomaron dos medidas para reducir el efecto de *overfitting* durante la etapa de entrenamiento.

- Incremento de datos: De forma artificial aumentamos el N de muestras, mediante el uso de transformaciones que preservan el valor de y (siendo y la clase objetivo). De forma aleatoria, algunas imágenes y sus landmarks asociados son espejados sobre sus coordenadas x y agregados al set de entrenamiento. Un ejemplo de esto se puede ver en Figura 6.5.

- Regularización: La complejidad del modelo fue penalizada utilizando *dropout* Hin-

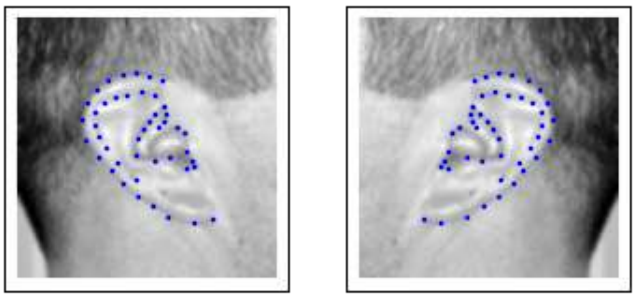


Figura 6.5: Landmarks asociados a una fotografía espejados sobre el eje x.

Tabla 6.2: Desempeño de las tres arquitecturas CNNs.

	Arch0	*Arch1*	*Arch2*
r^2	**0.709**	0.678	0.698
RMSE	**2.296**	2.415	2.338
EV	**0.976**	0.974	0.975
Pearson	**0.988**	0.987	0.988

ton et al. (2012), la cual fue explicada en detalle en la Sección 4.3.4, pero que básicamente consiste en anular salidas de unidades de capas ocultas con una cierta probabilidad. Esta técnica reduce adaptaciones complejas entre unidades de capas, ya que en etapa de entrenamiento, una unidad no puede depender de la activación de otra unidad en particular Krizhevsky et al. (2012b).

6.4. Resultados sobre landmarking automático en Pabellón Auditivo

El problema de ubicación de landmarks en forma automática puede ser pensado como un problema de regresión. Al usar este enfoque aplicamos métricas para evaluar el desempeño de diferentes CNNs contra landmarking manual (ground truth). En particular se trabajó con r^2, root mean square error (RMSE), variancia explicada (Explained Variance (EV)) y correlación de Pearson, detalladas en la Sección 5.3.

El desempeño del landmarking de las tres arquitecturas implementadas se puede observar en la Tabla 6.2. Se utilizó como línea de base un regresor básico que tiene como política

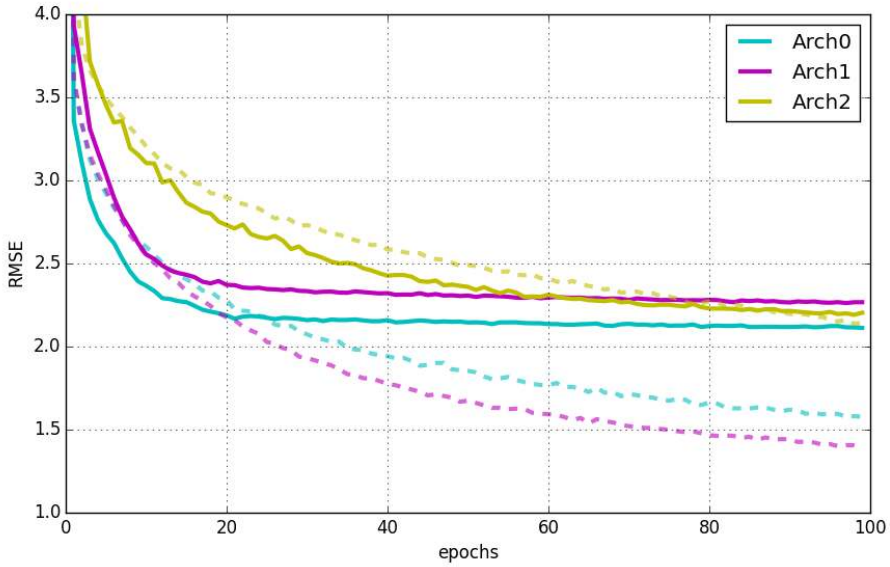


Figura 6.6: Curvas de aprendizaje de CNNs analizadas en la Tabla 6.2. Las líneas punteadas representan el RMSE sobre los valores de entrenamiento, y las líneas sólidas representan el error sobre el conjunto de validación de las tres redes.

Tabla 6.3: RMSE de cada landmark anatómico.

# Landmark	RMSE
1	1.8183
2	1.2216
3	1.08651
4	1.3291
5	2.4477
6	2.59746
7	1.17571

predecir el promedio de los valores de entrenamiento, el cual cuenta con un $r^2 = 0{,}003$.En la Tabla 6.3 se detalla los RMSE para cada coordenada de landmark.

Las métricas de regresión fueron calculadas utilizando las técnicas implementadas en Pedregosa et al. (2011). También se pueden observar en la Figura 6.6 las curvas de aprendizaje sobre los datos de entrenamiento y validación de las tres arquitecturas. La Figura 6.7 muestra los landmarks predichos sobre imágenes no vistas por la red en el

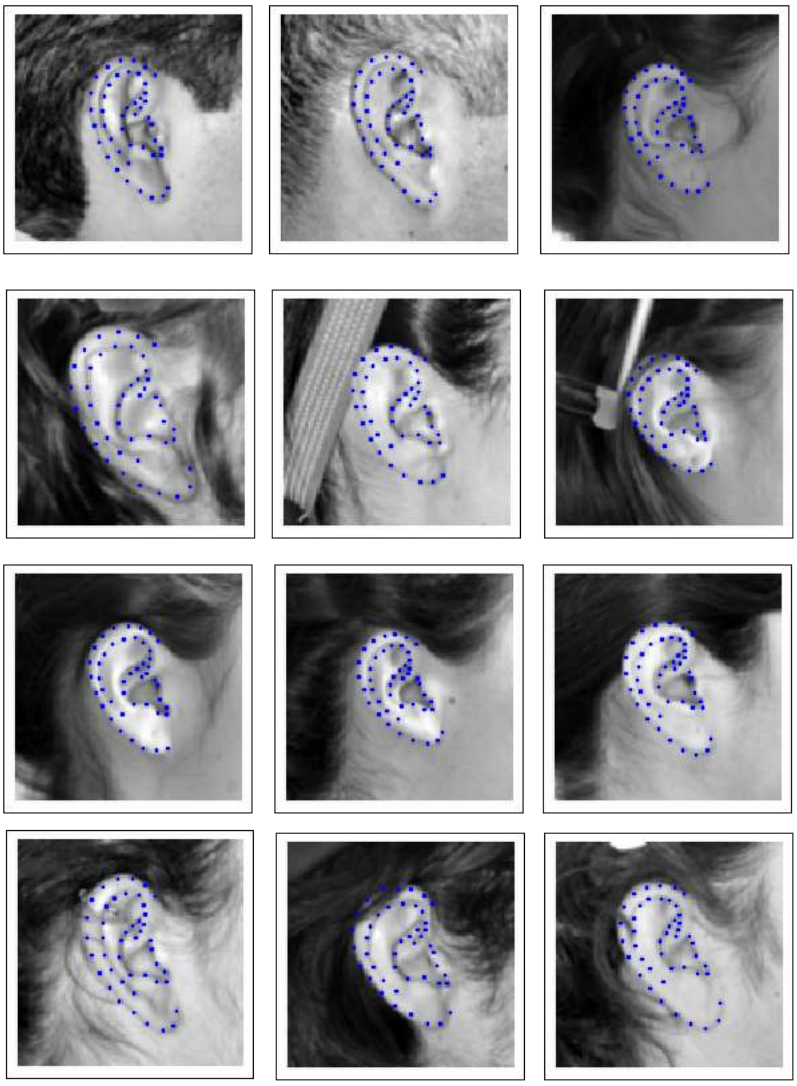


Figura 6.7: Resultados de nuestra mejor red sobre imágenes no vistas en la etapa de entrenamiento Cintas et al. (2016a).

conjunto de validación. Se puede observar que alguna de estas imágenes cuentan con obstrucción parcial de cabello, pese a lo cual el landmarking es correcto.

El conjunto completo de imágenes de validación landmarkeado por la CNN puede verse en (https://github.com/celiacintas/tests_landmarks/blob/master/testing_output_ears.ipynb). También se entrenaron redes para trabajar sobre la imagen completa, eliminando la búsqueda de ROI. En promedio, los resultados obtenidos fueron $r^2 = 0.884$, *RMSE* = 1.365, *EV* = 0.951 y *Correlación de Pearson* = 0.976, para landmarking sobre imágenes completas.

Algunos resultados sobre conjuntos de datos externos a CANDELA se pueden ver en la

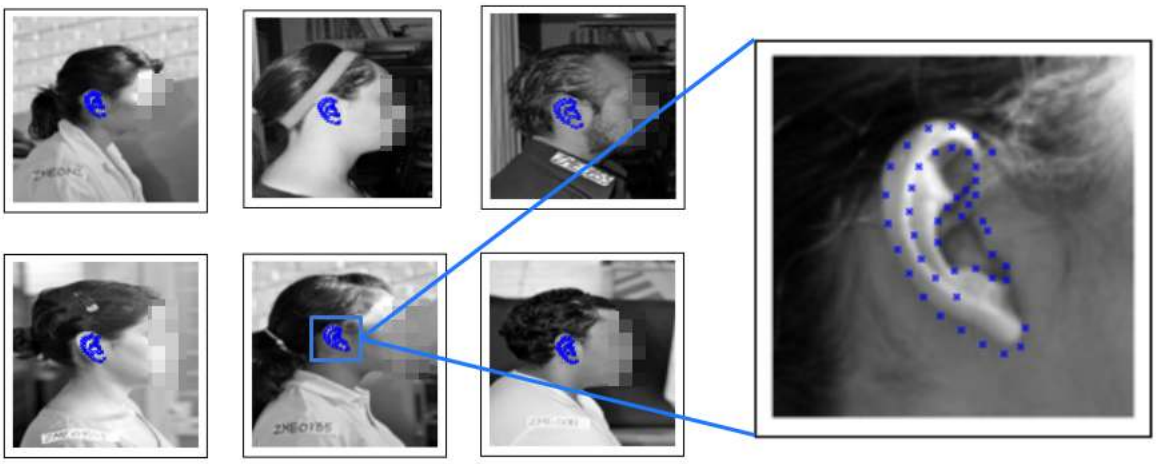


Figura 6.8: Resultados sobre imágenes seleccionadas de forma aleatoria en el conjunto de datos de CANDELA con fondo e iluminación no controlada Cintas et al. (2016a).

Figura 6.9. Si bien no se contaba con los landmarks para realizar métricas comparativas, se puede observar potencial sobre imágenes de otras fuentes externas a las de entrenamiento y validación. El landmarking fue evaluado además bajo imágenes menos controladas. Para esto se tomó de forma aleatoria un subconjunto de las imágenes de CANDELA de cara completa, sin iluminación controlada, fondo heterogéneo, etc. Los resultados se pueden ver en la Figura 6.8.

La red fue implementada en una PC de hardware convencional (single core Intel i7-5500 2.40GHz). El landmarking automático de una imagen requiere 4,68ms. Por otro lado el landmarking en modo *batch* de 684 imágenes demoró 1,04 segundos. En estas pruebas se decidió trabajar con la ROI de la oreja para disminuir la pérdida de información, pero es posible entrenar diferentes redes para trabajar con la imagen del rostro completo, en la cual las orejas se pueden ubicar sin ningún problema. En (https://github.com/celiacintas/tests_landmarks/blob/master/testing_output_ears.ipynb) se encuentran las pruebas y arquitectura de la red implementada sobre ipython notebooks, y la red entrenada se puede descargar y usar de (https://mega.nz/#F!XA8i1JbA!HaZQjjToUYnrF8xMNhWDJw).

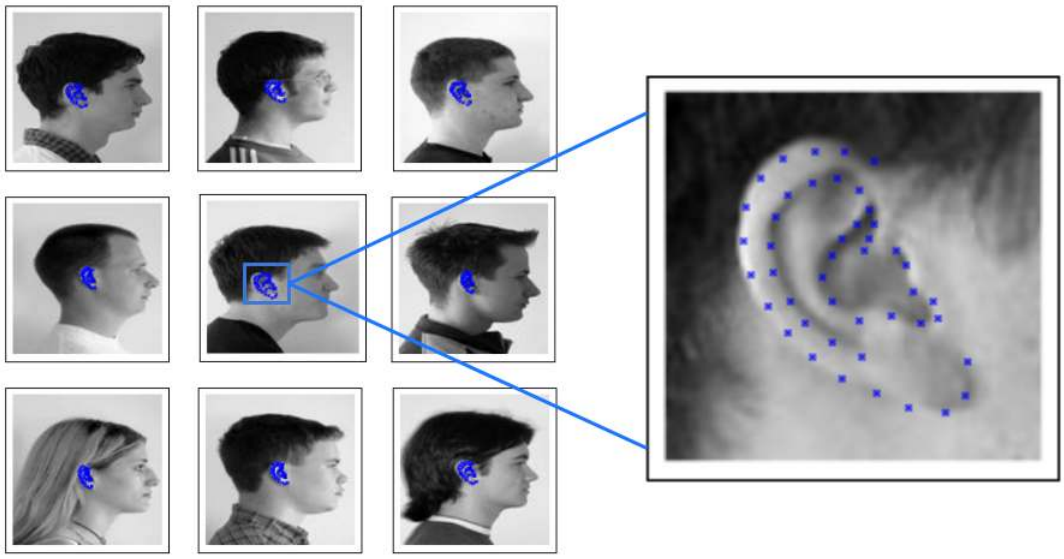


Figura 6.9: Resultados sobre imágenes seleccionadas de forma aleatoria de *CVL Face Database* http://www.lrv.fri.uni-lj.si/facedb.html (Solina et al., 2003) presentados en Cintas et al. (2016a)

.

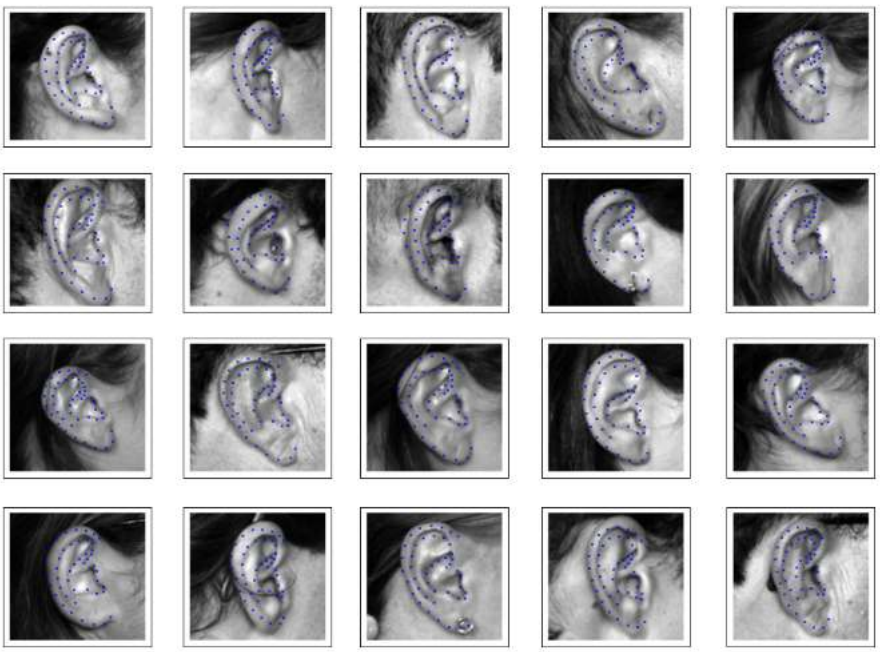


Figura 6.10: Resultados de la mejor arquitectura sobre imágenes pertenecientes a la base de datos AMI (http://www.ctim.es/research_works/ami_ear_database/).

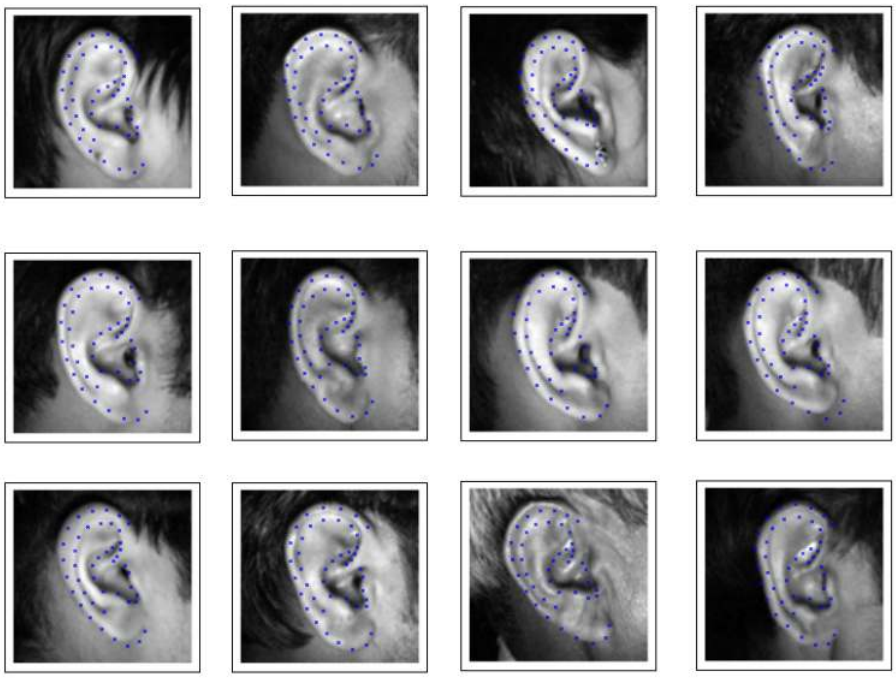


Figura 6.11: Resultados de la mejor arquitectura sobre imágenes pertenecientes a la base de datos IIT Delhi (Kumar and Wu, 2012).

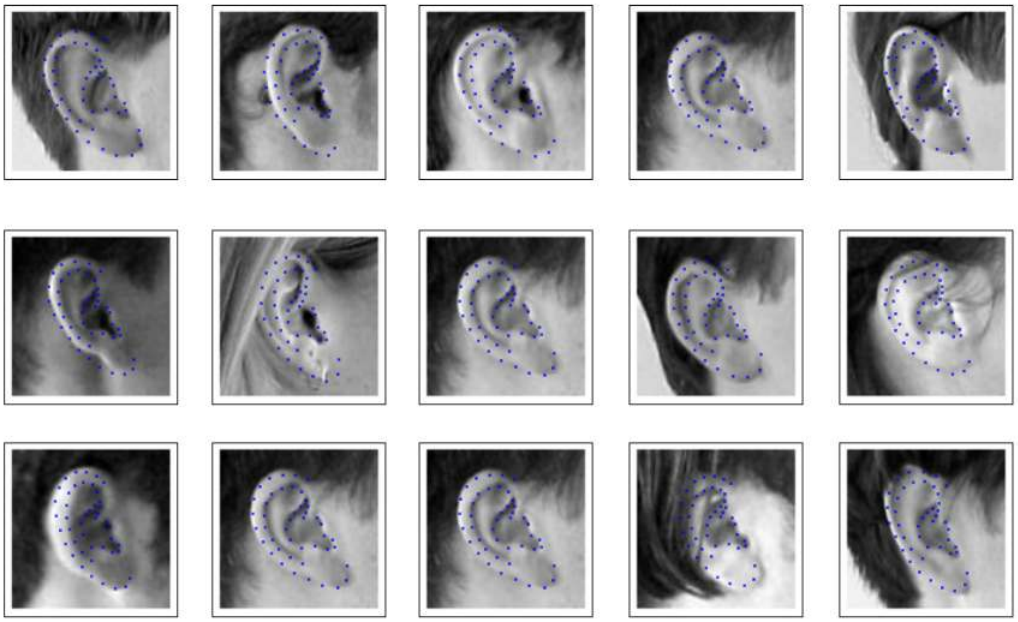


Figura 6.12: Resultados de la mejor arquitectura sobre imágenes pertenecientes a la base de datos CVL Solina et al. (2003).

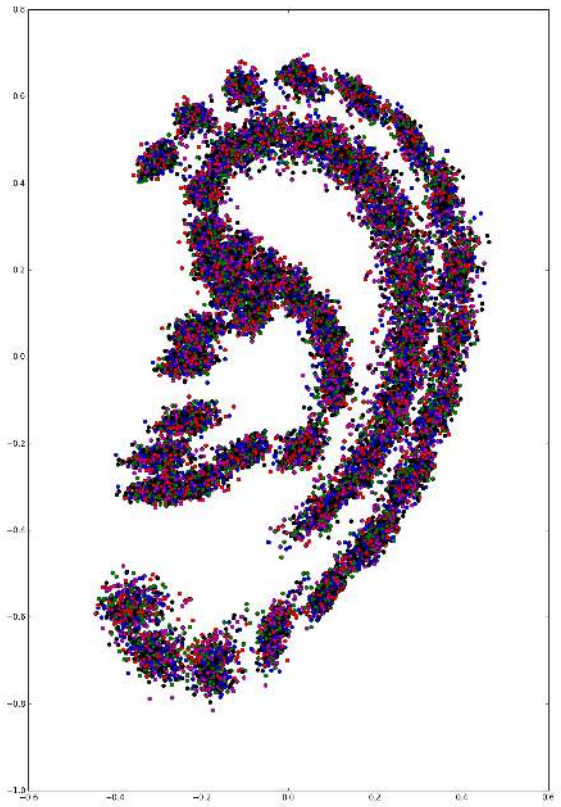


Figura 6.13: Landmarking de 15540 orejas de la muestra CANDELA, luego de la transformación de Procrustes Cintas et al. (2016a).

6.5. Biometría y aplicaciones forenses

Aunque el propósito principal de este trabajo era mostrar el potencial combinado de morfometría geométrica junto con *Deep Learning*, realizamos algunos experimentos previos, para evaluar si el *workflow* presentado podría ser utilizado con fines de reconocimiento. Para esto, agregamos un *Extremely Randomized Tree (ERT)*, como una última etapa de clasificación. Los ERTs son árboles de clasificación (o regresión), en los cuales los atributos y opciones de corte son parcial o totalmente aleatorios, a la hora de ramificarse en un nuevo nodo durante entrenamiento (Geurts et al., 2006). En un caso extremo, un ERT, construye árboles de clasificación aleatorios, cuyas estructuras son independientes de los valores de salida de elemento de entrenamiento.

6.5.1. Trabajos previos

La identificación automática de estructuras anatómicas de interés biométrico, tales como huellas digitales o patrones del iris, son utilizadas ampliamente en sistemas de control de acceso, investigación antropológica y hasta vigilancia. Estas estructuras tienen como desventaja que para la toma de las mismas se requiere de métodos intrusivos para su obtención.

Sobre este tema, tanto la vista lateral (o frontal) del rostro como la estructura del pabellón auditivo tienen varias ventajas, no sólo en que la información puede ser capturada a distancia con métodos no intrusivos (una fotografía) sino también porque el pabellón auditivo tienen menor variancia con la edad, y no es afectado por expresiones faciales. La evidencia empírica relacionada con el caracter único e individual del pabellón auditivo fue presentado por Iannarelli (1989). Por otro lado, en la mayoría de las propuestas en el estado del arte la captura de los atributos fenotípicos y forma de la oreja se basan en técnicas *ad hoc* que limitan su aplicación.

La mayoría de los algoritmos de detección de pabellón auditivo dependen de propiedades geométricas de la parte externa de la oreja, como la ocurrencia de ciertos bordes, orientaciones de curvas o frecuencia de patrones Pflug and Busch (2012). Un trabajo detallado de todos los métodos hasta el momento sobre reconocimiento e identificación de pabellón auditivo puede ser encontrado en Pflug and Busch (2012). Como veremos a continuación, ese sector del pabellón es precisamente el menos informativo desde el punto de vista de la identificación.

Uno de los métodos más utilizados son los modelos de forma. Estos modelos buscan reconocer distribuciones de índices de forma característicos del objeto en estudio, en este caso puntual, la superficie de la oreja. Por ejemplo, Chen and Bhanu (2005), propusieron detectar regiones de la imagen con curvaturas locales grandes. A esta técnica la denominaron *step edge magnitude*. Luego se aplica un *template matching* con formas típicas del contorno externo de la oreja (outer helix and anti-helix). En Chen and Bhanu (2007) se redujo la cantidad de orejas candidatas detectando regiones de piel antes de realizar el *template matching*.

Una aplicación similar, basada en el análisis de forma fue Attarchi et al. (2008). En este trabajo se utilizan contornos para la detección de la oreja. Primero se ubica el contorno

exterior de la oreja, buscando el borde interconectado más largo en la región de interés. Una vez ubicado, se realiza un triángulo de referencia basado en los puntos extremos del contorno. Finalmente, se utilizan propiedades del triangulo, con el baricentro, como un punto de referencia para el alineamiento de la imagen.

Por otro lado, Ansari and Gupta (2007) propuso el uso de las propiedades de los bordes, como por ejemplo concavidad o convexitud, para determinar localmente las partes de la oreja. De manera similar, Prakash and Gupta (2012) combinan segmentación de piel y bordes jerárquicos. Luego de detectados los bordes ubicados sobre regiones con piel, éstos son fraccionados en segmentos. En base a estos segmentos se crea un grafo de conectividad, integrando todos estos segmentos obtenidos previamente. Este grafo es utilizado para computar el *convex hull* del conjunto de segmentos que determinan la forma externa de la oreja.

Un aporte significativo se puede encontrar en Yan and Bowyer (2007), quienes desarrollaron un método de detección de orejas. Este método comienza detectando la concha (parte anatómica interna de la oreja, ver en Figura 6.1), la cual es utilizada como punto inicial para ASM (definido en la Sección 2.1), el cual es utilizado finalmente para determinar el contorno externo de la oreja. Todos los métodos mencionados anteriormente tienen como desventajas que se tratan de aproximaciones *ad-hoc*, requiriendo una ingeniería de características muy específica, sumado a la falta de invariancia bajo homografías y cambios de luminancia, lo cual los hace muy poco robustos para trabajar en ambientes no controlados.

Ademas de los métodos de forma extensamente utilizados, se encuentran otras aproximaciones que se basan en ideas de reconocimiento de patrones. Entre ellas, podemos mencionar Abaza et al. (2010) y Islam et al. (2008), los cuales utilizan clasificadores débiles basados en *Haar-wavelets* sobre regiones de las imágenes para encontrar correlaciones con patrones aprendidos previamente. Estos clasificadores débiles son combinados con *AdaBoost* Freund Robert Schapire (1999) para la ubicación de la oreja. Estos enfoques, contienen un *workflow* más robusto, ya que dividen la etapa de procesamiento de imágenes de la de identificación, tomando la ventaja de los beneficios asociados a los métodos de reconocimiento de patrones. En la Tabla 6.4 se pueden ver los métodos estudiados ordenados por el tipo de modelos diseñados y el tipo de datos tomados como entrada.

Tabla 6.4: Tabla de clasificación de métodos y tipos de datos.

Referencia (Autor)	Resumen	Tipo de método	Tipo de dato
Chen and Bhanu (2005)	Correspondencia de plantillas basados en histogramas de índices de forma	Modelo de forma	2D
Attarchi et al. (2008)	Detección de bordes y seguimiento de líneas	Modelo de forma	2D
Chen and Bhanu (2007)	Modelo de forma de la Helix	Modelo de forma	3D
Ansari and Gupta (2007)	Detección de bordes y estimación de curvatura	Modelo de forma	2D
Prakash and Gupta (2012)	Correspondencia entre grafos y color de piel	Modelo de forma	2D
Yan and Bowyer (2007)	ICP using model points	Modelo de forma	2D y 3D
Pflug et al. (2013)	Fusión de características e Información de contexto	Modelo de forma	2D y 3D
Liu et al. (2015)	Angulo entre oreja y rostro como vector de características	Modelo de forma	3D
Abaza et al. (2010)	*Cascaded adaboost*	Reconocimiento de Patrones	2D
Islam et al. (2008)	*Adaboost*	Reconocimiento de Patrones	2D
Yuan et al. (2016)	Representaciones no negativas basadas en diccionarios	Reconocimiento de Patrones	2D
Kumar and Chan (2013)	Información de orientación basada en la transformada de Radon	Reconocimiento de Patrones	2D
Kumar and Wu (2012)	Codificación de fase con filtros de registro Gabor	Reconocimiento de Patrones	2D
Cintas et al. (2016a)	CNN y morfometría geométrica	Reconocimiento de Patrones	2D

6.5.2. Identificación basada en landmarks y ERT

Random Forests

Se ha demostrado que los bosques aleatorios (random forests) alcanzan una exactitud tan buena como *Adaboost* y algunas veces mejor. Además el modelo de aprendizaje es relativamente más robusto frente a valores atípicos y ruidosos. Proporciona estimaciones internas de error, correlación e importancia de variables. Su entrenamiento es mucho más rápido que otros métodos como *bagging* o *boosting* Dietterich (2000). Es además simple y fácil de paralelizar Breiman (2001). Los ERTs son árboles de clasificación (o regresión) en los cuales los atributos y opciones de corte son parcial o totalmente aleatorios, a la hora de ramificarse en un nuevo nodo durante entrenamiento Geurts et al. (2006). En un caso extremo, un ERT construye árboles de forma totalmente aleatoria. Estas estructuras son independientes del valor de salida de la muestra de entrenamiento. El beneficio de esta aleatoriedad radica en que se pueden afinar comportamientos específicos mediante una apropiada parametrización.

Experimentos

Se entrenó un ERT con la configuración de landmarks de 1458 individuos. Cada individuo cuenta con 4 o 6 imágenes contando ambas orejas. Se trabajó con un total de 8354 imágenes con landmarks generados automáticamente utilizando la red detallada en

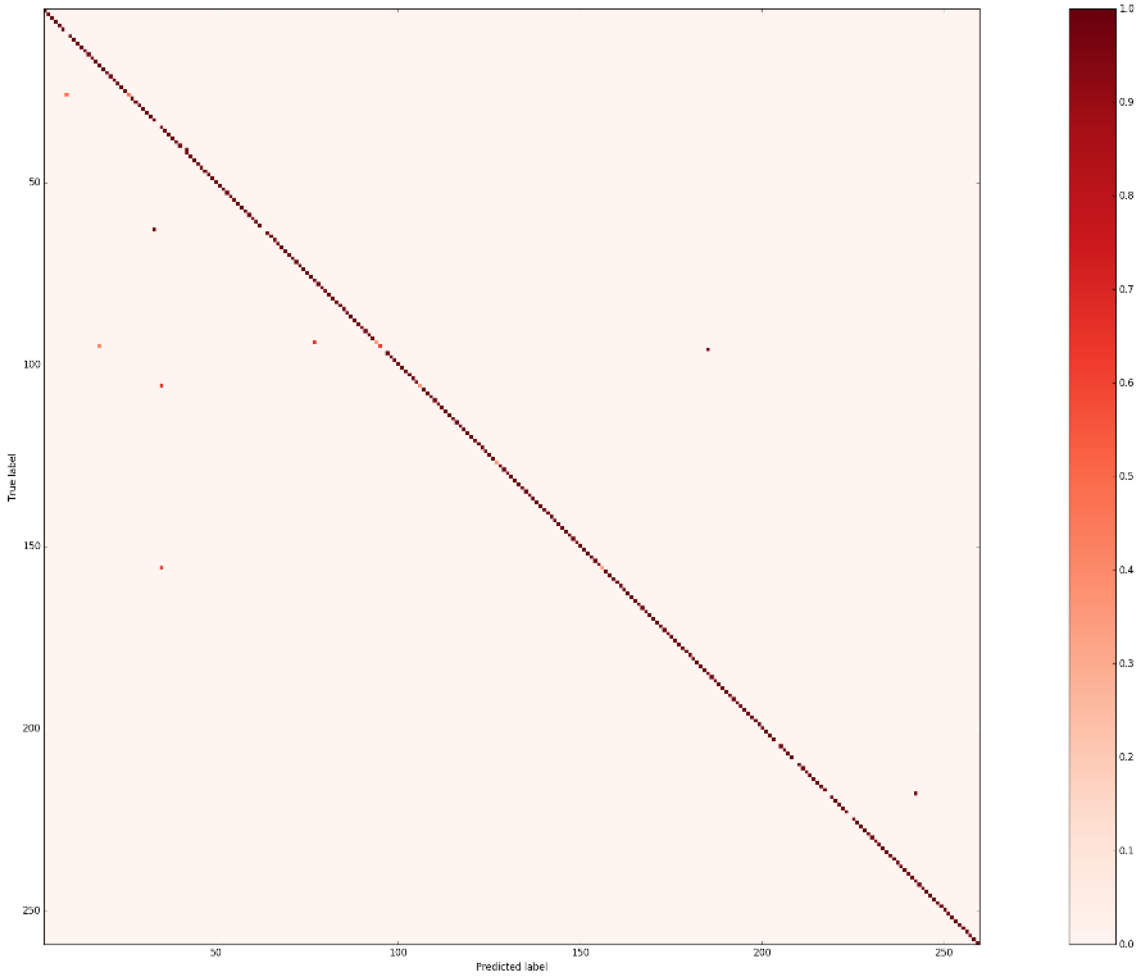


Figura 6.14: Matriz de confusión en el subconjunto de individuos id 5 al 248 Cintas et al. (2016a).

la Sección 6.3.3. Luego se aplicó *Generalized Procrustes Fit*, explicado en detalle en la Sección 3.2 para eliminar efectos de traslación, escala y rotación en las coordenadas de landmarks. En la Figura 6.13 se puede visualizar un subconjunto de las imágenes de orejas utilizadas en el entrenamiento.

El set de datos de entrenamiento cuenta con 6683 vectores de características v_i, formados por los 45 landmarks generados automáticamente ($v_i = [x_0, y_0, \cdots, x_{44}, y_{44}]$), y un valor de *target* t con la etiqueta asociada a el individuo. Las 1671 muestras restantes fueron resguardadas para realizar evaluaciones de performance. Los valores de reconocimiento obtenidos fueron (en promedio: *precision* 0,95, *recall* 0,90, *f1-score* 0,91 y *adjusted rand score ARI* 0,93).

Tabla 6.5: Accuracy del ERT en cada *fold* de entrenamiento.

Accuracy Score	# Fold
0.91202873	1
0.90125673	2
0.90125673	3
0.92040694	4
0.91083184	5
0.90604428	6
0.91322561	7
0.92339916	8
0.90724117	9
0.91861161	10

La matriz de confusión sobre un conjunto de datos de evaluación puede observarse en la Figura 6.14, y la curva ROC calculada puede visualizarse en la Figura 6.15. Para la evaluación se realizó un *K-fold* estratificado con 10 iteraciones y un tamaño de prueba del 20 % de la muestra. El valor promedio de exactitud fue del 0,9114 con un desvío estándar de $SD = 0{,}0146$. El valor para cada *fold* puede verse en detalle en la Tabla 6.5.

Los árboles de decisión son especialmente útiles a la hora de generar introspección sobre los datos, como éstos son clasificados, y la importancia de cada parámetro en el vector de características. Podemos obtener valores de pesos relativos o medir la importancia de cada dimensión en el espacio de características Wehenkel et al. (2006), tal como se explicó en la Sección 6.5.2. Gracias a esta propiedad fue posible analizar la contribución relativa de cada landmark en el proceso de reconocimiento, el detalle de las coordenadas. La lista de las diez características más importantes se puede ver en la Tabla 6.6. Como resultado se observó que las coordenadas de landmarks más importantes a la hora de selección corresponden a la parte anatómica interna de la oreja, lo cual sugiere que esta estructura es más informativa que la estructura externa (contorno de la oreja) a la hora de su uso como un posible valor discriminante.

Tabla 6.6: Peso relativo de coordenadas de landmarks en el proceso de reconocimiento. Para una referencia visual de la ubicación de los landmarks se puede ver la Figura 6.1.

Peso (en %)	# Landmark	Coordenada
1.5127	42	x
1.4552	36	y
1.4467	3	y
1.4442	43	x
1.3910	5	x
1.3798	40	y
1.3739	39	y
1.3671	35	y
1.3648	2	y
1.3641	1	y

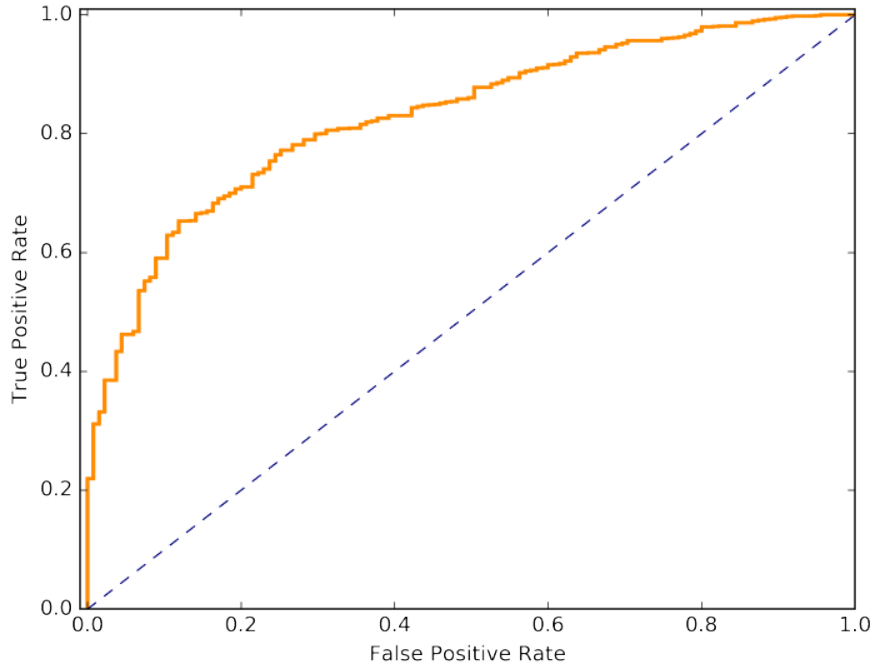


Figura 6.15: Curva ROC de identificación de individuos basados en la configuración de landmarks de pabellón auditivo.

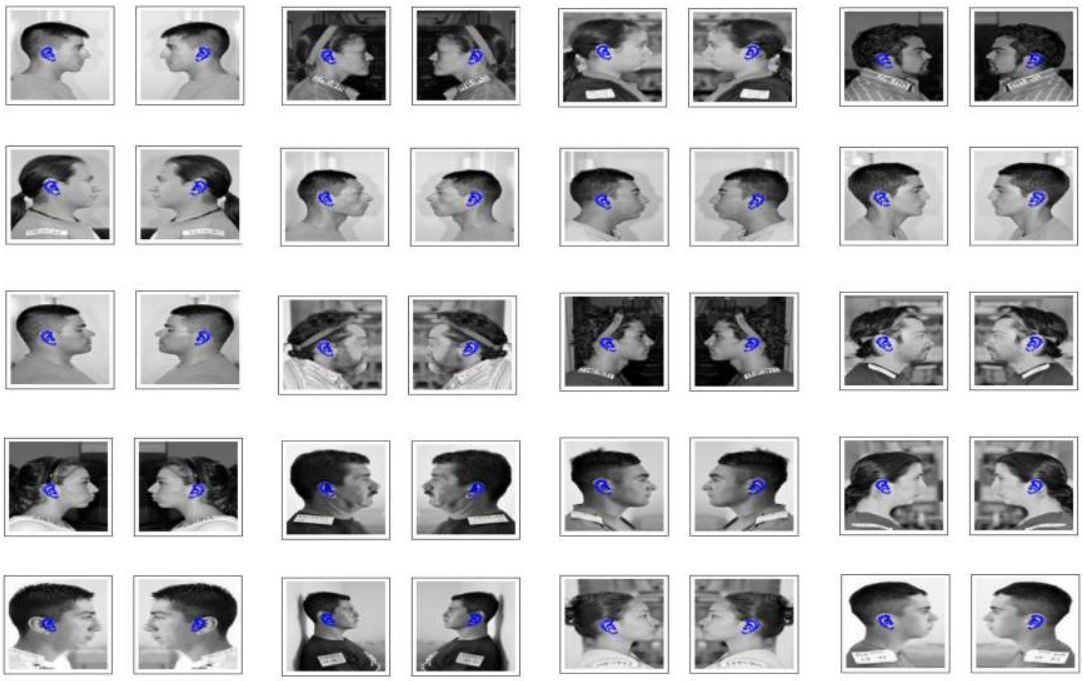


Figura 6.16: Imágenes procesadas automáticamente sobre la imagen completa.

6.6. Ubicación de orejas sobre imágenes faciales (CNN vs. Viola Jones)

Como se detalló en la Sección 6.3, en un comienzo se utilizó el método de Viola and Jones (2001) para ubicar las ROI en donde se encontraba la oreja, para aumentar la resolución del landmarking y no llevar partes innecesarias de la imagen a la CNN. A partir del trabajo Cintas et al. (2016a), se realizaron pruebas sobre la imagen completa para realizar landmarking. Si bien la calidad del mismo decrece, ya que se pierden detalles de la morfología del pabellón auditivo dada la resolución de la imagen, se puede usar como un detector de ROI dentro del rostro, utilizando la ubicación de los landmarks encontrados para calcular una ROI aproximada. En la Tabla 6.7 se puede ver una comparación de cantidad de ubicaciones de ROI correctamente ubicadas, no encontradas y ubicadas incorrectamente. Algunas imágenes con landmarks sobre el rostro completo se pueden observar en la Figura 6.16, en la esquina inferior se pueden ver algunas imágenes con landmarks posicionados incorrectamente.

Para validar este paso de pre-procesamiento, se tomaron de forma aleatoria 185 imágenes del conjunto de datos de CANDELA, en el 92,43 % de los casos la ROI fue encontrada

Tabla 6.7: Comparación entre el algoritmo de Viola-Jones y nuestra propuesta con CNN para la ubicación de ROI.

	ROI encontrada	ROI no encontrada	ROI mal ubicada
Viola Jones	92,43 %	1,62 %	5,95 %
CNN	97,29 %	0,0 %	2,70 %

correctamente, en el 1,62 % la ROI no fue encontrada y en el 5,95 % de los casos fue ubicada incorrectamente.

Capítulo 7

Landmarking Lateral

En este Capítulo abordamos otra aplicación del landmarking automático, en este caso el landmarking lateral o de perfiles de rostros. Se detallan las configuraciones de landmarks utilizadas, con su definición bio-antropológica bajo la estructura detallada en el Capítulo 3. Continuaremos con la descripción del *pipeline* utilizado para el landmarking automático, su diseño, implementación y pruebas realizadas para medir su desempeño. Además, analizaremos la selección de distintas configuraciones de landmarks y su impacto en el desempeño del landmarking automático y sus posibles repercusiones sobre la robustez de posibles vectores de características basados en landmarks y sobre estos ver cuáles landmarks cuentan con más información discriminante. Al final del Capítulo se presenta, como aplicación directa del landmarking automático lateral, la predicción de género a partir de una imagen de vista lateral. Se define el modelo y se muestran los resultados del mismo.

7.1. Configuración de landmarks: Vista Lateral

El tejido blando de la cara humana es una compleja geometría, compuesta por varios órganos, incluyendo, ojos, nariz, orejas, boca, etc. Dadas sus funciones biológicas principales, el rostro humano es un tópico central en varias investigaciones, con un gran rango de aplicaciones, incluyendo, antropología Gómez-Valdés et al. (2013), Quinto-Sánchez et al. (2015a), Schlager and Rüdell (2015), Paschetta et al. (2016), medicina genética Hammond (2007), Hammond et al. (2005), Weinberg et al. (2008), ciencias forenses

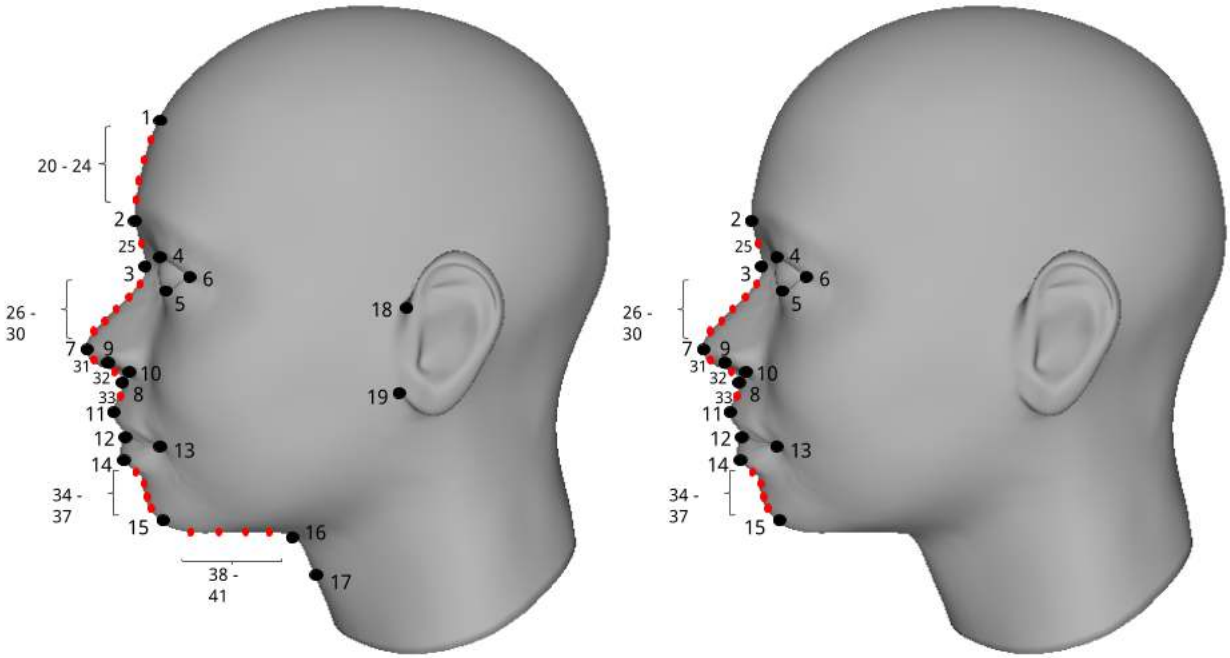


Figura 7.1: Configuraciones de Landmark y semi-landmarks y descripción anatómica.

Alexander et al. (2011), Kurniawan et al. (2014), Liu et al. (2015), Albert et al. (2007), envejecimiento Ramanathan et al. (2009), Fu et al. (2010) y genómica cuantitativa Liu et al. (2012), Adhikari et al. (2016).

Sin embargo, por un largo período de tiempo, importantes variables cuantitativas sobre rostros humanos no han podido ser utilizadas en toda su potencialidad, dado que estos estudios están basados sobre tediosas medidas manuales tomadas a partir de un set de coordenadas, las cuales son determinadas subjetivamente por el observador, dando lugar a errores de captura de datos entre individuos Segev et al. (2010), Kamoen et al. (2001). En la Tabla 7.1 se pueden observar los landmarks y semilandmarks definidos sobre la vista lateral, y cuáles de ellos fueron incluidos en las configuraciones evaluadas en este trabajo. Para una descripción gráfica, estas configuraciones se pueden ver en la Figura 7.1.

Tabla 7.1: Configuración de Landmarks y semi-landmarks sobre vista lateral.

Número	Nombre	Conf # 41	Conf # 27
1	Trichion	X	-
2	Sellion	X	X
3	Nasion	X	X
4	Palpebrale superiorus	X	X
5	Palpebrale inferiorus	X	X
6	Exocanthion	X	X
7	Domus	X	X
8	Subnasal	X	X
9	superior nostril axis	X	X
10	Inferior nostril axis	X	X
11	Labiale superious	X	X
12	Stomion	X	X
13	Cheilion	X	X
14	Labiale inferious	X	X
15	Gnathion	X	X
16		X	-
17		X	-
18	Otobasion superiorious	X	-
19	Otobasion inferiorous	X	-
20-24	Semi-landmarks	X	-
25	Semi-landmarks	X	X
26-30	Semi-landmarks	X	X
34-37	Semi-landmarks	X	X
38-41	Semi-landmarks	X	-

7.2. Conjuntos de datos de Landmarking Lateral

El subconjunto de datos utilizados para el landmarking lateral cuenta con 1658 imágenes para el entrenamiento, cada una de ellas con 41 landmarks y semi-landmarks provistos

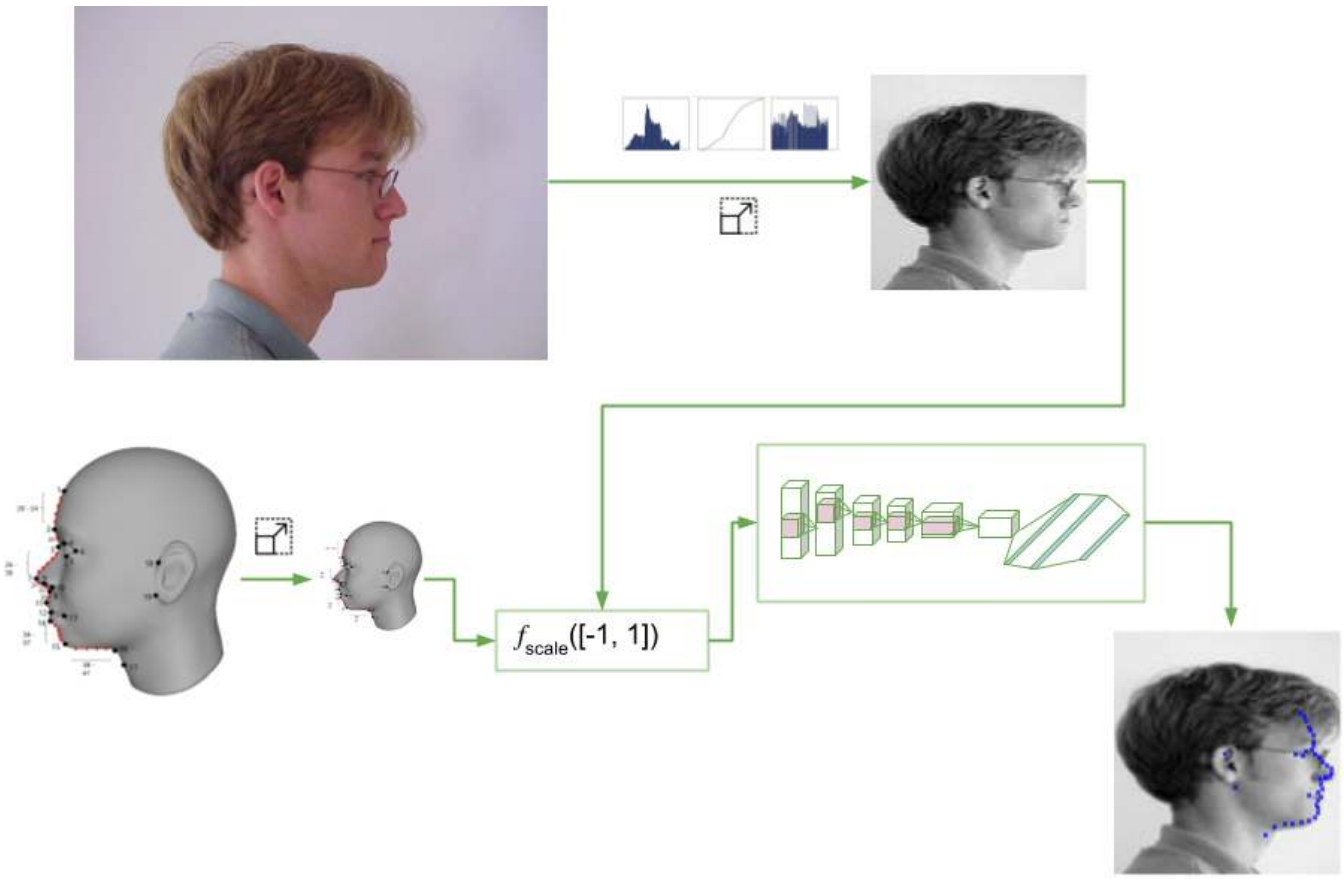


Figura 7.2: Visión general del *Pipeline* desarrollado para landmarking automático sobre la estructura de la vista lateral de la cara

por especialistas. También se cuenta con un conjunto de ejemplos de validación de 415 imágenes (25 % de la muestra total), tomadas con una permutación aleatoria utilizando *cross-validation*. El tipo de imagen consiste en la vista lateral de la persona con una resolución de 2136 × 3216 píxeles. Se puede observar un ejemplo de estas imágenes en la Figura 5.1.

7.3. Pipeline Desarrollado

Se entrenó una CNN con el conjunto de datos mencionado arriba, y además se realizaron pruebas con configuraciones basadas en subconjuntos de los 41 landmarks originalmente provistos. En esta sección detallamos todos los pasos del proceso, y por cada paso, describimos los formalismos utilizados.

7.3.1. Preprocesamiento de Imágenes

A diferencia del pipeline de pabellón auditivo visto en el Capítulo 6, en este caso no necesitamos ubicar ninguna ROI previa, ya que se trabaja con toda la imagen. A la imagen de entrada se le aplica una normalización de histograma, mediante el cual los valores de luminancia en la imagen se extienden para cubrir la mayor parte del rango dinámico. Los parámetros de estiramiento del histograma fueron programados para convertir en negro el 2 % de los píxeles y a blanco el 1 % de los píxeles como máximo en ambos casos. Cómo último paso, la imagen es remuestreada al tamaño final utilizado por la CNN de 96×96 píxeles mediante el uso de submuestreo bilineal. En la Figura 7.2 se puede observar el efecto de los pasos descriptos anteriormente.

7.3.2. Preprocesamiento de los landmarks

Los landmarks manuales fueron generados desde el software TPSdig, como se mencionó en la Sección 5.2. Este programa ubica el origen en el extremo superior izquierdo, por lo que por comodidad a la hora de implementación, antes de realizar cualquier modificación se invirtieron en el eje y todas las coordenadas para contar con el origen en la parte inferior izquierda. Las coordenadas se toman de un archivo con el siguiente formato:

```
id x00 y00 x01 y01 ... xnn ynn
```

Estas coordenadas se encuentran en el espacio de la imagen original de 2136×3216 pixels. Estas coordenadas tuvieron que ser escaladas por el mismo factor utilizado en la etapa de preprocesamiento de la imagen visto en la Sección 7.3.1. En términos generales se aplicaron las siguientes funciones:

$$(x_{new}, y_{new}) = (\frac{x_{old}}{dx}, \frac{y_{old}}{dy}), \tag{7.1}$$

donde x_{old} e y_{old} son las coordenadas originales respecto a la imagen completa. Los factores de escala se mantuvieron como variables diferentes, dado que si se utiliza la imagen completa al no ser cuadrada, los factores serán distintos.

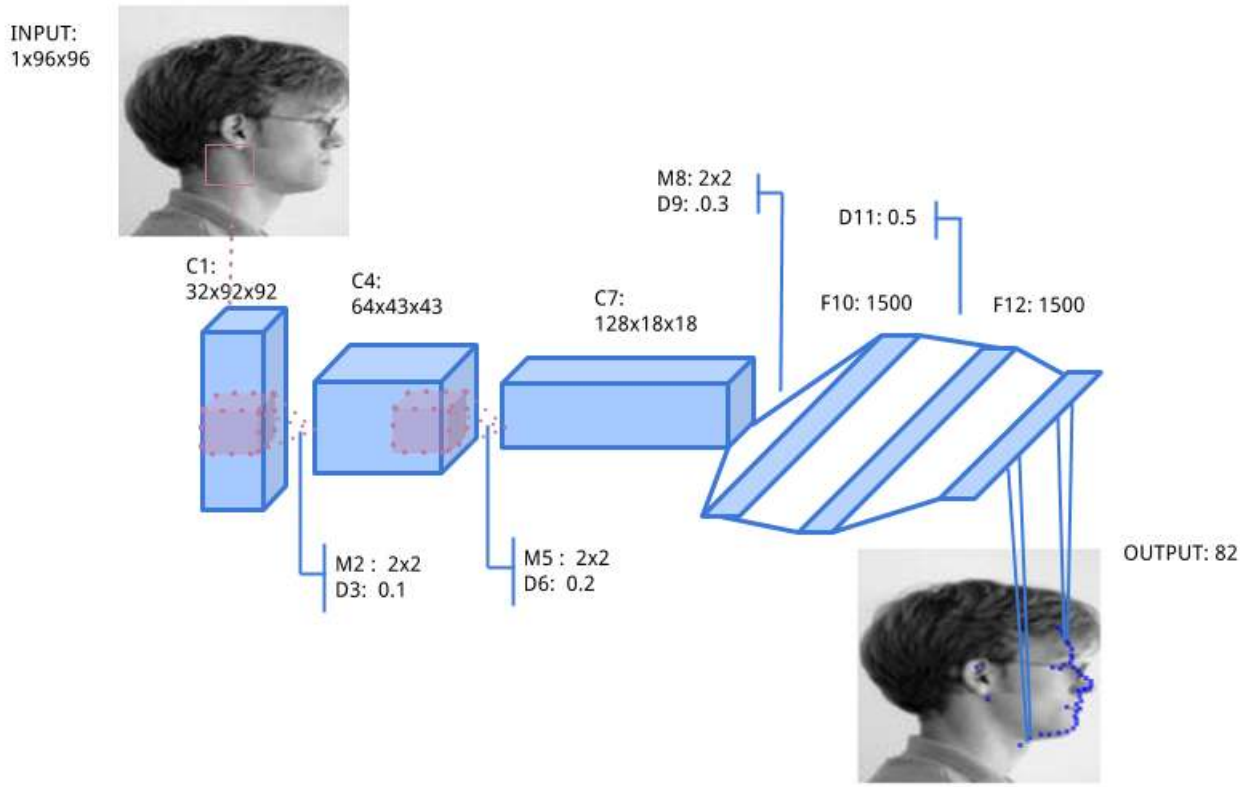


Figura 7.3: Arquitectura de CNN que obtuvo el mejor desempeño en landmarking lateral.

7.3.3. Elección de Arquitectura de ConvNet

Varias arquitecturas fueron implementadas, por lo que en esta Sección mostraremos la que obtuvo los mejores resultados. Estas arquitecturas fueron diseñadas y entrenadas con el propósito de realizar landmarking automático, principalmente para detectar e identificar partes anatómicas de la vista lateral sobre imágenes. Más allá de estudiar las arquitecturas de redes en este caso, consideramos que es interesante el análisis de la elección de distintas configuraciones de landmarks sobre una misma arquitectura para analizar variaciones en el entrenamiento y desempeño de la red. Como datos de entrada se tomaron imágenes de rostros, desde la vista lateral en escala de grises (un único canal) de 96×96 píxeles, estos valores fueron escalados a $[0, 1]$.

En la Figura 7.3 se puede observar la arquitectura de la red con mejor desempeño (*Arch0*). La arquitectura subyacente consiste en una capa de convolución con filtros cuadrados, seguida de una capa de *max-pooling* y *dropout*. Esta estructura es repetida tres veces para obtener características en distintos niveles de abstracción, con diferentes tamaños de filtros, cantidad de mapas de características y valores de probabilidad para las

capas de *dropout*.

Siguiendo la Figura 7.3, las capas de convolución C1, C2 y C3 tienen 32, 64 y 128 filtros de tamaño 5×5, 4×4 y 4×4 respectivamente. Todas las capas de *max-pooling* son de 2×2, y los valores de probabilidad usados para D1, D2 y D3 son, respectivamente, 0,1, 0,2 y 0,3. En la Figura 7.4 se pueden observar los filtros de la capa C1 y sus correspondientes mapas de características. Algo que podemos observar en ellos es que ninguno de sus nodos quedo .ªpagado", por lo que la tasa de aprendizaje es apropiada. Luego de la etapa de extracción de características, la arquitectura contiene dos capas densas con 1500 unidades cada una (F13 y F15 en el diagrama), y una capa de *dropout* (D14). La capa de salida contiene 82 unidades de salida (41 $[x, y]$ pares) para la posición de los landmarks predichos.

Al igual que la sección anterior, la implementación fue realizada en Python y Lasagne Dieleman et al. (2015b)[1]. Esto nos permite el uso de aceleración mediante GPU de una forma sencilla. El entrenamiento de esta red tomó 25 horas aproximadamente sobre una placa NVIDIA GeForce GTX 590. Una vez entrenada, la red puede ser utilizada en hardware convencional, incluso en sistemas embebidos.

[1]El código está disponible en https://github.com/celiacintas/tests_landmarks/blob/master/testing_output.ipynb.

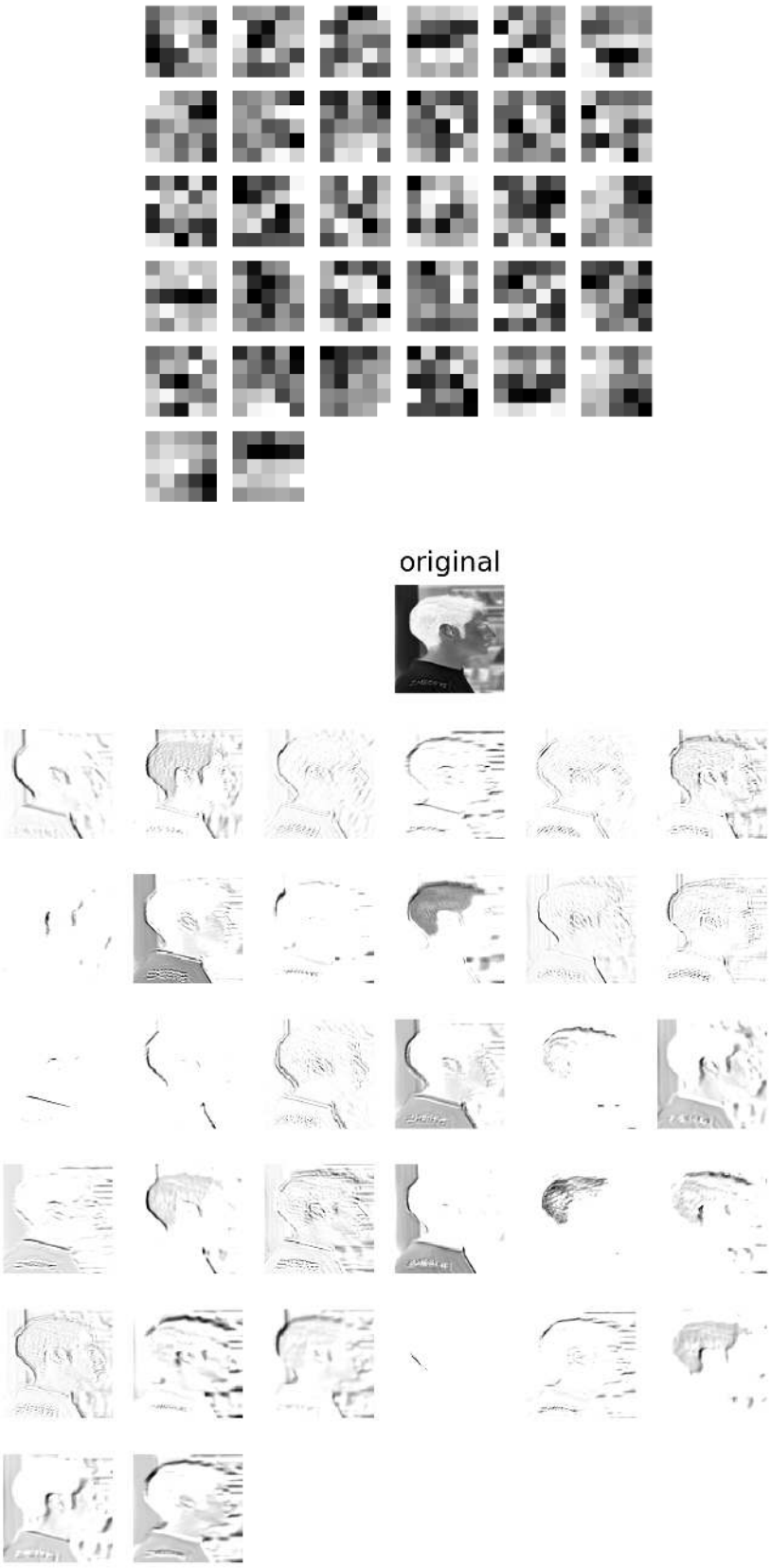


Figura 7.4: *Kernels* y mapas de características de la capa C1 de la red *net0* sobre una imagen de entrada X.

7.4. Resultados sobre landmarking lateral

El problema de ubicación de landmarks en forma automática puede ser pensado como un problema de regresión. Al usar este enfoque aplicamos métricas para evaluar el desempeño de diferentes CNNs contra el landmarking manual (ground truth). En particular se trabajó con r^2, error cuadrático medio (RMSE) y correlación de Pearson (detalladas en la Sección 5.3). Como se observa en la Figura 7.5, podemos notar que la correlación de Pearson es más dispersa en la *Net 1 # 41 Land. Conf* que en la *Net 0 # 27 Land. Conf*. Esto se debe a que la primera configuración de landmarks (Ver Figura 7.1), contiene landmarks ruidosos, difícilmente homólogos entre individuos, por lo que al investigador experto se le hace difícil localizarlos. La base de conocimiento arrastra este tipo de ruido, el cual se refleja en la relativamente baja correlación obtenida luego del entrenamiento de la red. En las otras métricas de calidad podemos corroborar que este problema se ve también reflejado, aunque en menor medida.

El desempeño del landmarking de las arquitecturas implementadas con diferentes configuraciones de landmarks y parámetros de base se puede observar en la Tabla 7.3. Se utilizó como línea de base un regresor básico que tiene como política predecir el promedio de los valores de entrenamiento. En la Tabla 7.2 se detallas los RMSE para cada coordenada de landmark.

Las métricas de regresión fueron calculadas utilizando Pedregosa et al. (2011). También se pueden observar en la Figura 7.6 las curvas de error sobre los datos de entrenamiento y validación de las dos arquitecturas. La Figura 7.7 muestra los landmarks predichos sobre imágenes no vistas por la red *#41 Land. Conf.* (el conjunto de validación). Se puede observar que alguna de estas imágenes cuentan con fondos heterogéneos e iluminación no controlada (por ejemplo las siluetas de las sombras).

Algunos resultados sobre conjuntos de datos externos a CANDELA se pueden ver en la Figura 7.8, si bien no se contaban con los landmarks para realizar métricas comparativas, se puede observar potencial sobre imágenes de otras fuentes externas a las de entrenamiento y validación.

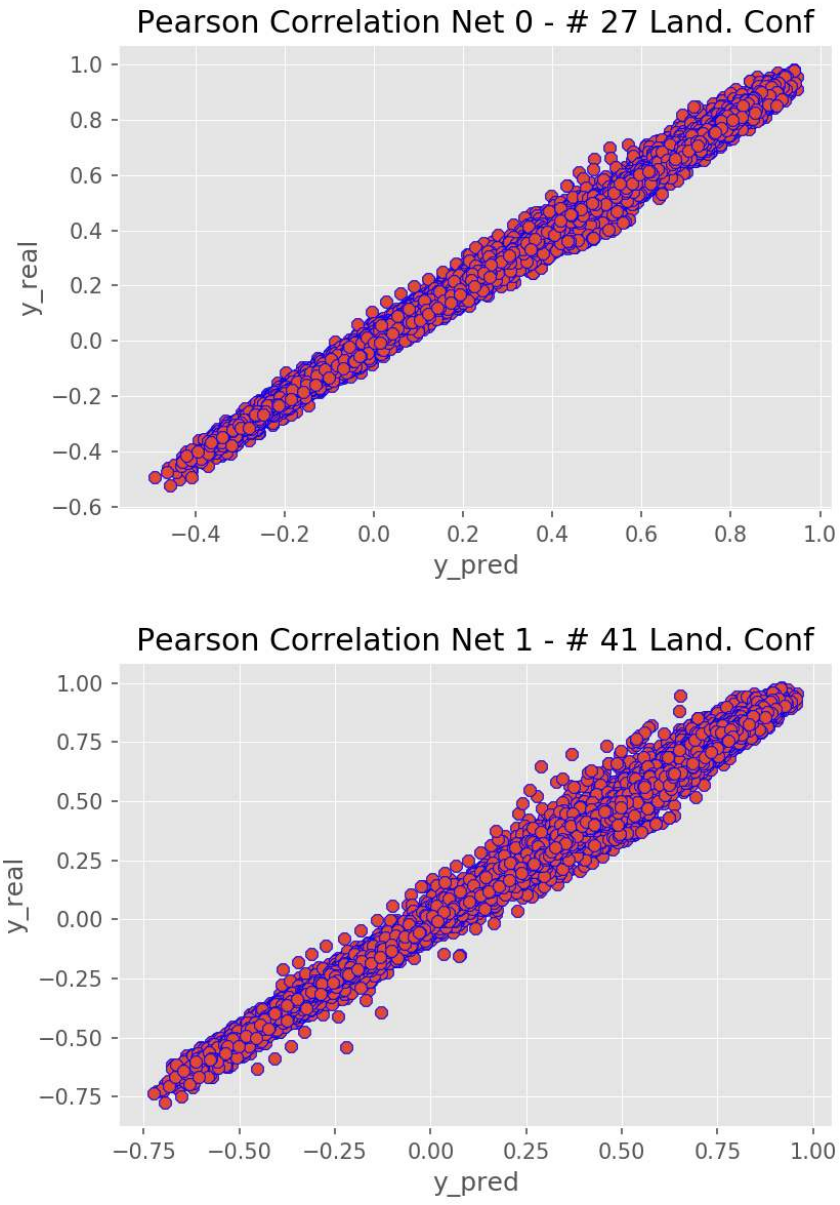


Figura 7.5: Correlación de Pearson sobre datos reales vs predichos de las dos redes con diferentes configuraciones de landmarks

Tabla 7.2: RMSE de cada Landmark y Semi-Landmark de la Configuración de Vista Lateral.

Landmark #	RMSE
1	1.13398
2	1.07418
3	0.903374
4	0.877164
5	0.921844
6	1.1937
7	1.06892
8	1.11192
9	1.07502
10	1.12419
11	1.13046
12	1.14933
13	1.21603
14	1.65964
24	1.05838
25	1.07302
26	1.08066
27	1.08303
28	1.10548
29	1.13774
30	1.14198
31	1.10529
32	1.02747
33	1.27866
34	1.37326
35	1.44251
36	1.53532

Tabla 7.3: Desempeño de CNNs sobre dos configuraciones de landmarks distintas y la linea base.

	Arch0 # 41 Land. Conf	*Arch1 # 27 Land. Conf.*	Baseline
r^2	0.9276	**0.9384**	*0.001*
RMSE	1.3512	**1.2182**	*5.026*

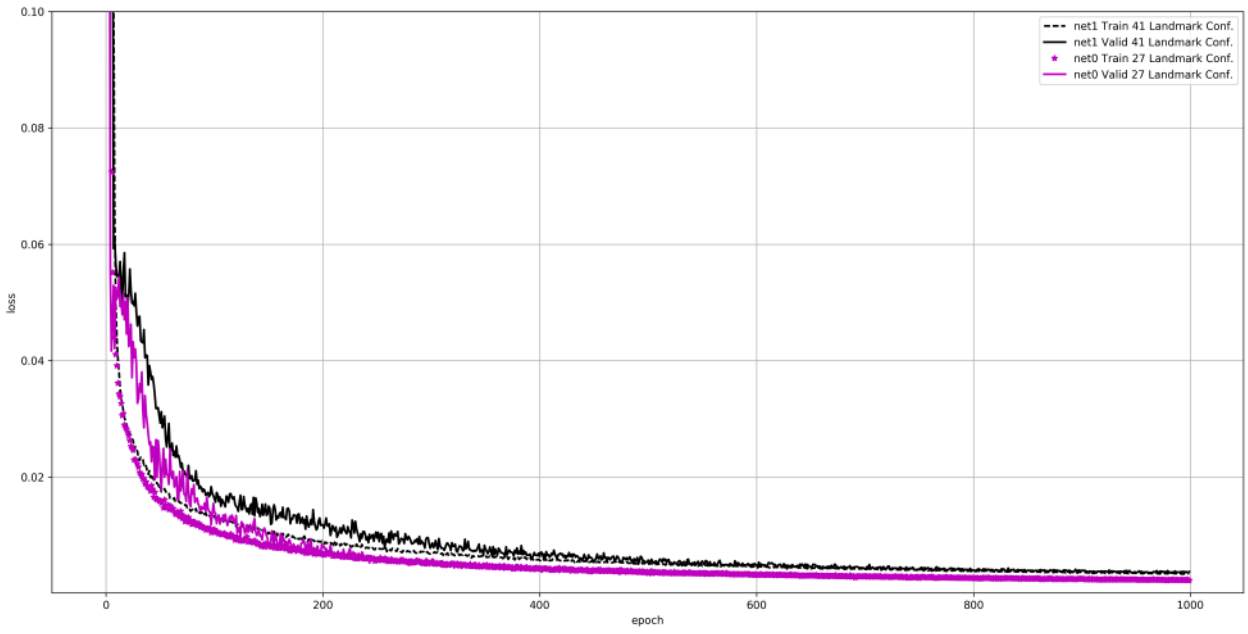


Figura 7.6: Curvas de pérdida para CNNs #41 Land. Conf. y #27 Land. Conf.

Figura 7.7: Resultados de nuestra mejor red sobre imágenes no vistas en la etapa de entrenamiento.

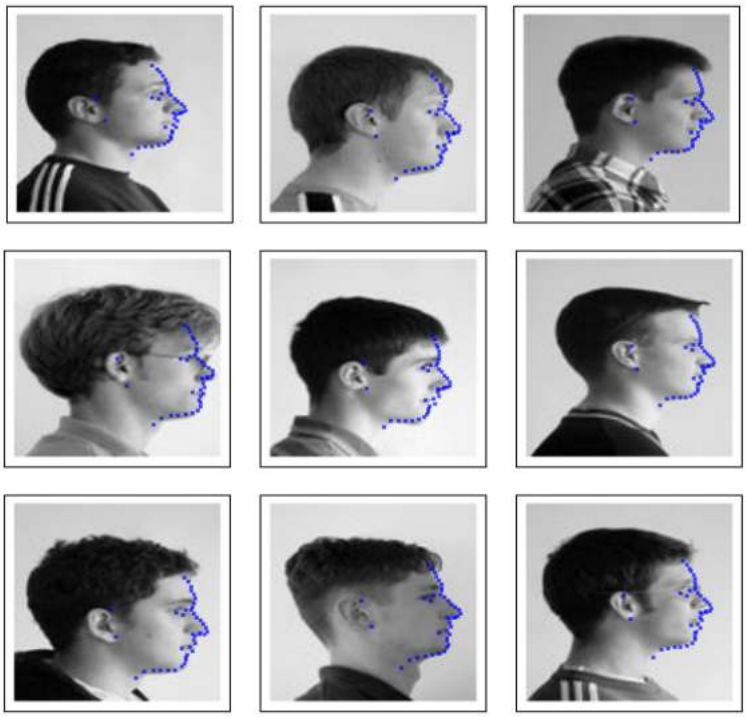


Figura 7.8: Resultados de utilizar la mejor arquitectura sobre imágenes pertenecientes a una base de datos externa (Solina et al., 2003), ver Sección 5.2 para más detalle.

7.5. Clasificación de género vía landmarking automático

La determinación de género a partir de imágenes faciales ha despertado el interés en el ámbito de la investigación básica y aplicada. Desde el punto de vista de la investigación básica, se busca el entendimiento de cómo el ser humano reconoce de manera inmediata el género de un individuo Bruce et al. (1993) mientras que desde la visión computacional sigue siendo un desafío Ng et al. (2015). Entre algunas de las dificultades que emergen se pueden enumerar: la posición de la cámara, pose del rostro, iluminación, fondos, oclusión parcial (cabello, anteojos, gorros), etc. Sumándole a esto el factor intrínseco de la variabilidad humana, etnicidad, edad y expresiones faciales. Desde la perspectiva aplicada de la investigación, se tiene un interés biométrico Mäkinen and Raisamo (2008), de seguridad y vigilancia, adaptabilidad de interfaces hombre-maquina. En particular, en el ámbito comercial se busca tener publicidad inteligente diseñada para atraer la atención de determinado género o sistemas de recolección de datos para análisis de mercado Caldwell (2011).

7.5.1. Trabajos previos

Dado que la mayoría de las aproximaciones a la determinación de género (binaria) son dadas sobre imágenes de la cara desde vista frontal con ciertos grados de rotación Ng et al. (2015), Gutta et al. (2000), Jain et al. (2005), Tivive and Bouzerdoum (2006), Alexandre (2010), Hussain et al. (2013), se dará un breve resumen del estado del arte de estas técnicas (para un detallado panorama se recomienda Ng et al. (2015)) y también se detallará los avances que hay sobre el reconocimiento de personas en vista lateral, dado que son potenciales técnicas a aplicarse en la determinación de género sobre vista lateral.

El reconocimiento de género sobre imágenes faciales respeta un *workflow* genérico, el cual comienza con (a) la detección del rostro en la imagen, se continúa con (b) preprocesamiento de la ROI, (c) extracción de características y finalmente (d) una clasificación binaria. Para el paso (a) se busca tener como salida una ROI en donde sólo se encuentren píxeles correspondientes al rostro, usualmente se utiliza Viola and Jones (2001), para una revisión detallada de métodos de detección de caras el lector se puede referir a Zhang and Zhang (2010). Luego de que se obtiene la ROI se procesan los píxeles

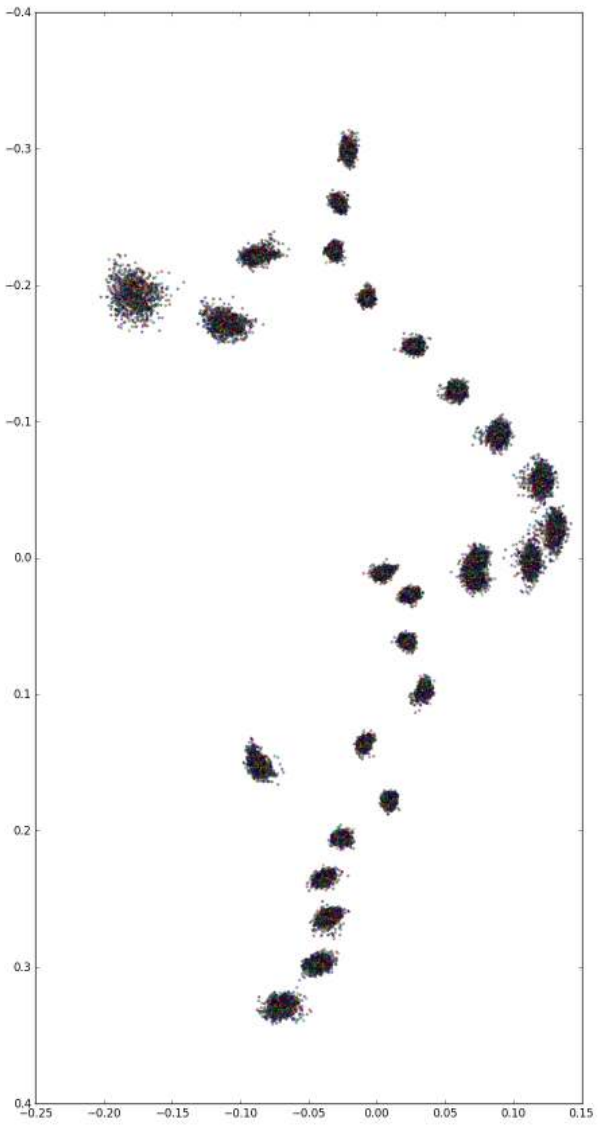


Figura 7.9: Configuración de # 27 landmarks sobre 1000 imágenes luego de aplicar GPA utilizadas en el entrenamiento del ERT.

antes de proseguir con el extractor de características. Algunos procedimientos clásicos de (b) suelen ser:

- Eliminar píxeles correspondientes a cabello o cuello, para que sólo queden píxeles asociados a la cara.
- Alineación geométrica de rostros Mäkinen and Raisamo (2008).
- Normalización de la imagen mediante ecualización del histograma.

En el paso (c) se busca obtener descriptores representativos por lo que las características

más discriminantes son llevadas a clasificación. Este paso suele ser separado en dos grandes categorías, por un lado, las basadas en geometría y los modelos basados en apariencia BenAbdelkader and Griffin (2005), Li et al. (2012). Los métodos basados en geometría suelen utilizar landmarks para describir la forma del objeto a clasificar mientras que los modelos basados en apariencia toman información de la textura de la imagen. A continuación se listan los métodos de extracción de características más utilizados:

- Distancias entre landmarks Brunelli and Poggio (1993), Fellous (1997).
- Intensidad de píxeles Baluja and Rowley (2007), Abdi et al. (1995).
- Características rectangulares Viola and Jones (2001) (Visto en detalle en la Sección 2.3).
- Patrones locales binarios (LBP) Ojala et al. (2002).
- Características SIFT Lowe (1999, 2004) (Visto en detalle en la Sección 2.2).

Finalmente tenemos el paso de clasificación (d), el cual recibe las características discriminantes obtenidas en (c) y se utiliza un modelo de clasificación binaria para determinar el género. Algunos ejemplos de los clasificadores más ampliamente utilizados para determinación de género suelen ser SVM, Adaboost, redes neuronales, etc.

Hasta aquí vimos un panorama de la clasificación de género a partir de imágenes faciales frontales, a continuación veremos el estado del arte de reconocimiento de individuos utilizando información de la vista lateral.

La mayoría de las investigaciones realizadas para reconocimiento de personas sobre imágenes en la vista lateral son basadas en textura o en información geométrica. Los métodos basados en textura trabajan directamente sobre los datos en crudo, accediendo a cada valor de pixel y su vecindad, para obtener un vector de características basados en texturas relevantes, por ejemplo, segmentos de bordes entre regiones con distintos descriptores de textura. Dado que estas operaciones no son invariantes bajo cambios geométricos o de iluminación, el desempeño de detección de perfiles está fuertemente ligado a las condiciones de la toma de los datos (posición de la cámara, iluminación de ambiente, etc.), es decir, no trabajan adecuadamente bajo ambientes no controlados, (in the open).

Por otro lado, los métodos geométricos utilizan características de la forma y tamaño de la cara humana para realizar *pipelines* más robustos ante las variaciones enumeradas en el párrafo anterior. Estos métodos se basan en reconocer características específicas, como los landmarks anatómicos, o en buscar equivalencias entre la imagen obtenida y plantillas conocidas.

La vista lateral de la cara como característica para reconocimiento de personas fue propuesto por Galton (1910). Harmon and Hunt (1977) publicó un trabajo pionero en reconocimiento sobre vista lateral. La geometría allí utilizada para la vista lateral fue un vector de características de 10 dimensiones, basado en la ubicación de 9 puntos fiduciarios sobre el perfil del rostro. En trabajos posteriores se incrementó la cantidad de puntos fiduciarios a 11, para el armado de un vector de características más robusto, reportando una exactitud del 96 % (utilizando landmarks manuales). Bhanu and Zhou (2004) propuso un método basado en correspondencias de contorno y "distorsión" de tiempo dinámica para reconocimiento de caras en vista lateral. De acuerdo al valor de curvatura de cada punto en la vista lateral, se ubican los landmarks pronasal y nasion. Éstos son luego utilizados para encontrar la ROI del rostro. Se recorre el contorno del rostro y se computa la curvatura local utilizando *time warping*. Partiendo de una galería de imágenes laterales previa, se calcula un valor de similitud entre el nuevo perfil y los ya existentes en la galería. Este método fue evaluado en dos diferentes bases de datos de vista lateral, reportando una tasa de reconocimiento del 90 % en los mejores casos.

Zhou and Bhanu (2005) extendieron su trabajo a reconocimiento facial (en vista lateral) sobre vídeos. Esto se hizo gracias a la reconstrucción de imágenes de alta resolución basadas en una secuencia de fotos de baja resolución. En estos métodos es crucial calcular los puntos basados en la curvatura dado que el desempeño depende de pre-procesamiento de escala-espacio Kakadiaris et al. (2008). Lipoščak and Loncaric (1999) también presentaron un método de filtrado sobre escala basado en landmarks que mostraba una tasa de reconocimiento de 90 %.

Todos los trabajos previos encontrados en la literatura realizan sus evaluaciones sobre conjuntos de datos muy pequeños, entre 30 a 44 personas, contando con 60 a 290 imágenes en total. Además se utiliza una cantidad limitada de landmarks (entre 5 a 20 puntos fiduciarios), ubicados mayoritariamente en áreas donde las expresión facial puede

alterar el resultado. Una configuración más densa de landmarks, como se presenta en este capítulo, produce una representación más robusta del contorno del rostro, la cual reduce significativamente los efectos de las expresiones faciales.

7.5.2. Solución Propuesta

Al igual que en el Capítulo anterior, agregamos un *Extremely Randomized Tree (ERT)*, como una última etapa de clasificación. Se entrenaron dos ERT diferentes con las dos configuraciones de landmarks de 1907 individuos, con 4 o 6 imágenes por individuo (ambos lados). Se trabajó con un total de 10843 imágenes con landmarks generados automáticamente con la red detallada en la Sección 7.3.3, un total de 6255 imágenes de hombres y 4588 de mujeres. Luego se aplicó *Generalized Procrustes Fit* (GPA) para eliminar efectos de traslación, escala y rotación en las coordenadas de landmarks. En la Figura 7.9 se puede observar un subconjunto de los perfiles utilizados en el entrenamiento obtenidas de la red *Net 0 # 27 Land. Conf*. El set de datos de entrenamiento cuenta con 7590 (4368 hombres y 3222 mujeres) vectores de características v_i, formados, uno por 41 landmarks generados automáticamente ($v_i = [x_0, y_0, \cdots, x_{41}, y_{41}]$) y un valor de *target* t con la etiqueta asociada a el individuo y el otro con 27 landmarks. Las 3253 muestras restantes fueron resguardadas para realizar evaluaciones de desempeño. Los valores de clasificación obtenidos para *# 27 Conf. Land.* fueron (en promedio: precisión 0,84, exactitud 0,83, *recall* 0,83, *f1-score* 0,80.

El diagrama de Sankey sobre los datos de validación se pueden observar en la Figura 7.10, este nos indica la cantidad de muestras correctamente e incorrectamente predichas. Para la evaluación se realizó un *K-fold* estratificado con 10 iteraciones y un tamaño de prueba del 20 % de la muestra. El valor promedio de exactitud fue del 0,836 con un desvío estándar de $SD = 0{,}0090$. El valor para cada *fold* para *# 27 Conf. Land.* puede verse en detalle en la Tabla 7.4. Para el propósito de determinación de género también se compararon los dos diferentes vectores de características, los resultados se muestran en la Tabla 7.5.

Al igual que en el Capítulo anterior, se analizó la contribución relativa de cada landmark en el proceso de clasificación utilizando árboles de clasificación. En la Figura 7.11 se pueden visualizar todos los landmarks con sus pesos correspondientes, en violeta pode-

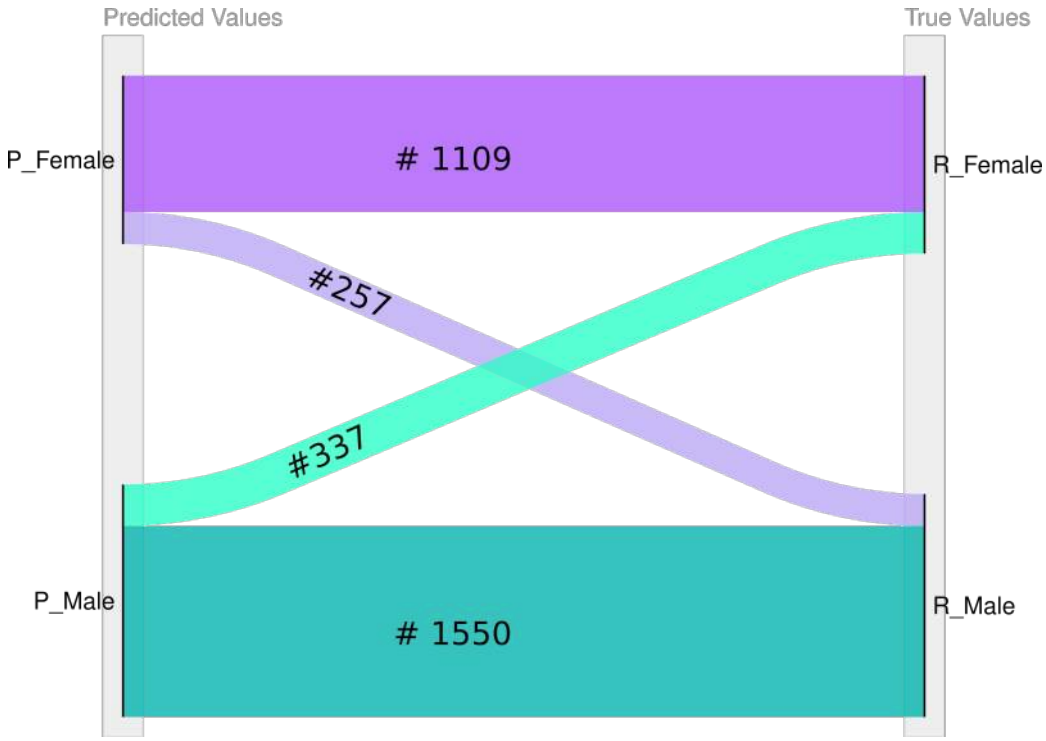


Figura 7.10: Diagrama de Sankey, visualizando el flujo de elementos de validación clasificados correcta e incorrectamente.

Tabla 7.4: Exactitud lograda para cada *fold* del ERT utilizando la *# 27 Conf. Land.*.

Exactitud	# Fold
0.82203781	1
0.82757031	2
0.84232365	3
0.85016136	4
0.84324574	5
0.82941448	6
0.83909636	7
0.82664822	8
0.84186261	9
0.84601199	10

mos observar coordenadas de landmarks que se encuentran en la red entrenada con *# 41 Conf. Land* y que fueron eliminados en *# 27 Conf. Land*. Cabe destacar que estos landmarks (en violeta), pertenecen a landmarks correspondientes al área de la frente y mentón

Tabla 7.5: Performance de las dos configuraciones de landmarks luego de una validación cruzada de 10 folds.

	Configuración # 27		Configuración # 41	
	Promedio	SD	Promedio	SD
exactitud	0.8368	0.0090	**0.8785**	0.0054
recall	0.8463	0.0157	**0.8806**	0.0104
f1	0.8146	0.0095	**0.8599**	0.0070

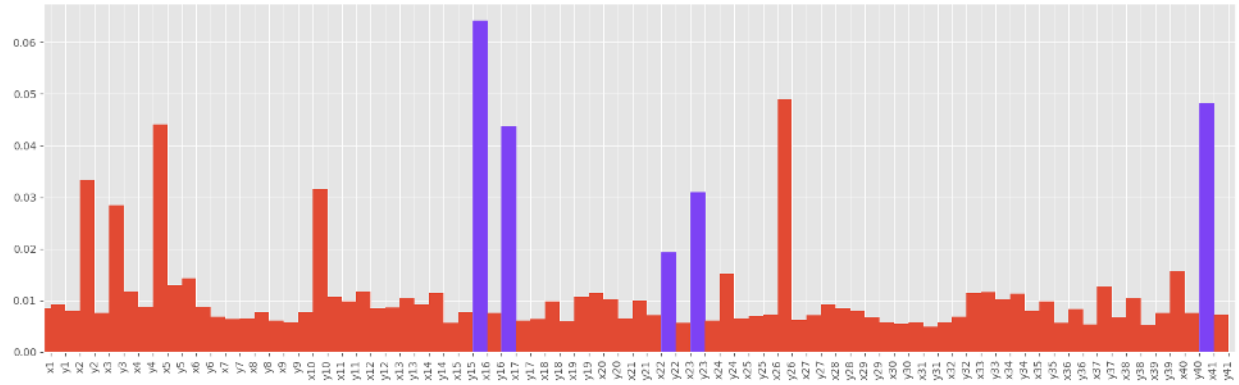


Figura 7.11: Gráfico de barras sobre la importancia de las coordenadas de landmarks a la hora de clasificación. En color violeta landmarks que se encuentran en la *# 41 Conf. Land* y fueron excluidos de # 27 Conf. Land.

(Figura 7.11), los cuales son potenciales indicadores de dimorfismo sexual[2] Velemínská et al. (2012), Kesterke et al. (2016). Esto provee una posible explicación del por qué la configuración *# 41 Conf. Land* es mejor como vector de características para clasificador de género.

[2]El dimorfismo sexual es definido como las variaciones en la fisonomía externa, como forma o tamaño, entre hombres y mujeres.

Capítulo 8

Landmarking Corporal 3D

El uso de dispositivos digitales 3D para capturar información relevante sobre la forma del cuerpo humano esta creciendo rápidamente y no se limita únicamente a estudios ergonómicos Han and Nam (2011), Park et al. (2015), diseño de ropa Paquette (1996), D'Apuzzo (2007), sino que ha tenido gran impacto en aplicaciones relacionadas a la salud Ben Azouz et al. (2006), Wells et al. (2015, 2008), Treleaven and Wells (2007). La toma de medidas antropométricas corporales es la base de muchas aplicaciones, incluyendo control del sobrepeso, *body-building*, deportología, diseño de indumentaria, *virtual try-on*, hasta verificación de identidad biométrica. Para evaluar parámetros cuantitativos de los cuerpos escaneados, es necesario , identificar landmarks, realizar determinadas medidas, y extraer descriptores geométricos de tamaño y forma, que puedan luego ser comparados entre diferentes sujetos y correlacionarlos, por ejemplo con datos de diagnóstico Lovato et al. (2014). Para la obtención de estas medidas se propuso una nueva aproximación desde la morfometría geométrica tomando un conjunto de landmarks definidos en la Tabla 8.1 y visualizados en la Figura 8.1, los cuales ademas de proveer las medidas clásicas de la antropometría, permiten extraer un conjunto más amplio de mediciones.

El landmarking 3D también puede ser utilizado como un paso intermedio a una posible reconstrucción 3D, dado que su ubicación puede servir como información para alinear un modelo 3D objetivo al recien obtenido Hirshberg et al. (2011). Un problema relevante a la hora de landmarking automático sobre cuerpos 3D, es que varios puntos anatómicos no siempre tienen una caracterización geométrica. La geometría local no sólo está pobremente caracterizada sino que suele variar entre sujetos con diferente estructura corporal.

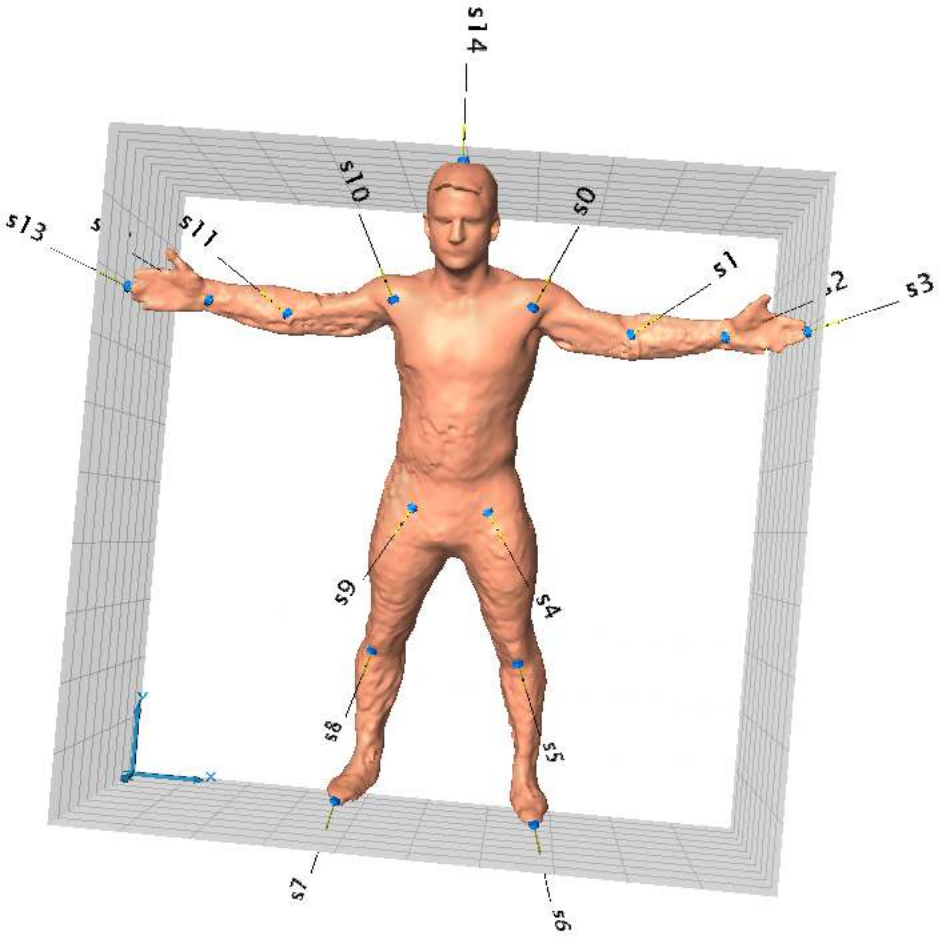


Figura 8.1: Configuración de landmarks corporales 3D.

Se necesita, por lo tanto, complementar las estrategias aplicadas en otros contextos con maneras novedosas de obtener evaluaciones precisas de la geometría local en 3D.

8.1. Configuración de landmarks corporales 3D

Para esta aplicación se requiere una configuración de landmarks que ayude a delimitar la topología del cuerpo humano. En pasos ulteriores, los landmarks son utilizados como guías para calcular parámetros geométricos de la superficie y servir de apoyo a las tareas de reconstrucción. Además, se utiliza una configuración de landmarks que proporciona de manera directa las distancias y otras medidas utilizadas en la antropometría clásica para el seguimiento de la forma corporal (en nutrición, monitoreo del sobrepeso, etc.).

Tabla 8.1: Configuración y definición anatómica de landmarks corporales 3D.

Número	Definición anatómica
0, 10	Articulación del húmero
1, 11	Extremo proximal del radio
2, 12	Punto extremo distal del radio
3, 13	Punto extremo del dedo mayor
4, 9	Punto de articulación de la cabeza del femur en la cadera
5, 8	Punto medio de la rótula
6, 7	Extremo distal segundo dedo
14	Punto extremo de la cabeza

8.2. El conjunto de datos 3D

El conjunto de datos cuenta con 450 modelos sintéticos (225 modelos femeninos y 225 masculinos) antropométricamente correctos que han sido generados mediante el programa Poser (ver más abajo) para diferentes contexturas físicas y edades. Además, se cuenta con 149 modelos reales obtenidos durante una colecta de datos[1], realizado con el aval del comité de ética del Hospital Zonal Andrés Isola, Área Programática Norte, Puerto Madryn. Pueden verse en la Figura 8.2 algunos ejemplos de estos datos.

8.2.1. Origen de datos 3D

Malla 3D Por un lado, las mallas de cuerpos reales fueron tomadas con el sensor Structure[2] este sensor 3D cuenta con tecnología similar a la encontrada en Microsoft Kinect Sensor. Posee un rango operativo de 0.40 mts a 3.5 mts, precisión de dato de profundidad de 0.5mm en rango cercano, y bajo ruido entre capturas. El entorno de desarrollo de este sensor es multi-plataforma (iOS, Windows, Android y Linux). Para iOS, específicamente se provee un SDK con APIs de alto nivel para desarrollo de aplicaciones 3D. Para Windows y Linux se cuenta con OpenNI2[3]. Por otro lado, los cuerpos sintéticos fueron realizados con el programa de software *Poser*. Poser es un

[1] http://cites-gss.com/startups/2016/3dlabs/

[2] https://structure.io/

[3] https://structure.io/openni

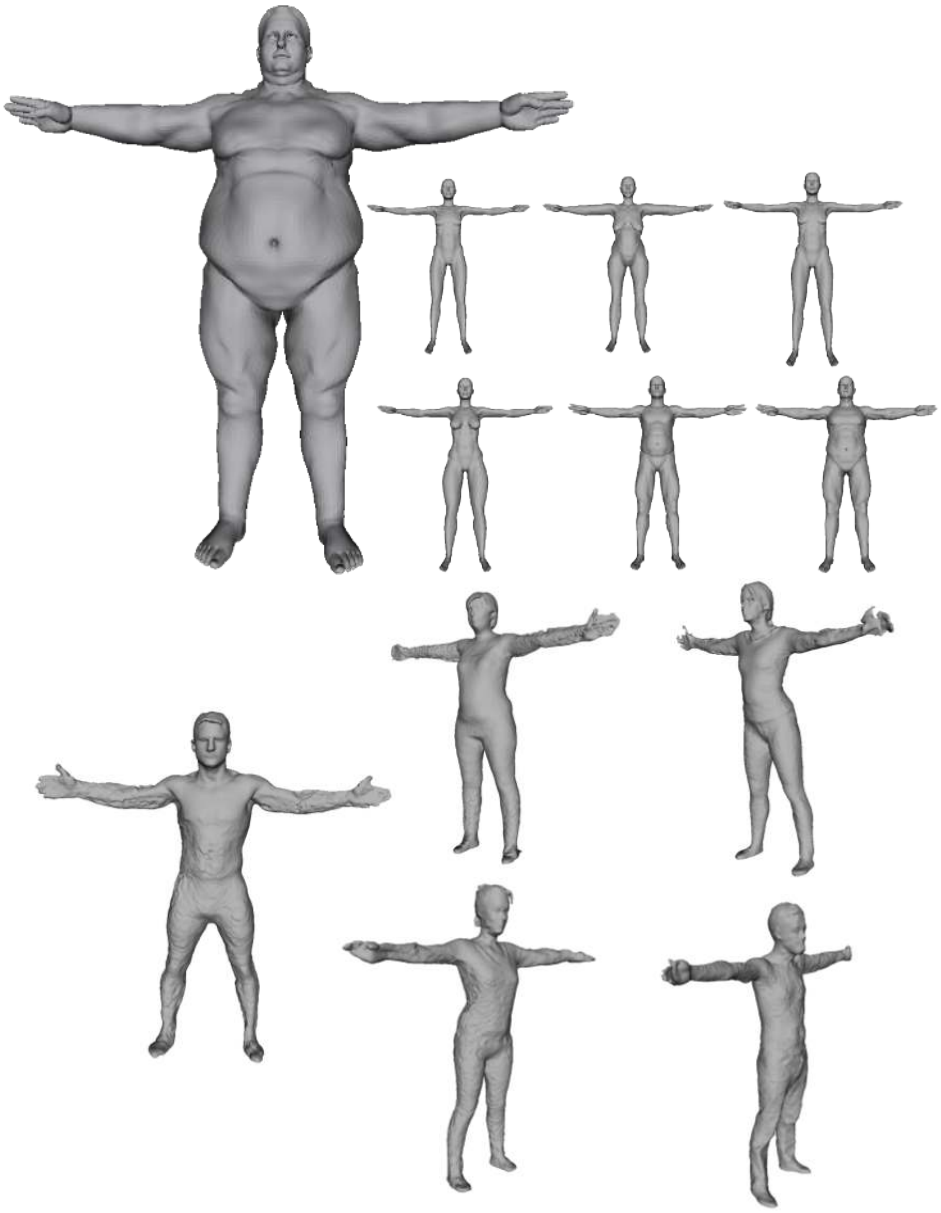


Figura 8.2: Ejemplos de modelos sintéticos (arriba) y modelos reales (abajo) utilizados. Colecta CITES.

software de rendering para posición, renderizado y animación de cuerpos humanos. De forma nativa usa OBJ formatos.

Landmarks 3D Las coordenadas de los landmarks fueron obtenidas por una profesional del área de antropología física utilizando una herramienta de anotación manual 3D llamada *Landmark*[4].

[4] http://www.idav.ucdavis.edu/research/EvoMorph

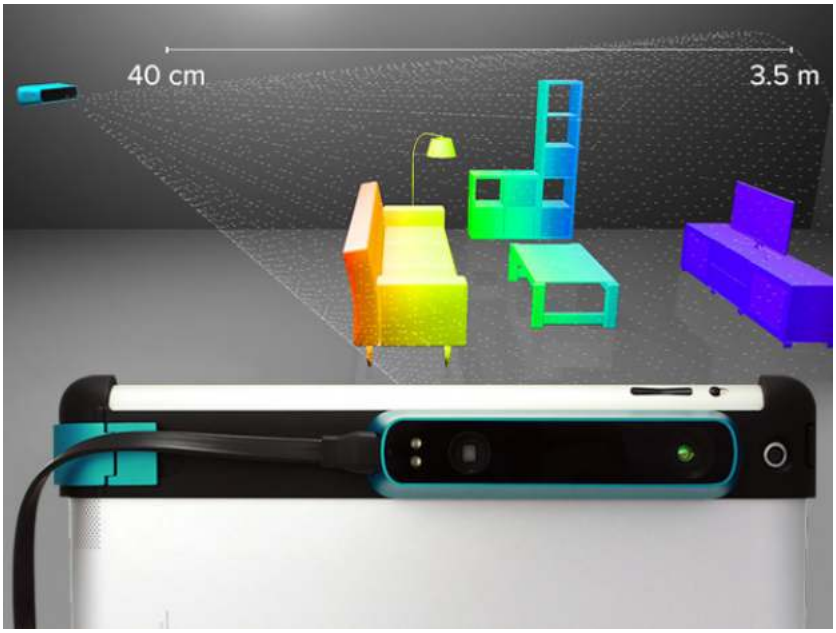


Figura 8.3: Imagen ilustrativa de Sensor Structure.

8.3. Pipeline Desarrollado

La idea principal consiste en diseñar y entrenar una red con un conjunto de ejemplos de nubes de puntos no ordenadas, asociadas con coordenadas en tres dimensiones (x, y, z). Se entrenó una red con un conjunto de 600 nubes no ordenadas, asociadas con 15 landmarks y semilandmarks ubicados manualmente. En esta sección detallamos todos los pasos del proceso, y por cada paso describimos los formalismos utilizados. En la Figura 8.4 se puede observar una visión general de los pasos realizados.

Obtención de vértices En este experimento sólo se trabajo con nubes de puntos no ordenadas, el procesamiento de las mismas fue hecho con la biblioteca PCL[5].

PCA El Análisis de Componentes Principales sobre una matriz de datos nos permite extraer los patrones dominantes de variación. Este método es utilizado para descomponer conjuntos de datos multivariados (con alta dimensionalidad) en un conjunto de componentes ortogonales que explican una proporción determinada de la varianza total. Específicamente, se utilizó un PCA aleatorio Halko et al. (2011), dada la gran cantidad de datos con la que se contaba, si se tiene en cuenta que cada escaneo produce 100187 vértices. Sea $\lambda_1, \lambda_2, \cdots \lambda_n$ (ordenados en valor decreciente)

[5]http://www.pointclouds.org/

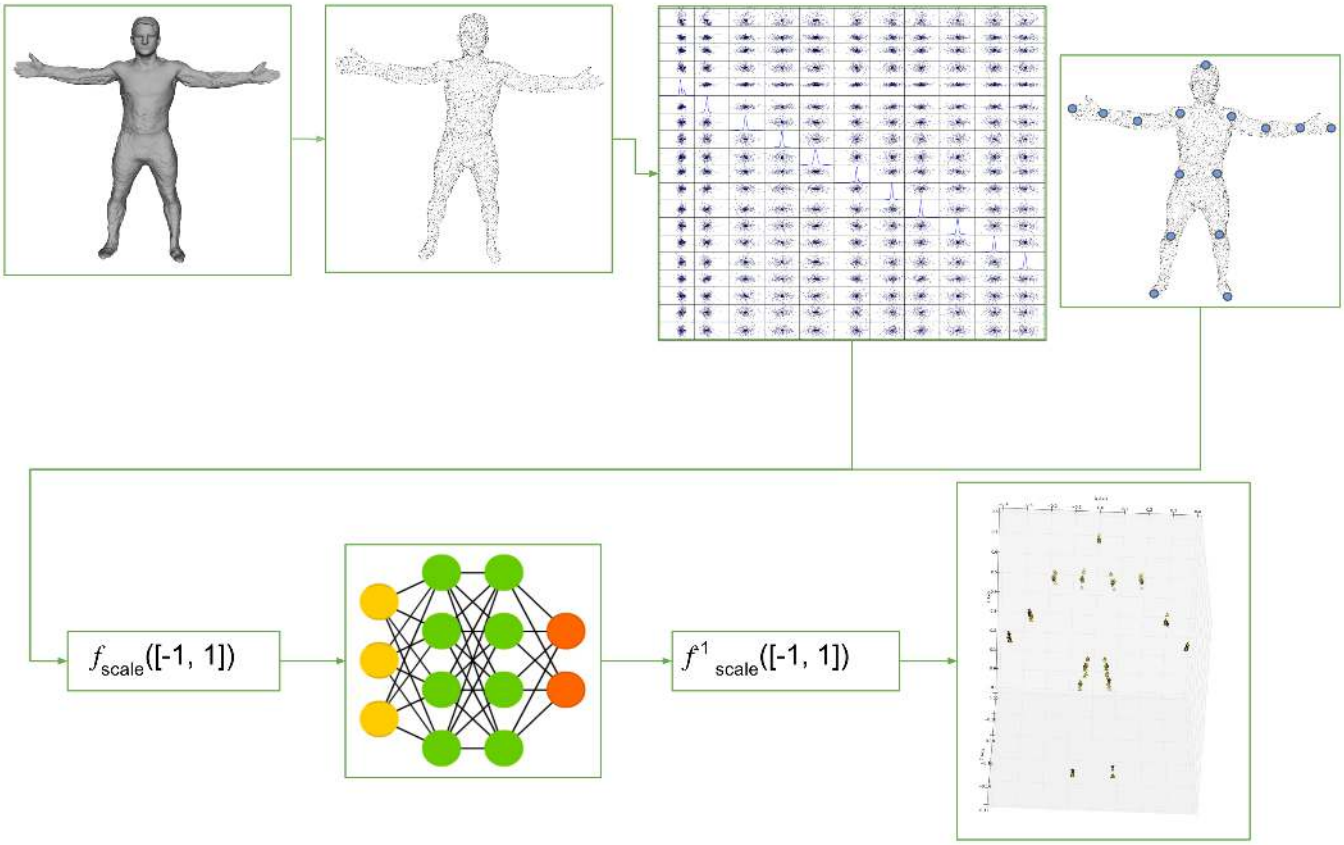


Figura 8.4: Estructura descriptiva del pipeline de landmarking corporal 3D.

los autovalores de $\sum$, siendo λ_j es el autovalor correspondiente al autovector u_j. Si tomamos k PCs, el porcentaje de variación obtenido esta definido como:

$$\frac{\sum_{j=1}^{k} \lambda_{j=1}}{\sum_{j}^{n} \lambda_j} \tag{8.1}$$

En la Figura 8.5 se pueden observar los 10 PCs retenidos según la Ecuación 8.1, los cuales explican el 0.852 de la variabilidad de las reconstrucciones. La decisión de la elección de la cantidad de PCs se basó en la evaluación de la red entrenada con diferentes cantidades de PCs. En las Figuras 8.6 y 8.7 se pueden observar las variaciones de r^2 y $RMSE$ en función de la cantidad de PCs.

Escalado Se utilizó un estimador que escala y traslada cada característica de forma individual para que se encuentre en un determinado rango (en nuestro caso $[-1, 1]$). Esto se aplicó tanto a los 10 PCs como a las 15 coordenadas de landmarks del tipo (x, y, z).

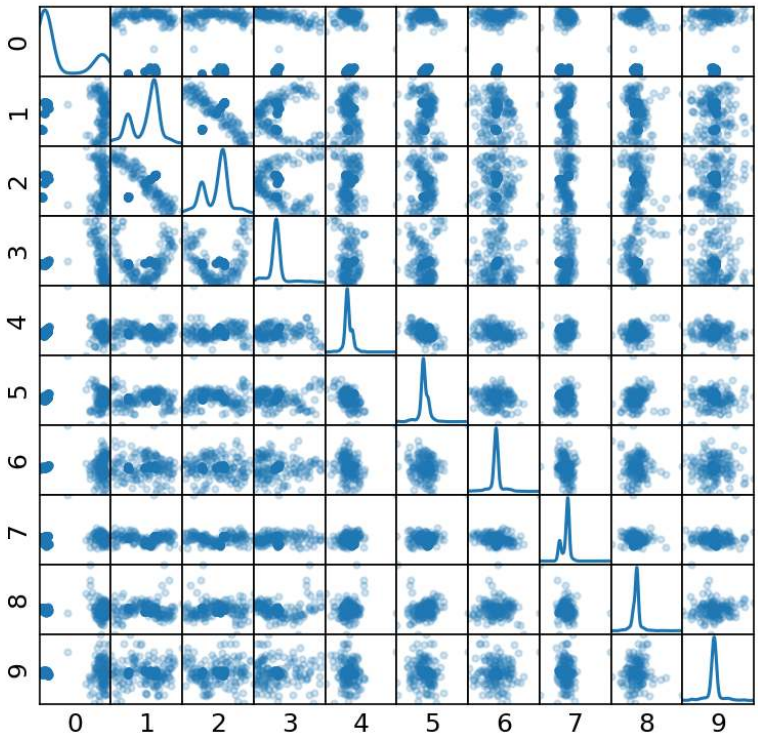


Figura 8.5: 10 PCs que explican el 85 % de la variabilidad de la muestra mixta (reales y sintéticos). En las primeros PCs se diferencia claramente los sintéticos de los reales.

Red Neuronal Dado que se contaba con un conjunto de datos reducido (10 PCs por individuo, con 480 individuos para el entrenamiento) se realizaron pruebas con estructuras de redes clásicas del tipo *perceptron*, con la salvedad de que se usó *Dropout* como técnica de regularización (Explicada en detalle en la Sección 4.3.4). La arquitectura contiene dos capas lineales completamente conectadas y una capa *dropout* intermedia. La capa de salida cuenta con 45 unidades (15 coordeandas (x, y, z)) una por cada posición predicha de landmarks y semi-landmarks. El entrenamiento se iniciaba con un $\eta = 0{,}6$ y decae de forma lineal a lo largo de todos los *epochs* hasta un $\eta = 0{,}001$.

Transformación a coordenadas reales Se aplica la inversa del estimador utilizado en el escalado para la entrada de coordenadas, las cuales son devueltas a la escala real

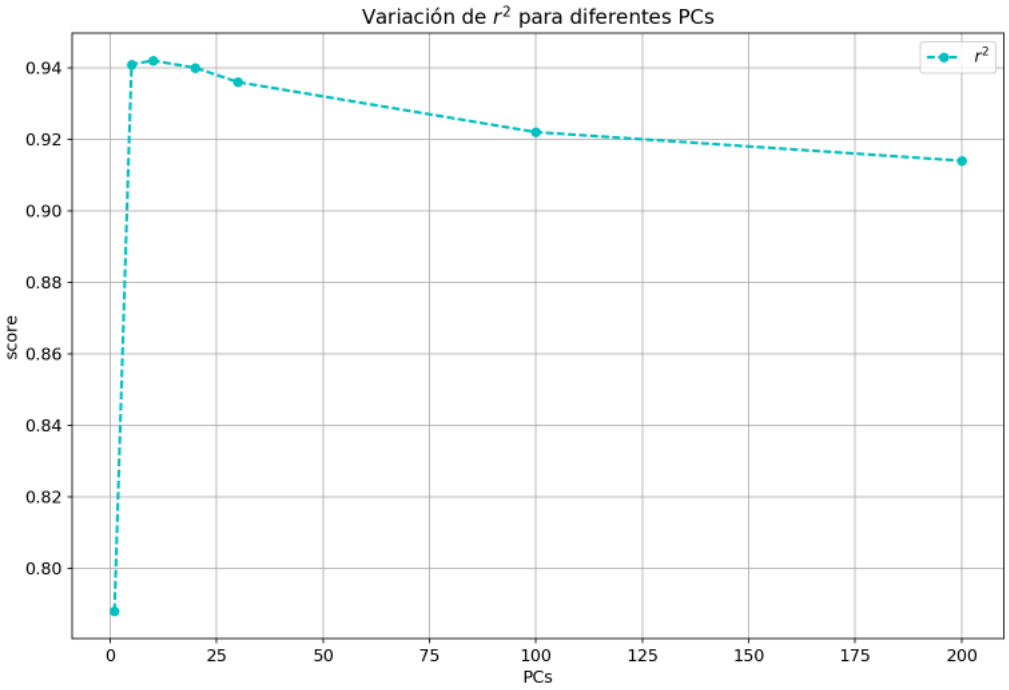


Figura 8.6: Variación de r^2 en función de la cantidad de PCs como entrada de la red.

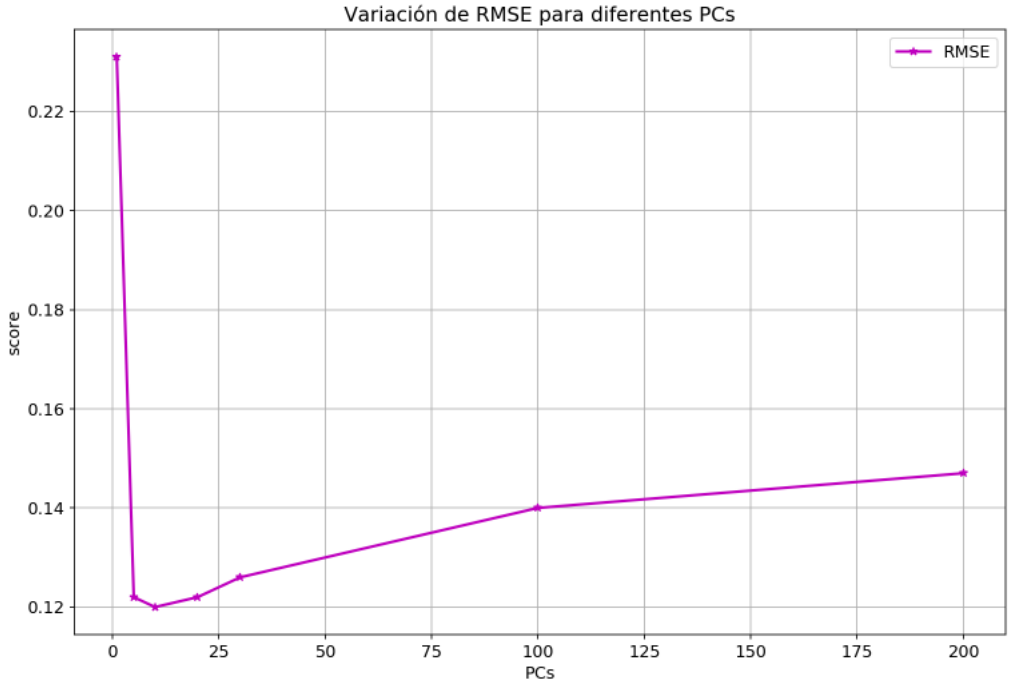


Figura 8.7: Variación de *RMSE* en función de la cantidad de PCs como entrada de la red.

asociada a la nube de puntos.

8.4. Resultados sobre landmarking automático 3D

Como observamos en capítulos anteriores la ubicación de landmarks en forma automática puede ser pensada como un problema de regresión. Al usar este enfoque aplicamos métricas para evaluar el desempeño de las diferentes redes entrenadas contra el landmarking manual (ground truth), en particular se trabajó con r^2, *root mean square error* (RMSE) y correlación de Pearson (detalladas en la Sección 5.3). El desempeño del land-

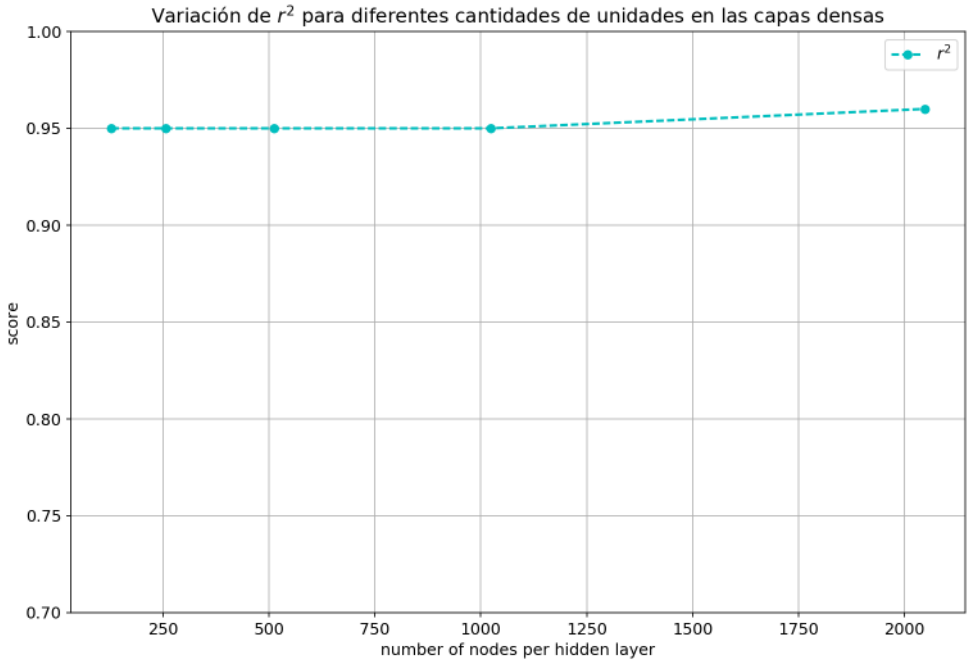


Figura 8.8: Variación de r^2 en función de la cantidad de nodos en las capas ocultas de la red.

marking de las redes implementadas se puede ver en la Tabla 8.2, en este caso se decidió utilizar la misma arquitectura en la comparación de las redes Perceptron con 128 nodos por capa densa, ya que al incrementar la cantidad de nodos en la red surgen varios factores negativos, si bien mejora levemente el r^2 (Figura 8.8), las curvas de pérdida muestran señales de sobre-ajuste a medida que se aumenta la cantidad de nodos (Figura 8.9) y el tiempo de entrenamiento crece a medida de que la arquitectura de la red se vuelve más compleja (Figura 8.10). En la Tabla 8.2 se analiza su entrenamiento variando el parámetro η. Para una versión gráfica de estos resultados ver Figura 8.11. Se puede observar que si bien la red *net5* tiene mejor r^2, se advirtió una diferencia significativa entre las curvas de validación y entrenamiento de la red *net5*, lo que nos indica un posible *overfitting* por lo

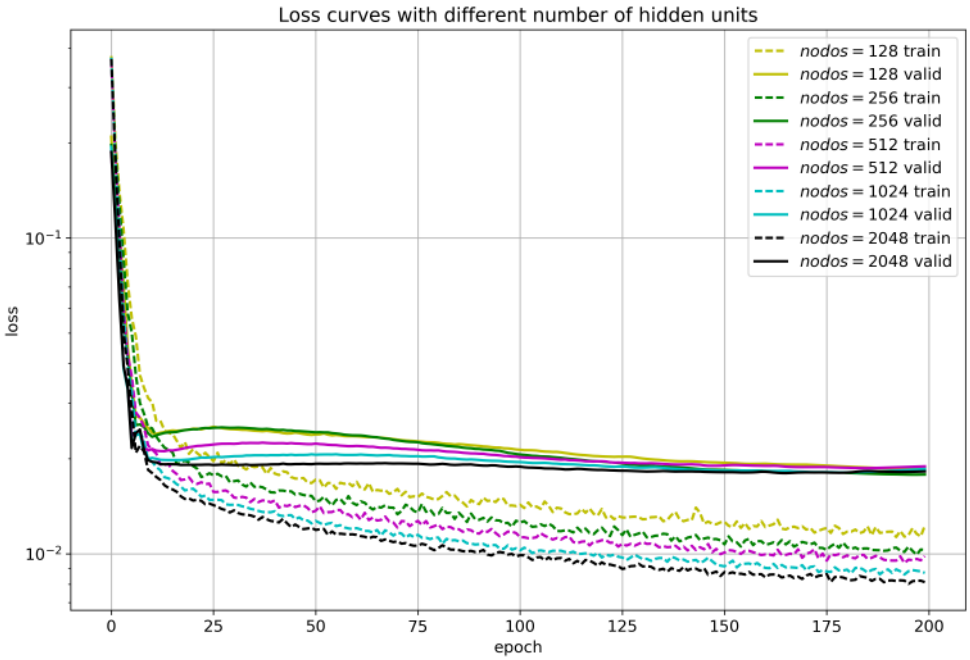


Figura 8.9: Curvas del pérdida de varias redes con diferentes cantidades de nodos en las capas ocultas. En las líneas punteadas se observa la curva de entrenamiento y la línea solida la validación.

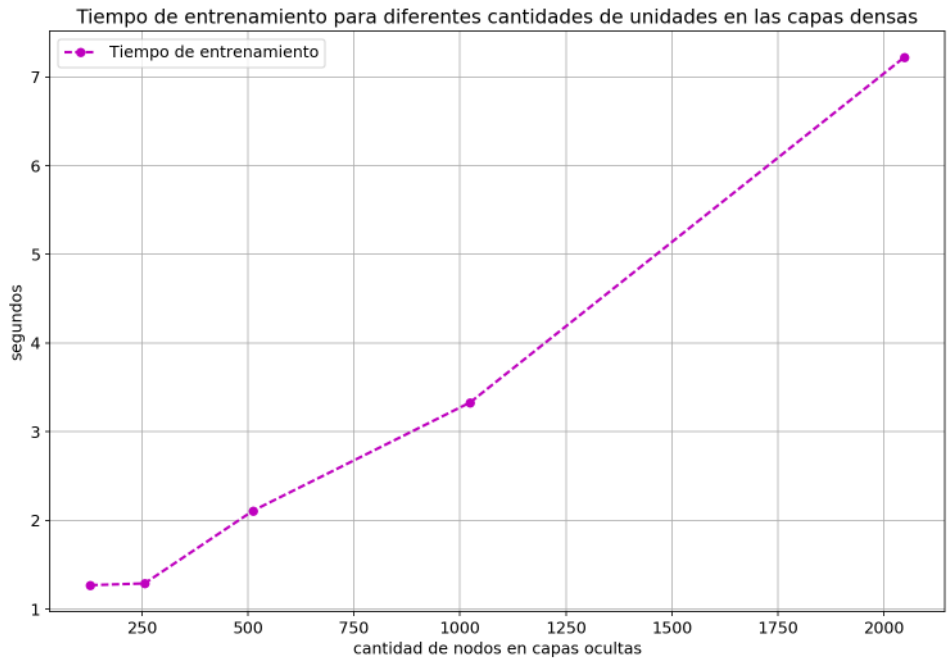


Figura 8.10: Curva de tiempo de entrenamiento de varias redes con diferentes cantidades de nodos en las capas ocultas.

Tabla 8.2: Comparación entre redes variando el parámetro η. Para una versión gráfica de estos resultados ver Figura 8.11.

# de red	RMSE	MAE	r^2	η
net1	0.12146	0.0737	0.94177	0.06
net2	0.11691	0.0697	0.94605	0.08
net3	0.11351	0.0664	0.94914	0.1
net4	**0.10449**	**0.0574**	**0.95691**	**0.6**
net5	0.10207	0.0541	0.95887	0.8

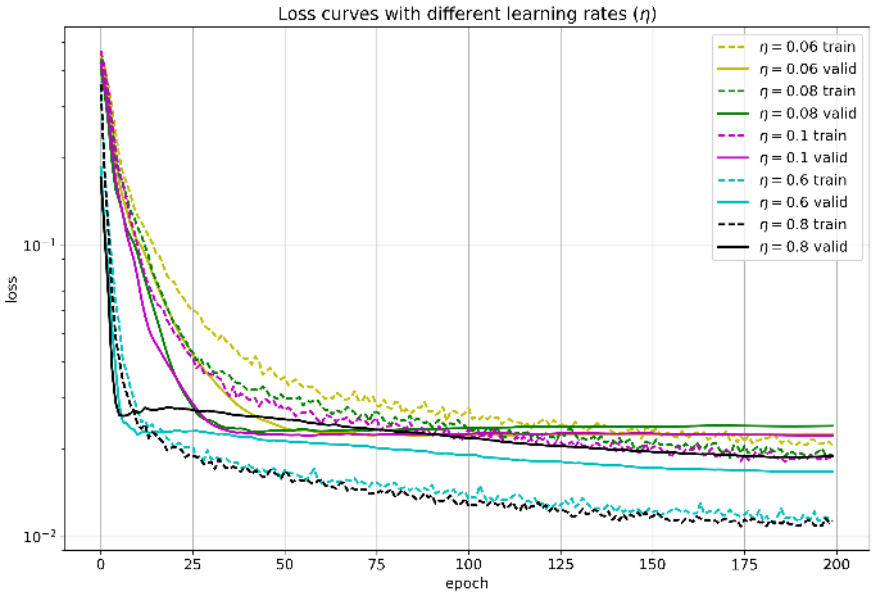


Figura 8.11: Curvas del pérdida de varias redes con diferente tasa de aprendizaje (η) pero misma estructura. En las líneas punteadas se observa la curva de entrenamiento y la línea sólida la validación.

que se decidió optar por la red *net4*. La red fue implementada en una PC con hardware convencional (single core Intel i7-5500 2.40 GHz). El landmarking automático de una nube de puntos (representada por sus primeros 10 PCs) tardó 198μs en promedio, por otro lado, el landmarking en modo batch de 120 individuos demoró 237μs. En el repositorio https://github.com/celiacintas/tests_landmarks, se encuentra el código de las pruebas realizadas sobre ipython notebooks. Si bien estos son primeras aproximaciones

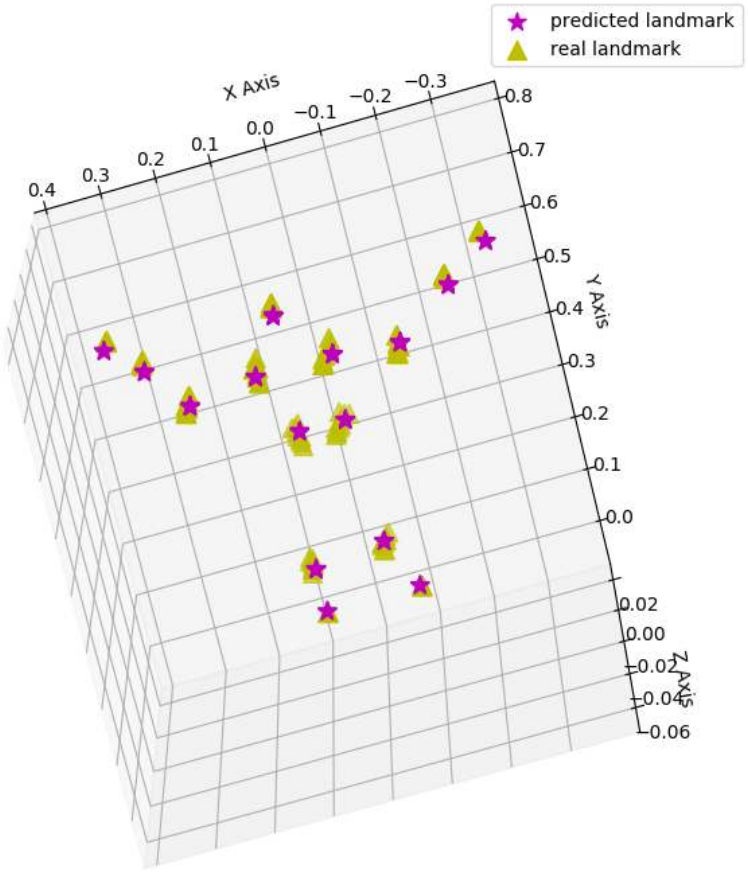


Figura 8.12: Algunos landmarks conformando el esqueleto, con su posición real y la predicha.

al problema de landmarking 3D, a partir de una mayor cantidad de datos migraremos a arquitecturas jerárquicas, para eliminar pasos de procesamiento. Actualmente nos encontramos expandiendo los resultados de esta Sección con experimentos sobre Deep learning sobre point clouds Qi et al. (2016).

8.5. Aplicaciones del escaneo corporal 3D

Dado que la toma de datos puede realizarse con dispositivos móviles de uso popular, y luego procesarlos utilizando técnicas de fotogrametría y *point clouds*, se pensó en el desarrollo de un servicio en la nube (SaaS) para realizar escaneo 3D y obtención de variables antropométricas, que permita cuantificar la forma corporal con alta exactitud. Los datos se gestionan desde la nube en forma segura y encriptada, pudiendo evaluar seguimientos, computar tendencias, elaborar reportes, impresión 3D, entre otras. Esta

Figura 8.13: Modelo obtenido con la aplicación de 3DLab.

tecnología tiene distintas aplicaciones en mercados globales de gran impacto:

- Bodybuilding y deportología, realización de evaluaciones físico-mecánicas, seguimiento de la evolución de rutinas de musculación.
- Outfit e-commerce, construcción automática de modelos 3D personalizados (avatar) para venta online de indumentaria.
- Medicina, seguimiento y evaluación de condiciones asociadas al sobrepeso, monitoreo y estimación de condiciones médicas (displasias, malformaciones, dermatología, prótesis, etc.)

En ese sentido, con un subsidio de la aceleradora de negocios CITES (del grupo SANCOR Seguros) se desarrolló un prototipo completo (minimum viable product) denominado 3DLabs (ver Figura 8.13, el cual, junto con la Gerencia de Vinculación del CONICET y se presentó a la convocatoria EMPRETECNO EBT de la ANPCyT.

Parte IV

Conclusiones

Capítulo 9

Conclusiones Generales

9.1. Conclusiones

La ubicación de landmarks sobre estructuras faciales, es un paso intermedio muy importante para muchas operaciones subsecuentes de análisis antropométrico, desde aspectos biométricos (Lanitis et al., 1995, Wiskott et al., 1997, Campadelli et al., 2003), la animación facial (Kang Liu et al., 2008), la interacción hombre-máquina, la determinación del punto de mirada, la comprensión de expresiones faciales (Tian et al., 2001, Pantic and Rothkrantz, 2000), la reconstrucción 3D de rostros (Ying et al., 2006), los videojuegos, y estudios de antropología física Paschetta et al. (2016), Quinto-Sánchez et al. (2015b), Quinto-Sánchez et al. (2017). Pese a ello no hay aún un uso masivo de las importantes variables cuantitativas antropométricas, principalmente por el límite que impone realizar tediosas mediciones manuales, un proceso lento y que origina grandes errores (Segev et al., 2010, Kamoen et al., 2001).

Por ello se propuso investigar técnicas avanzadas en procesamiento de imágenes para recolectar datos morfométricos de una manera abarcativa y masiva. Este método está basado en la Morfometría Geométrica y utiliza Redes Convolucionadas (CNN) para la realización del landmarking. Luego de entrenar la red con landmarks colocados de forma manual junto con sus imágenes, el algoritmo es capaz de ubicar la posición de landmarks y semi-landmarks de forma adecuada. Se emplearon ERT para probar la viabilidad de utilizar landmarks para tareas de reconocimiento e identificación. Se mostró que el método es lo suficientemente flexible como para poder ser aplicado en varios contextos diferentes.

Toda la implementación fue desarrollada con herramientas libres y abiertas y el código se encuentra disponible en https://github.com/celiacintas/deeplandmarking y https://github.com/celiacintas/tests_landmarks, por lo que el modelo preentrenado y el código pueden ser descargados y utilizados por la comunidad. Los fundamentos metodológicos de esta propuesta la hacen más robusta y flexible que otras aproximaciones *ad-hoc* existentes en la literatura, específicamente frente a situaciones típicas como por ejemplo homografías, remuestreo, o transformaciones de luminancia.

Nuestra propuesta es lo suficientemente general como para ser adaptada a otras características antropológicas físicas. Por dicha razón, varias líneas de investigación se abrieron gracias a los resultados aquí expuestos, incluyendo estudios de rasgos de interés biomédico y forense, biometría, análisis de percepción, por mencionar alguno de ellos. En cuanto a esto último, se han descrito una serie de trastornos que afectan al desarrollo del pabellón auditivo humano, que se producen aisladamente o como parte de síndromes complejos con múltiples órganos afectados (Faris, 2011, Cox et al., 2014). Por ejemplo, hemos reportado recientemente asociaciones genómicas de siete regiones genéticas, incluyendo el gen Edar, con fenotipos categóricos macroscópicos en el oído externo Adhikari et al. (2015). Por lo tanto, se necesitan mejoras adicionales en la captura de un fenotipo tan complejo para complementar la comprensión de su base genética y no genética.

Como resultado de este trabajo doctoral se realizaron varias publicaciones, donde se consignan los avances realizados en conjunción con esta línea de trabajo, incluyendo la presentación de un landmarking facial de vista frontal (Cintas et al., 2014a,b, 2013b), el análisis sobre la asimetría fluctuante asociado a ancestría genética (Quinto-Sánchez et al., 2015b), un estudio sobre las asimetrías fluctuantes a lo largo de la escala socioeconómica (Quinto-Sánchez et al., 2017), el desarrollo de un sistema de landmarking automático del pabellón auditivo (Cintas et al., 2016a), el cual explora las posibilidades de utilizar vectores de características como identificación y una cooperación a futuro para un trabajo sobre asociación de genoma completo identificando siete zonas del genoma altamente relacionadas con la variación del pabellón auditivo en humanos y ratones (Adhikari et al., 2015). Además se contó con una serie de resultados relacionados, logrados asimismo durante el período de la realización del doctorado:

- Estudios comparativos sobre la localización automática de regiones de interés utili-

zando redes neuronales, y su ventajosa comparación con el algoritmo más difundido (Viola-Jones).

- Un método para reconocimiento biométrico a partir de una configuración de landmarks emplazados sobre el pabellón auditivo y ERT (Cintas et al., 2016a).
- Análisis comparativo sobre la importancia relativa de los landmarks a la hora del ensamblado del vector de características (Cintas et al., 2016a).
- Evaluación de configuraciones de landmarks sobre vista lateral y su ulterior impacto como vectores de características.
- Un método para clasificación de género a partir de imágenes de vista lateral utilizando landmarking automático.

Aunque el propósito principal del trabajo doctoral era mostrar el potencial combinado de Morfometría Geométrica junto con Deep Learning, realizamos investigaciones adicionales, evaluando la capacidad del *workflow* desarrollado para otros fines en reconocimiento e identificación de personas. Para ello se emplearon árboles de decisión, los cuales son especialmente útiles para determinar la importancia de los datos, cómo son clasificados, y paras obtener valores de pesos relativos o medir la importancia de cada dimensión en el espacio de características Wehenkel et al. (2006). Gracias a esta característica fue posible analizar la contribución relativa de cada landmark en el proceso de reconocimiento.

Uno de los resultados más importantes obtenidos está relacionado con el landmarking del pabellón auditivo humano. Se observó que las coordenadas de landmarks más importantes a la hora de selección corresponden a la parte anatómica interna de la oreja, lo cual sugiere que esta estructura es más informativa que la estructura externa (contorno de la oreja) como un posible valor discriminante. Estas observaciones no habían sido realizadas con anterioridad en la comunidad científica especializada en la biometría, por lo cual existe una gran cantidad de oportunidades científicas relacionadas a estos resultados.

Otro resultado está aplicado a la determinación de género en personas a partir de imágenes faciales, el cual es un paso importante de pre-procesamiento para varias aplicaciones, como por ejemplo en interacción humano-computadora, sistemas de vigilancia, marketing, etc. En los trabajos previos en la literatura se realizan evaluaciones sobre conjuntos de

datos muy pequeños, entre 30 a 44 personas, contando con 60 a 290 imágenes de todos los individuos. En nuestros trabajos contamos con varios miles de personas, por lo cual la significatividad de nuestros resultados es considerablemente mayor. Además, el estado del arte es utilizar una cantidad limitada de landmarks (entre 5 a 20 puntos fiduciarios), ubicados mayoritariamente en áreas donde las expresión facial puede alterar el resultado. En nuestro caso se empleó una configuración más densa de landmarks, lo cual produce una representación más robusta del contorno del rostro, la cual reduce significativamente los efectos de las expresiones faciales.

9.2. Trabajos en curso y futuros

En estos momentos nos encontramos definiendo varias líneas de investigación las cuales surgen naturalmente de los resultados mostrados en la presente tesis. Por un lado, el meta-modelado de CNN, el cual los permitirá el desarrollo generalizado de fenotipado automático de diferentes estructuras biológicas de manera transparente. Generando una biblioteca de landmarking de alto nivel que podrá ser utilizada con sólo el conjunto de entrenamiento de imágenes y sus landmarks asociados, ver la Sección 9.2.1. Por otro lado, el diseño de CNN para diferentes aplicaciones bioantropológicas como se detalla en la Sección 9.2.2.

9.2.1. Meta-modelado de Arquitecturas de CNNs

Finalmente, como resultado del diseño de diversas arquitecturas de CNN se observó que ésto requiere experiencia y diseño humano, pese a ser una tarea exploratoria. Cada arquitectura es diseñada bajo cuidadosa experimentación y modificada o heredada por modelos previos. Surge entonces como objetivo primordial de trabajo el desarrollo de algoritmos de meta-modelado utilizando analíticos visuales para generar estructuras de CNN con un elevado desempeño para clasificación Baker et al. (2016), Shahriari et al. (2016), Pinto et al. (2009). Otro trabajo específico en desarrollo es el diseño de algoritmos para meta-modelado de estructuras de CNN para fenotipado automático. Asimismo se están estudiando nuevas técnicas de *data augmentation* específicas para los conjuntos de datos de fenotipos actuales, para aumentar la capacidad de generalización.

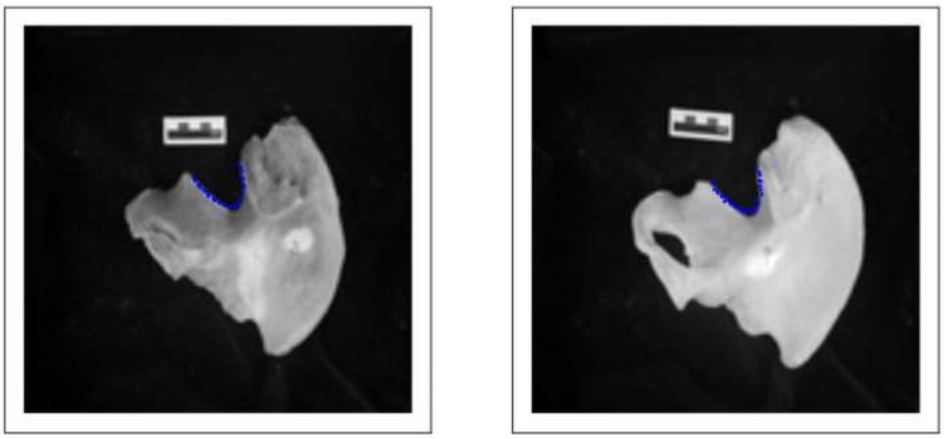


Figura 9.1: Semilandmarks para calcular escotadura ciática.

9.2.2. Aplicaciones bioantropológicas

Por el lado de las aplicaciones bioantropológicas, actualmente se está desarrollando una aplicación de escaneo corporal 3D, la cual es uno de los pasos previos a la obtención de medidas antropométricas es el landmarking automático como se definió en la Sección 8.4. Otra aplicación importante de nuestro framework de análisis esta siendo realizada para la estimación de género en la reconstrucción del perfil biológico de un esqueleto no identificado (arqueológico o contemporáneo) y en forma consecuente para la identificación positiva de restos óseos recuperados de una escena forense. Se procesaron 130 imágenes de huesos iliacos (45.4 % femeninos y 54.6 % masculinos) pertenecientes a la colección del Departamento de Anatomía, Facultad de Medicina de la Universidad de México (UNAM). Se utilizaron semilandmarks sobre la escotadura ciática como vector de características para definir el sexo Gómez-Valdés et al. (2012). Actualmente estamos diseñando nuevas arquitecturas de redes de análisis, así como métodos de *data augmentation* para trabajar sobre estos datos con mayor significatividad y de forma automática. Algunos resultados preliminares e ilustrativos se pueden ver en la Figura 9.1. También sobre datos óseos se está comenzando a desarrollar un modelo para el landmarking automático sobre cráneos. Para ello se cuenta con dos configuraciones de landmarks distintas, por un lado landmarks situados sobre la cara y sobre el contorno del cráneo. Estas configuraciones fueron utilizadas en el trabajo de De Azevedo et al. (2011), en el cual se realizó un estudio sobre posibles escenarios en las fases iniciales de las migraciones de humanos desde Asia al Nuevo Mundo. Se trabajó con datos morfométricos para analizar asociaciones entre la

Figura 9.2: Semilandmarks para calcular contorno del craneo.

matriz de distancias craneométricas y diferentes matrices geográficas, reflejando distintos escenarios para el poblamiento del Nuevo Mundo. Algunas imágenes preliminares de landmarking automático se pueden observar en la Figura 9.2.

Anexos

Anexo A

Implementación de Redes Neuronales con Theano y Lasagne

A.1. Theano

Theano (Theano Development Team (2016)) es una biblioteca de Python que permite definir, optimizar y evaluar expresiones matemáticas que tienen como entrada arreglos multi-dimensionales de forma eficiente.

Theano puede calcular automáticamente la diferenciación simbólica de expresiones complejas, ignorar las variables que no son necesarias para calcular la salida final, reutilizar resultados parciales para evitar cálculos redundantes, aplicar simplificaciones matemáticas, calcular las operaciones en su lugar cuando sea posible para minimizar el uso de memoria y aplicar optimización numérica de estabilidad para superar o minimizar el error debido a las aproximaciones de hardware. Para lograrlo, las expresiones matemáticas definidas por el usuario se almacenan como un grafo de variables y operaciones, que se optimiza en tiempo de compilación.

La API de Theano simula NumPy Oliphant (2007), van der Walt et al. (2011), una biblioteca de Python extensamente utilizada que proporciona un tipo de datos de matriz n-dimensional y muchas funciones para indexar, reestructurar y realizar cálculos elementales en arreglos. Esto permite a los usuarios de Python cambiar rápidamente a Theano utilizando una sintaxis y un conjunto de instrucciones conocidas, ampliadas con funciones avanzadas, como computación automática de gradientes, mejoras numéricas de estabi-

lidad y optimización, y generar un código de alto rendimiento tanto para la CPU como para la GPU (ademas de que es posible generar contextos para trabajar con múltiples placas gráficas de forma semi-transparente), sin requerir cambios en el código de usuario. Theano también ha sido diseñado para poder ser extendido de manera fácil y rápida a través de la definición de grafos de ejecución personalizados desarrollados en Python, C ++ o CUDA Theano Development Team (2016).

A.1.1. Aritmética de Convolución con Theano

La forma de la salida de una capa de convolución esta afectada por la forma de su entrada, el kernel, el relleno de ceros y su paso. La relación entre estos componentes no es tan fácil de vislumbrar. En contraste con las capas densas, las cuales su tamaño de salida es independiente del tamaño de entrada. La definición de la operación de Convolución se revisó de forma detallada en la Sección 4.4.2. Aquí veremos cuestiones prácticas para su implementación y comprender la entrada y salida de estas capas.

La colección de *kernels* que definen una convolución discreta de N dimensiones tienen la forma de (n, m, k_1, k_N), donde:

1. $n \equiv$ cantidad de mapas de característica de salida
2. $m \equiv$ cantidad de mapas de característica de entrada
3. $k_j \equiv$ tamaño del kernel sobre el eje j

Las siguientes propiedades infieren en el tamaño de la salida de una capa de convolución o_j sobre el eje j.

1. i_j, es el tamaño de entrada sobre el eje j.
2. k_j, es el tamaño del kernel sobre el eje j.
3. s_j, es la distancia entre dos posiciones consecutivas de un kernel (también llamado paso) sobre el eje j.
4. p_j, cantidad de ceros concatenados al comienzo y fin de cada eje (también llamado relleno de ceros), sobre el eje j.

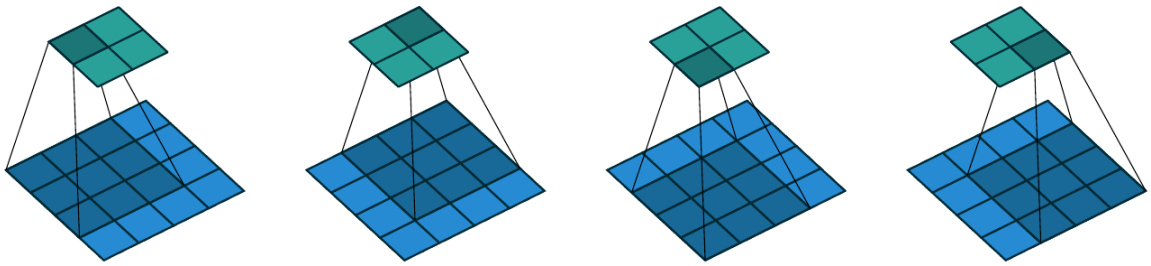


Figura A.1: Ejemplo de convolución con un kernel de 3×3 sobre una matriz de entrada de 5×5 usando un paso unitario de 1×1 y sin ceros agregados Dumoulin and Visin (2016).

Un ejemplo se puede observar en la Figura A.1.

Para simplificar los ejemplos tomaremos convoluciones en 2D ($N = 2$), las entradas ($i_1 = i_2 = i$), kernels ($k_1 = k_2 = k$), paso ($s_1 = s_2 = s$) y el relleno de ceros ($p_1 = p_2 = p$) son cuadradas. Ahora analicemos distintas relaciones entre estos valores:

Relación 1 Primero veamos el caso más simple, donde no contamos con relleno de ceros y el paso es unitario. Sea cualquier i y k donde $s = 1$ y $p = 0$,

$$o = (i - k) + 1 \tag{A.1}$$

Relación 2 Ahora, basándonos en la relación anterior, utilizamos el relleno de ceros. Sea i, k y p donde $s = 1$,

$$o = (i - k) + 2p + 1 \tag{A.2}$$

Relación 3 Cuando deseamos que el tamaño de entrada sea igual que el de salida. Tenemos que para cualquier i, con k impar ($k = 2n + 1$, donde $s = 1$ y $p = \frac{k}{2} = n$.

$$o = i + 2\left\lfloor \frac{k}{2} \right\rfloor - (k - 1) = i + 2n - 2n = i \tag{A.3}$$

Relación 4 Cuando se busca tener una salida más grande que la entrada dado cualquier i, k y $p = k - 1$ y $s = 1$,

$$o = i + 2(k - 1) - (k - 1) = i + (k - 1) \tag{A.4}$$

Relación 5 Hasta ahora todas las relaciones que analizamos cuentan con un paso unitario, ahora analicemos esto para pasos más grandes. Sea cualquier i, k, $s > 1$ y $p = 0$

$$o = \left\lfloor \frac{i - k}{s} \right\rfloor + 1. \tag{A.5}$$

Relación 6 Por último vemos el caso más general, donde se cuenta ceros agregados en los bordes y pasos no unitarios.

$$o = \left\lfloor \frac{i + 2p - k}{s} \right\rfloor + 1 \tag{A.6}$$

Ahora veamos un ejemplo de una capa de convolución con Theano. La entrada cuenta con 1 mapa de características (Imagen de 128×128 de un único canal). A esta entrada aplicamos un filtro constituido por dos kernels de 4×4.

```
from theano.tensor.nnet import conv2d, sigmoid
from theano import shared
import numpy as np

rng = np.random.RandomState(23465)

# Creamos tensor de 4D el cual contendrá la imagen
input_l = T.tensor4(name='input')
# inicializamos con variables el filtro con los dos kernels de 4x4
w_shp = (2, 1, 4, 4)
w_bound = np.sqrt(1 * 4 * 4)
W = shared( np.asarray(
            rng.uniform(
                low=-1.0 / w_bound,
                high=1.0 / w_bound,
                size=w_shp),
            dtype=input_l.dtype), name ='W')
# definimos nuestro bias
b_shp = (2,)
b = shared(np.zeros(b_shp, dtype=input_l.dtype), name ='b')
```

```
# creamos la convolución y aplicamos una función de activación
conv_out = conv2d(input_1, W)
output = sigmoid(conv_out + b.dimshuffle('x', 0, 'x', 'x'))
# definimos la función en theano
f = function([input_1], output)
```

Ahora que tenemos todo definido, apliquemos las funciones creadas (ver resultado en la Figura A.2):

```
import cv2
import matplotlib.pyplot as plt


img = cv2.imread('falcon_#1.jpg', 0)
img = img / 255.
img = img.astype(np.float32)
# colocamos la imagen en forma de tensor 4D
img_ = img.reshape(1, 1, 128, 128)
filtered_img = f(img_)
plt.subplot(1, 3, 1)
plt.imshow(img, cmap='gray')
plt.subplot(1, 3, 2)
plt.imshow(filtered_img[0, 0, :, :], cmap='gray')
pylab.subplot(1, 3, 3)
plt.imshow(filtered_img[0, 1, :, :], cmap='gray')
plt.tight_layout()
plt.show()
```

Para más información sobre este apartado recomendamos la lectura de Dumoulin and Visin (2016), Theano Development Team (2016)

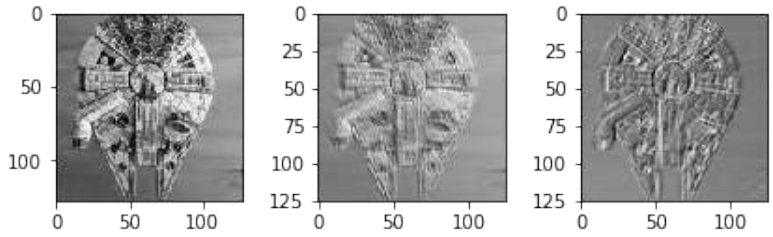


Figura A.2: Aplicación de filtro con dos kernels de 4×4 sobre entrada de 128×128 con Theano.

A.2. Lasagne

Lasagne (Dieleman et al. (2015b)) es una biblioteca de Python que permite construir y entrenar redes neuronales en Theano Theano Development Team (2016). Sus principales características son:

1. Soporta redes de feed-forward tales como Redes Neuronales Convolucionales CNN, redes recurrentes incluyendo Memoria de Corto Plazo (LSTM), y cualquier combinación de las mismas.

2. Permite arquitecturas de múltiples entradas y salidas múltiples, incluyendo clasificadores auxiliares.

3. Varios métodos de optimización implementados, entre ellos Nesterov momentum, RMSprop y ADAM.

4. Conjunto de funciones de costo clasicas implementadas. Además, cuenta con la posibilidad de definir funciones propias de manera simple sin necesidad de derivar los gradientes gracias al compilador de expresiones de Theano.

5. Soporte transparente de CPUs y GPUs gracias al compilador de expresiones de Theano.

Lasagne es una excelente biblioteca cuando lo que se busca es flexibilidad en el desarrollo de los modelos, en términos de definir funciones objetivo personalizadas, selección de las muestras a entrenar y generación artificial de datos en tiempo de entrenamiento. Lasagne esta desarrollada sobre Theano, habilitando la definición de redes de forma sencilla sin perder acceso a las variables de Theano.

El conjunto de datos y código mostrado en este Anexo se encuentra disponible en (https://github.com/celiacintas/star_wars_hackathon)

A.3. CNN con Lasagne

Primero veremos los tipos de capas disponibles en Lasagne para el armado de CNNs. Luego daremos un ejemplo sencillo de el ensamblado de una red, normalización de entradas, su entrenamiento y visualización de la red.

Dentro de los tipos de Capas que veremos en el ejemplo tenemos:

Capas Densas Todas las unidades ocultas están conectadas con todas las unidades de entrada.

```
(DenseLayer, {'num_units': 256, 'nonlinearity': rectify})
```

Capas de Convolución En las capas de convolución, las unidades están organizadas en *feature maps*, en las cuales cada unidad esta conectada a patches locales de los *feature maps* pertenecientes a la capa anterior mediante un conjunto de pesos, llamados *filter bank*. Todas las unidades dentro de un *feature map* comparten el mismo *filter bank*.

```
(Conv2DLayer, {'num_filters': 32, 'filter_size': 2, 'W': GlorotUniform()})
```

Max Pooling Para reducir la dimensionalidad de los *feature maps* una capa de pooling es ubicada entre las capas de convolución. Las capas de pooling eliminan los valores no máximos calculando una función de agregación, comunmente se utiliza el maximo o el promedio sobre pequeñas regiones de la entrada. El propósito general de las capas de pooling es reducir el costo computacional en las capas ulteriores, reduciendo el tamaño de los futuros mapas de características y otorgando una forma de invariancia traslacional.

```
(MaxPool2DLayer, {'pool_size': 2})
```

Dropout El termino de dropout se refiere al descarte de unidades y sus conexiones (ya sea en capas ocultas o no) en una red neuronal, este descarte es solo temporal. La

forma de elección de descarte es aleatoria, a cada unidad se le asocia un valor de probabilidad p independiente del resto entre $[0, 1]$.

```
(DropoutLayer, {'p': 0.5})
```

Ahora que tenemos la definición de todas las capas a utilizar veremos como se realiza el diseño de la red. Ademas de pasar el orden y tipo de capas que contará la red, se debe indicar que tipo de optimización se empleará, determinar la tasa de aprendizaje y su comportamiento a lo largo del entrenamiento y la función objetivo a minimizar.

```
def create_net(max_epochs=20):
    return NeuralNet(
        layers=[
                (InputLayer, {'shape': (None, 1, 128, 128)}),
                (Conv2DLayer, {'num_filters': 16, 'filter_size': 3,
                               'W': lasagne.init.GlorotUniform()}),
                (MaxPool2DLayer, {'pool_size': 2}),
                (DropoutLayer, {'p': 0.5}),
                (Conv2DLayer, {'num_filters': 16, 'filter_size': 3,
                               'W': lasagne.init.GlorotUniform()}),
                (MaxPool2DLayer, {'pool_size': 2}),
                (DenseLayer, {'num_units': 100, 'nonlinearity':rectify}),
                (DropoutLayer, {'p': 0.5}),
                (DenseLayer, {'num_units': 3, 'nonlinearity':softmax}),
        ],
        update=nesterov_momentum,
        update_learning_rate=theano.shared(np.float32(0.03)),
        update_momentum=theano.shared(np.float32(0.9)),
        regression=False,
        objective_loss_function=categorical_crossentropy,
        batch_iterator_train=BatchIterator(batch_size=512),
        on_epoch_finished=[
            AdjustVariable('update_learning_rate', start=0.03, stop=0.001),
```

```
            AdjustVariable('update_momentum', start=0.9, stop=0.9999)
        ],
        max_epochs=max_epochs,
        verbose=1)
```

A.3.1. Normalización de valores de entrada

Dados los motivos explicados en la Sección 4.3.5, es deseable normalizar los valores de entrada de nuestra red. Supongamos que los datos a consumir por nuestra CNN son imágenes en escala de grises, por lo que trabajaremos sobre un único canal y por simplicidad diremos que su dimensión es 128×128 (Figura A.3). Es deseable que los valores de las imágenes se encuentren en rango de $[0, 1]$ para una rápida convergencia y ademas deben estar en formato de tensores 4D y en float32 para su funcionamiento en Theano bajo GPUs. A continuación se muestra un breve ejemplo de código para lograr lo mencionado anteriormente.

```
import numpy as np
from sklearn.utils import shuffle

def load(images_dir="../data/all/"):
    """Load images and target class for Falcon, Lambda and K-wing ships."""
    df = images_to_file(images_dir)
    df['Image'] = df['Image'].apply(lambda im: np.fromstring(im, sep=' '))
    X = np.vstack(df['Image'].values) / 255.
    X = X.astype(np.float32)
    y = df['Model_enc'].values
    X, y = shuffle(X, y, random_state=42)
    y = y.astype(np.int32)

    return X, y
```

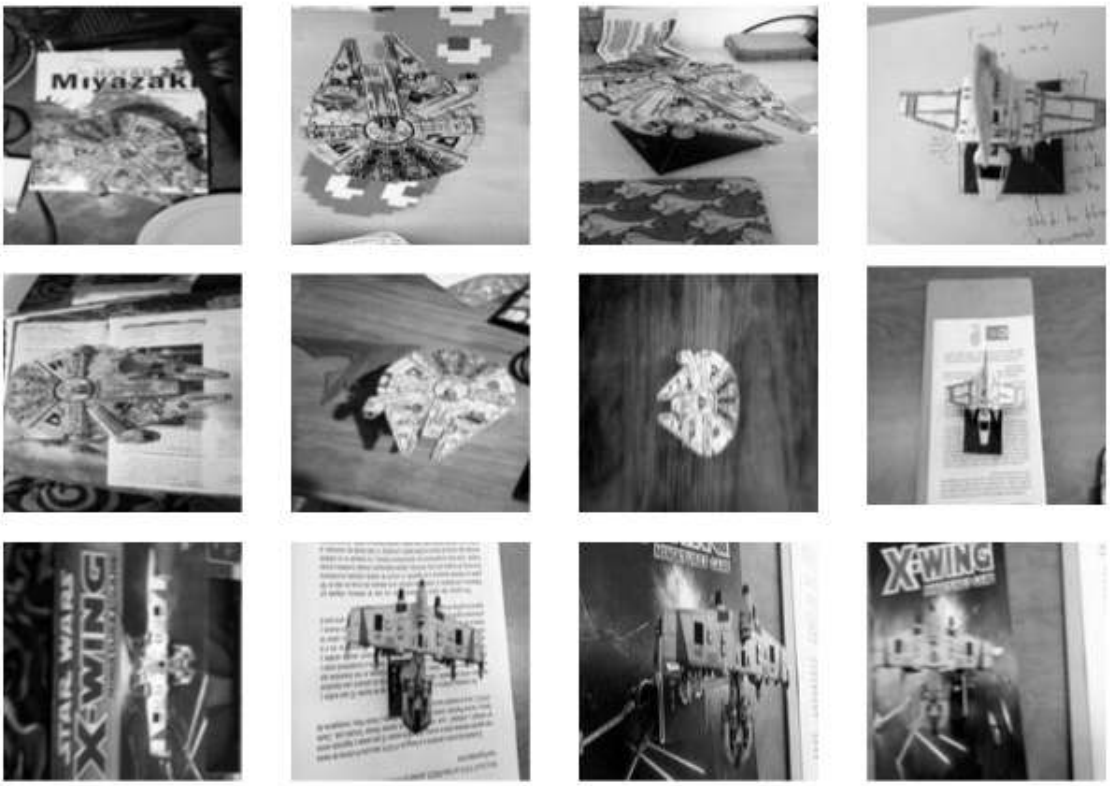


Figura A.3: Ejemplo de Imágenes que se utilizarán como dato de entrada de nuestra ConvNet.

A.3.2. Entrenamiento

Para el entrenamiento se cuentan con funciones como *fit* que permiten entrenar la red con un conjunto de datos de entrenamiento con sus etiquetas asociadas. Un ejemplo de entrenamiento de la red definida anteriormente con 20 *epochs*, se pueden ver en el siguiente código y su correspondiente salida o también de manera gráfica en la Figura A.4.

```
from sklearn.cross_validation import train_test_split
x, y = load()
x = x.reshape(-1, 1, 128, 128)
X_train, X_test, y_train, y_test = train_test_split(x, y,
                                                    test_size=0.3,
                                                    random_state=42)
net.fit(X_train, y_train)
```

```
# Neural Network with 1442883 learnable parameters

## Layer information

  #  name        size
---  ----------  ----------
  0  input0      1x128x128
  1  conv2d1     16x126x126
  2  maxpool2d2  16x63x63
  3  conv2d3     16x61x61
  4  maxpool2d4  16x30x30
  5  dense5      100
  6  dropout6    100
  7  dense7      3

  epoch    trn loss    val loss    trn/val    valid acc  dur
-------  ----------  ----------  ---------  -----------  -----
      1     1.08048     1.02025    1.05903      0.63590  3.25s
      2     0.94312     1.07926    0.87386      0.47179  3.23s
      3     0.78020     0.52369    1.48980      0.81026  3.23s
      4     0.48067     0.31679    1.51731      0.89231  3.23s
      5     0.32176     0.23017    1.39794      0.92650  3.25s
      6     0.20488     0.13029    1.57248      0.95385  3.23s
      7     0.15567     0.10442    1.49076      0.96410  3.23s
      8     0.13133     0.10713    1.22593      0.96581  3.23s
      9     0.08860     0.07329    1.20877      0.97607  3.23s
     10     0.09492     0.07117    1.33381      0.97607  3.23s
     11     0.07474     0.06678    1.11915      0.98291  3.23s
     12     0.06174     0.07607    0.81164      0.97778  3.23s
     13     0.05950     0.05635    1.05589      0.98120  3.23s
     14     0.05217     0.06147    0.84883      0.98120  3.23s
```

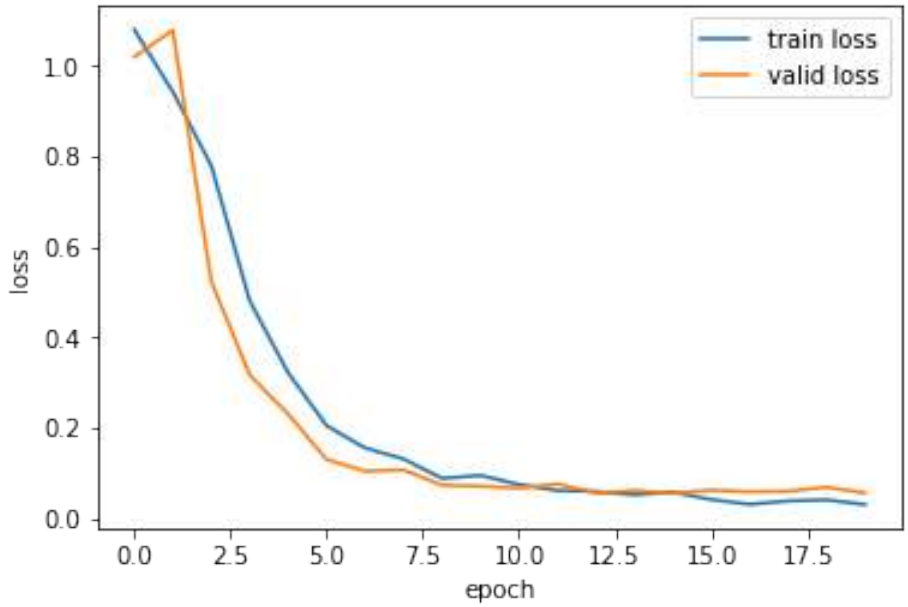


Figura A.4: Gráfico de la función de perdida. En azul la curva de entrenamiento y naranja la curva de validación.

15	0.05787	0.05515	1.04931	0.97949	3.23s
16	0.04165	0.06229	0.66868	0.98120	3.23s
17	0.03111	0.05882	0.52888	0.98120	3.23s
18	0.03856	0.06006	0.64200	0.97607	3.23s
19	0.04087	0.06890	0.59319	0.98291	3.23s
20	0.03075	0.05620	0.54719	0.98632	3.23s

A.3.3. Visualización de CNNs

Una visualización interesante para tener introspección sobre las decisiones que toma la red, es por ejemplo, ver que partes o píxeles son importantes a la hora de clasificar una imagen X Nouri (2014), Zeiler and Fergus (2013). Un ejemplo de la salida de esta visualización se puede observar en la Figura A.5.

```
h = visualize.plot_occlusion(net, X, [2], square_length=3)
```

Otra visualización importante es poder ver los kernels de una capa de convolución específica y su salida como puede observase en la Figura A.6.

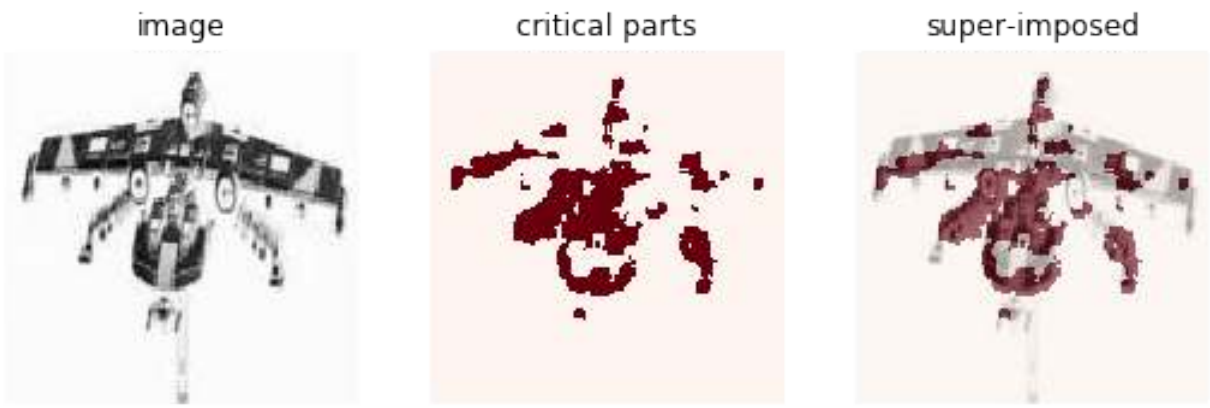


Figura A.5: Importancia de pixels a la hora de clasificar una imagen X de entrada.

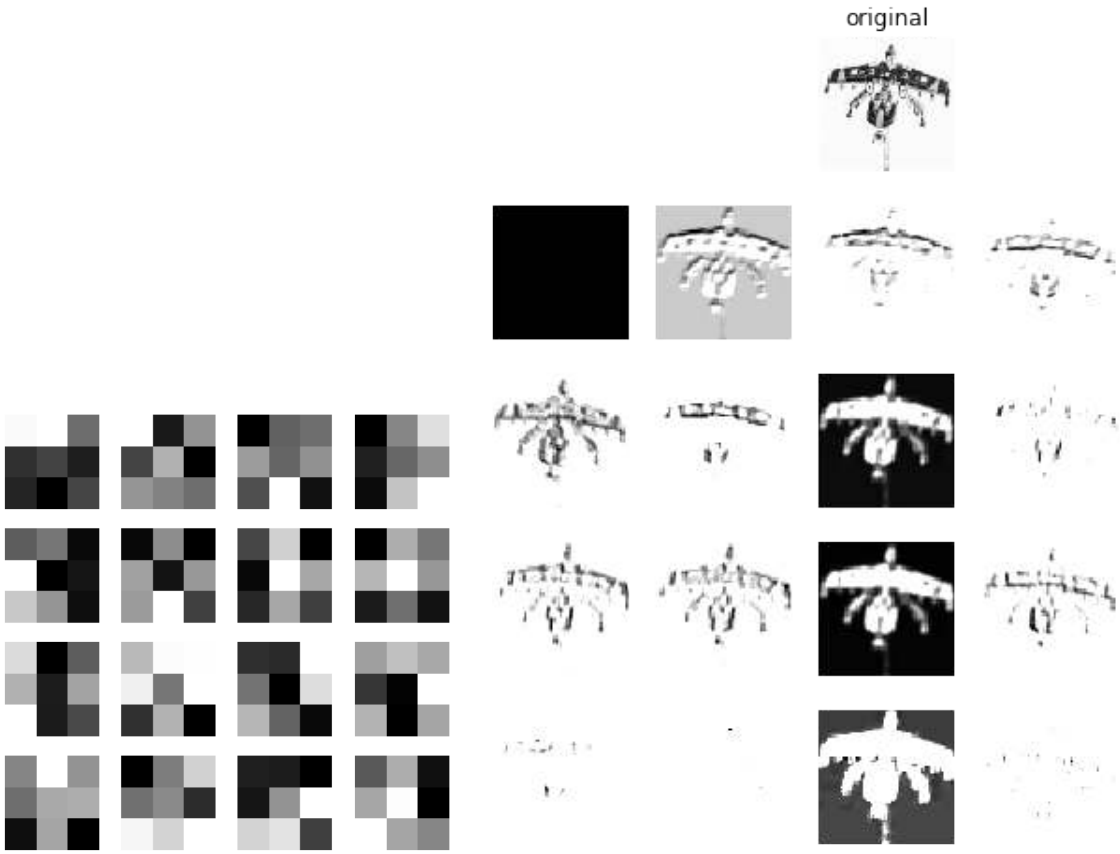


Figura A.6: kernels y mapa de características de la capa de convolución sobre una imagen de entrada X.

```
visualize.plot_conv_weights(net.layers_[1])
visualize.plot_conv_activity(net.layers_['conv2d3'], X)
```

Bibliografía

Stephen Milborrow. *Locating Facial Features with Active Shape Models*. PhD thesis, University of Cape Town, 2007.

Stephen Milborrow, Tom E. Bishop, and Fred Nicolls. Multiview active shape models with SIFT descriptors for the 300-W face landmark challenge. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 378–385, 2013. ISBN 9781479930227. doi: 10.1109/ICCVW.2013.57.

Stephen Milborrow and Fred Nicolls. Active Shape Models with SIFT Descriptors and MARS. *VISAPP*, (i), 2014.

Jonathan Long, Ning Zhang, and Trevor Darrell. Do Convnets Learn Correspondence? *Advances in Neural Information . . .*, pages 1601–1609, 2014. ISSN 03029743. doi: 10.1007/978-3-642-33863-2{_}51. URL http://papers.nips.cc/paper/5420-do-convnets-learn-correspondence.pdf$\delimiter"026E30F$nhttp://papers.nips.cc/paper/5420-do-convnets-learn-correspondence.

Paul Viola and Michael Jones. Robust real-time object detection. *International Journal of Computer Vision*, 57:137–154, 2001. ISSN 09205691. doi: http://dx.doi.org/10.1023/B:VISI.0000013087.49260.fb. URL http://scholar.google.com/scholar?hl=en{&}btnG=Search{&}q=intitle:Robust+Real-time+Object+Detection{#}0.

Rolando González-José, Ignacio Escapa, Walter a Neves, Rubén Cúneo, and Héctor M Pucciarelli. Cladistic analysis of continuous modularized traits provides phylogenetic signals in Homo evolution. *Nature*, 453(7196):775–8, 2008a. ISSN 1476-4687. doi: 10.1038/nature06891. URL http://www.ncbi.nlm.nih.gov/pubmed/18454137.

Fred L. Bookstein, Philipp Gunz, Philipp Mitteroecker, Hermann Prossinger, Katrin Sch??fer, and Horst Seidler. Cranial integration in Homo: Singular warps analysis of the midsagittal plane in ontogeny and evolution. *Journal of Human Evolution*, 44(2):167–187, 2003. ISSN 00472484. doi: 10.1016/S0047-2484(02)00201-4.

DE Slice. Modern morphometrics. In *Modern morphometrics in physical anthropology*, pages 1–46. 2005. ISBN 0306486970. doi: 10.1007/0-387-27614-9{_}1. URL http://link.springer.com/chapter/10.1007/0-387-27614-9{_}1.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553): 436–444, 2015. ISSN 0028-0836. doi: 10.1038/nature14539. URL http://dx.doi.org/10.1038/nature14539$\delimiter"026E30F$n10.1038/nature14539.

Aravindh Mahendran and Andrea Vedaldi. Understanding deep image representations by inverting them. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 07-12-June-2015, pages 5188–5196, 2015. ISBN 9781467369640. doi: 10.1109/CVPR.2015.7299155.

Celia Cintas, Mirsha Quinto-Sánchez, Victor Acuña, Carolina Paschetta, Soledad De Azevedo, Caio Silva de Cerqueira, Virginia Ramallo, Carla Gallo, Giovanni Poletti, Maria Catira Bortolini, Samuel Canizales-Quinteros, Rothhammer Francisco, Gabriel Bedoya, Andres Ruiz-Linares, Rolando Gonzaléz-José, and Claudio Delrieux. Automatic ear detection and feature extraction using Geometric Morphometrics and Convolutional Neural Networks. *IET Biometrics*, dec 2016a. ISSN 2047-4938. doi: 10.1049/iet-bmt.2016.0002. URL http://digital-library.theiet.org/content/journals/10.1049/iet-bmt.2016.0002.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. *MIT press*, page 1, 2016. ISSN 1548-7091. doi: 10.1038/nmeth.3707. URL http://files.sig2d.org/sig2d14.pdf{#}page=5.

Haohan Wang and Bhiksha Raj. On the Origin of Deep Learning. feb 2017. URL http://arxiv.org/abs/1702.07800.

Mirsha Emmanuel Quinto Sanchez. *Asimetría facial un estudio de bioantropología integrativa en Poblaciones Cosmopolitas Latinoamericanas*. PhD thesis, 2015. URL http://www.fcnym.unlp.edu.ar/postgrado/verTesis.php?id=2422.

Ruma Purkait and Priyanka Singh. A test of individuality of human external ear pattern: its application in the field of personal identification. *Forensic science international*, 178 (2-3):112–8, 2008. ISSN 1872-6283. doi: 10.1016/j.forsciint.2008.02.009.

Ilker Ercan, Senem Turan Ozdemir, Abdullah Etoz, Deniz Sigirli, R. Shane Tubbs, Marios Loukas, and Ibrahim Guney. Facial asymmetry in young healthy subjects evaluated by statistical shape analysis. *Journal of Anatomy*, 213(6):663–669, 2008. ISSN 00218782. doi: 10.1111/j.1469-7580.2008.01002.x.

Franc Solina, Peter Peer, Borut Batagelj, Samo Juvan, and Jure Kovač. Color-based face detection in the"15 seconds of fame.ªrt installation. *Proceedings of Mirage 2003, Conference on Computer Vision / Computer Graphics*, pages 38–47, 2003.

Ajay Kumar and Chenye Wu. Automated human identification using ear imaging. *Pattern Recognition*, 45(3):956–968, 2012.

Vincent Dumoulin and Francesco Visin. A guide to convolution arithmetic for deep learning. *CoRR*, abs/1603.07285, 2016. URL http://dblp.uni-trier.de/db/journals/corr/corr1603.html#DumoulinV16.

A Lanitis, CJ Taylor, and TF Cootes. Automatic face identification system using flexible appearance models. *Image and Vision Computing*, 13(5):393–401, 1995. ISSN 0262-8856. doi: 10.1016/0262-8856(95)99726-H. URL http://www.sciencedirect.com/science/article/pii/026288569599726H$\delimiter"026E30F$nhttp://www.sciencedirect.com/science/article/pii/026288569599726H/pdf?md5=44ed418408e2f7e84e0e9a1c9a447ec0{&}pid=1-s2.0-026288569599726H-main.pdf.

Laurenz Wiskott, Jean Marc Fellous, Norbert Krüger, and Christoph Von der Malsburg. Face recognition by elastic bunch graph matching. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes*

in Bioinformatics), volume 1296, pages 456–463, 1997. ISBN 3540634606. doi: 10.1007/3-540-63460-6{_}150.

Paola Campadelli, Raffaella Lanzarotti, and Chiara Savazzi. A feature-based face recognition system. In *Proceedings - 12th International Conference on Image Analysis and Processing, ICIAP 2003*, pages 68–73, 2003. ISBN 0769519482. doi: 10.1109/ICIAP.2003.1234027.

Kang Kang Liu, Axel Weissenfeld, Joern Ostermann, and Xinghan Xinghan Luo. Robust AAM building for morphing in an image-based facial animation system. In *2008 IEEE International Conference on Multimedia and Expo*, pages 933–936. IEEE, jun 2008. ISBN 978-1-4244-2570-9. doi: 10.1109/ICME.2008.4607589. URL http://ieeexplore.ieee.org/document/4607589/.

Ying Li Tian, Takeo Kanade, and Jeffrey F. Conn. Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2):97–115, 2001. ISSN 01628828. doi: 10.1109/34.908962.

M Pantic and L J M Rothkrantz. Automatic analysis of facial expressions: The state of the art. *Ieee Transactions on Pattern Analysis and Machine Intelligence*, 22(12): 1424–1445, 2000. ISSN 01628828. doi: 10.1109/34.895976. URL <GotoISI>://WOS:000165901900006$\delimiter"026E30F$nhttp://ieeexplore.ieee.org/ielx5/34/19391/00895976.pdf?tp={&}arnumber=895976{&}isnumber=19391.

Zheng Ying, Chang Jianglong, Zheng Zhigang, and Wang Zengfu. 3D face reconstruction from stereo: A model based approach. In *Proceedings - International Conference on Image Processing, ICIP*, volume 3, 2006. ISBN 1424414377. doi: 10.1109/ICIP.2007.4379247.

Jianya Guo, Xi Mei, and Kun Tang. Automatic landmark annotation and dense correspondence registration for 3D human facial images. *BMC bioinformatics*, 14(1):232, 2013. ISSN 1471-2105. doi: 10.1186/1471-2105-14-232. URL http://www.biomedcentral.com/1471-2105/14/232.

Jorge Gómez-Valdés, Tábita Hünemeier, Mirsha Quinto-Sánchez, Carolina Paschetta, Soledad de Azevedo, Marina F González, Neus Martínez-Abadías, Mireia Esparza,

Héctor M Pucciarelli, Francisco M Salzano, Claiton H D Bau, Maria Cátira Bortolini, and Rolando González-José. Lack of Support for the Association between Facial Shape and Aggression: A Reappraisal Based on a Worldwide Population Genetics Perspective. *PLOS ONE*, 8(1):1–8, 2013. doi: 10.1371/journal.pone.0052317. URL http://dx.doi.org/10.1371{%}2Fjournal.pone.0052317.

Mirsha Quinto-Sánchez, Kaustubh Adhikari, Victor Acuña-Alonzo, Celia Cintas, Caio Cesar Silva de Cerqueira, Virginia Ramallo, Lucia Castillo, Arodi Farrera, Claudia Jaramillo, Williams Arias, Macarena Fuentes, Paola Everardo, Francisco de Avila, Jorge Gomez-Valdés, Tábita Hünemeier, Shara Gibbon, Carla Gallo, Giovanni Poletti, Javier Rosique, Maria Cátira Bortolini, Samuel Canizales-Quinteros, Francisco Rothhammer, Gabriel Bedoya, Andres Ruiz-Linares, and Rolando González-José. Facial asymmetry and genetic ancestry in latin american admixed populations. *American Journal of Physical Anthropology*, 157(1):58–70, 2015a. ISSN 1096-8644. doi: 10.1002/ajpa.22688. URL http://dx.doi.org/10.1002/ajpa.22688.

Stefan Schlager and Alexandra Rüdell. Analysis of the human osseous nasal shape—population differences and sexual dimorphism. *American Journal of Physical Anthropology*, 157(4):571–581, 2015. ISSN 1096-8644. doi: 10.1002/ajpa.22749. URL http://dx.doi.org/10.1002/ajpa.22749.

Carolina Paschetta, Soledad De Azevedo, Marina González, Mirsha Quinto-Sánchez, Celia Cintas, Hugo Varela, Jorge Gómez-Valdés, Gabriela Sánchez-Mejorada, and Rolando González-José. Shifts in subsistence type and its impact on the human skull's morphological integration. *American Journal of Human Biology*, 28(1):118–128, jan 2016. ISSN 10420533. doi: 10.1002/ajhb.22746. URL http://doi.wiley.com/10.1002/ajhb.22746.

P Hammond. The use of 3D face shape modelling in dysmorphology. *Archives of Disease in Childhood*, 92(12):1120–1126, 2007. ISSN 0003-9888. doi: 10.1136/adc.2006.103507. URL http://adc.bmj.com.pklibresources.health.wa.gov.au/content/92/12/1120.full.pdf?sid=a07ee8a2-98c9-41b6-8bfe-94f4a7cf14b3.

P Hammond, T J Hutton, J E Allanson, B Buxton, L E Campbell, J Clayton-Smith,

D Donnai, A Karmiloff-Smith, K Metcalfe, K C Murphy, M Patton, B Pober, K Prescott, P Scambler, A Shaw, A C M Smith, A F Stevens, I K Temple, R Hennekam, and M Tassabehji. Discriminating power of localized three-dimensional facial morphology. *American Journal of Human Genetics*, 77(6):999–1010, 2005. ISSN 0002-9297. doi: 10.1086/498396.

Seth M. Weinberg, Katherine Neiswanger, Joan T. Richtsmeier, Brion S. Maher, Mark P. Mooney, Michael I. Siegel, and Mary L. Marazita. Three-dimensional morphometric analysis of craniofacial shape in the unaffected relatives of individuals with nonsyndromic orofacial clefts: A possible marker for genetic susceptibility. *American Journal of Medical Genetics Part A*, 146A(4):409–420, feb 2008. ISSN 15524825. doi: 10.1002/ajmg.a.32177. URL http://www.ncbi.nlm.nih.gov/pubmed/18203157http://doi.wiley.com/10.1002/ajmg.a.32177.

K Skaria Alexander, David J Stott, Branavan Sivakumar, and Norbert Kang. A morphometric study of the human ear. *Journal of plastic, reconstructive & aesthetic surgery : JPRAS*, 64(1):41–7, jan 2011. ISSN 1878-0539. doi: 10.1016/j.bjps.2010.04.005. URL http://www.ncbi.nlm.nih.gov/pubmed/20447883.

Fajri Kurniawan, Mohd Shafry, Mohd Rahim, and Mohammed S Khalil. Geometrical and Eigenvector Features for Ear Recognition. pages 57–62, 2014.

Yahui Liu, Bob Zhang, and David Zhang. Ear-parotic face angle: A unique feature for 3D ear recognition. *Pattern Recognition Letters*, 53:9–15, feb 2015. ISSN 01678655. doi: 10.1016/j.patrec.2014.10.014. URL http://linkinghub.elsevier.com/retrieve/pii/S0167865514003316.

A. Midori Albert, Karl Ricanek, and Eric Patterson. A review of the literature on the aging adult skull and face: Implications for forensic science research and applications. *Forensic Science International*, 172(1):1–9, oct 2007. ISSN 03790738. doi: 10.1016/j.forsciint.2007.03.015. URL http://www.ncbi.nlm.nih.gov/pubmed/17434276http://linkinghub.elsevier.com/retrieve/pii/S0379073807001624.

Narayanan Ramanathan, Rama Chellappa, and Soma Biswas. Computational methods

for modeling facial aging: A survey. *Journal of Visual Languages & Computing*, 20(3): 131–144, 2009. ISSN 1045926X. doi: 10.1016/j.jvlc.2009.01.011.

Yun Fu, Guodong Guo, and Thomas S. Huang. Age synthesis and estimation via faces: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(11): 1955–1976, 2010. ISSN 01628828. doi: 10.1109/TPAMI.2010.36.

Fan Liu, Fedde van der Lijn, Claudia Schurmann, Gu Zhu, M. Mallar Chakravarty, Pirro G. Hysi, Andreas Wollstein, Oscar Lao, Marleen de Bruijne, M. Arfan Ikram, Aad van der Lugt, Fernando Rivadeneira, André G. Uitterlinden, Albert Hofman, Wiro J. Niessen, Georg Homuth, Greig de Zubicaray, Katie L. McMahon, Paul M. Thompson, Amro Daboul, Ralf Puls, Katrin Hegenscheid, Liisa Bevan, Zdenka Pausova, Sarah E. Medland, Grant W. Montgomery, Margaret J. Wright, Carol Wicking, Stefan Boehringer, Timothy D. Spector, Tomáš Paus, Nicholas G. Martin, Reiner Biffar, and Manfred Kayser. A Genome-Wide Association Study Identifies Five Loci Influencing Facial Morphology in Europeans. *PLoS Genetics*, 8, 2012. ISSN 15537390. doi: 10.1371/journal.pgen.1002932.

Kaustubh Adhikari, Macarena Fuentes-guajardo, Mirsha Quinto-sa, Victor Acun, Claudia Jaramillo, William Arias, Rodrigo Barquera Lozano, Virginia Ramallo, Caio C Silva De Cerqueira, Malena Hurtado, Valeria Villegas, Vanessa Granja, Carla Gallo, Giovanni Poletti, Lavinia Schuler-faccini, Francisco M Salzano, Samuel Canizales-quinteros, Michael Cheeseman, Javier Rosique, and Gabriel Bedoya. A genome-wide association scan implicates DCHS2, RUNX2, GLI3, PAX1 and EDAR in human facial variation. *Nature Communications*, 7(May):1–11, 2016. ISSN 2041-1723. doi: 10.1038/ncomms11616.

Eitan Segev, Hemo Yoram, Shlomo Wientroub, Ovadia Dror, Michael Fishkin, David M Steinberg, Shlomo Hayek, E Segev, Á Y Hemo, Á S Wientroub, Á D Ovadia, Á M Fishkin, Á S Hayek, Y Hemo, S Wientroub, D Ovadia, M Fishkin, S Hayek, and D M Steinberg. Intra-and interobserver reliability analysis of digital radiographic measurements for pediatric orthopedic parameters using a novel PACS integrated computer software program. *J Child Orthop*, 2010. doi: 10.1007/s11832-010-0259-5.

A Kamoen, L Dermaut, and R Verbeeck. The clinical significance of error measurement

in the interpretation of treatment results. *European journal of orthodontics*, 23(5): 569–78, oct 2001. ISSN 0141-5387. URL http://www.ncbi.nlm.nih.gov/pubmed/11668876.

Colin R. Goodall and Kanti V. Mardia. A geometrical derivation of the shape density. *Advances in Applied Probability*, 23(3):496–514, 1991. ISSN 00018678. URL http://www.jstor.org/stable/1427619.

Andrés Ruiz-Linares, Kaustubh Adhikari, Victor Acuña-Alonzo, Mirsha Quinto-Sanchez, Claudia Jaramillo, William Arias, Macarena Fuentes, María Pizarro, Paola Everardo, Francisco de Avila, Jorge Gómez-Valdés, Paola León-Mimila, Tábita Hunemeier, Virginia Ramallo, Caio C Silva de Cerqueira, Mari-Wyn Burley, Esra Konca, Marcelo Zagonel de Oliveira, Mauricio Roberto Veronez, Marta Rubio-Codina, Orazio Attanasio, Sahra Gibbon, Nicolas Ray, Carla Gallo, Giovanni Poletti, Javier Rosique, Lavinia Schuler-Faccini, Francisco M Salzano, Maria-Cátira Bortolini, Samuel Canizales-Quinteros, Francisco Rothhammer, Gabriel Bedoya, David Balding, and Rolando Gonzalez-José. Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. *PLoS genetics*, 10(9):e1004572, sep 2014. ISSN 1553-7404. doi: 10.1371/journal.pgen.1004572. URL http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1004572.

Caio Cesar Silva De Cerqueira, Tábita Hünemeier, Jorge Gomez-Valdés, Virgínia Ramallo, Carla Daiana Volasko-Krause, Ana Angélica Leal Barbosa, Pedro Vargas-Pinilla, Rodrigo Ciconet Dornelles, Danaê Longo, Francisco Rothhammer, Gabriel Bedoya, Samuel Canizales-Quinteros, Victor Acuña-Alonzo, Carla Gallo, Giovanni Poletti, Rolando González-José, Francisco Mauro Salzano, Sídia Maria Callegari-Jacques, Lavínia Schuler-Faccini, Andrés Ruiz-Linares, and Maria Cátira Bortolini. Implications of the admixture process in skin color molecular assessment. *PLoS ONE*, 9(5), 2014. ISSN 19326203. doi: 10.1371/journal.pone.0096886.

Mirsha Quinto-Sánchez, Kaustubh Adhikari, Victor Acuña-Alonzo, Celia Cintas, Caio Cesar Silva De Cerqueira, Virginia Ramallo, Lucia Castillo, Arodi Farrera, Claudia Jaramillo, Williams Arias, Macarena Fuentes, Paola Everardo, Francisco De Avila, Jorge Gomez-Valdés, Tábita Hünemeier, Shara Gibbon, Carla Gallo, Giovanni Poletti, Javier Rosique,

Maria Cátira Bortolini, Samuel Canizales-Quinteros, Francisco Rothhammer, Gabriel Bedoya, Andres Ruiz-Linares, and Rolando González-José. Facial asymmetry and genetic ancestry in Latin American admixed populations. *American Journal of Physical Anthropology*, 157(1):58–70, 2015b. ISSN 10968644. doi: 10.1002/ajpa.22688.

Mirsha Quinto-Sánchez, Celia Cintas, Caio Cesar Silva de Cerqueira, Virginia Ramallo, Victor Acuña-Alonzo, Kaustubh Adhikari, Lucía Castillo, Jorge Gomez-Valdés, Paola Everardo, Francisco De Avila, Tábita Hünemeier, Claudia Jaramillo, Williams Arias, Macarena Fuentes, Carla Gallo, Giovani Poletti, Lavinia Schuler-Faccini, Maria Cátira Bortolini, Samuel Canizales-Quinteros, Francisco Rothhammer, Gabriel Bedoya, Javier Rosique, Andrés Ruiz-Linares, and Rolando González-José. Socioeconomic Status Is Not Related with Facial Fluctuating Asymmetry: Evidence from Latin-American Populations. *PLOS ONE*, 12(1):e0169287, jan 2017. ISSN 1932-6203. doi: 10.1371/journal.pone.0169287. URL http://dx.plos.org/10.1371/journal.pone.0169287.

Kaustubh Adhikari, Guillermo Reales, Andrew J. P. Smith, Esra Konka, Jutta Palmen, Mirsha Quinto-Sanchez, Victor Acuña-Alonzo, Claudia Jaramillo, William Arias, Macarena Fuentes, María Pizarro, Rodrigo Barquera Lozano, Gastón Macín Pérez, Jorge Gómez-Valdés, Hugo Villamil-Ramírez, Tábita Hunemeier, Virginia Ramallo, Caio C. Silva de Cerqueira, Malena Hurtado, Valeria Villegas, Vanessa Granja, Carla Gallo, Giovanni Poletti, Lavinia Schuler-Faccini, Francisco M. Salzano, Maria-Cátira Bortolini, Samuel Canizales-Quinteros, Francisco Rothhammer, Gabriel Bedoya, Rosario Calderón, Javier Rosique, Michael Cheeseman, Mahmood F. Bhutta, Steve E. Humphries, Rolando Gonzalez-José, Denis Headon, David Balding, and Andrés Ruiz-Linares. A genome-wide association study identifies multiple loci for variation in human ear morphology. *Nature Communications*, 6(May):7500, 2015. ISSN 2041-1723. doi: 10.1038/ncomms8500. URL http://www.nature.com/doifinder/10.1038/ncomms8500.

A. Carvajal-Rodríguez, P. CONDE-PADÍN, and E. ROLÁN-ALVAREZ. Decomposing shell form into size and shape by geometric morphometric methods in two sympatric ecotypes of littorina saxatilis. *Journal of Molluscan Studies*, 71(4):313–318, 2005. doi: 10.1093/mollus/eyi037. URL +http://dx.doi.org/10.1093/mollus/eyi037.

Alexey Shipunov and Richard Bateman. Geometric morphometrics as a tool for un-

derstanding {\em Dactylorhiza} {(Orchidaceae)} diversity in European Russia. *Biological Journal of the Linnean Society*, 85(1):1–12, 2005. ISSN 0024-4066. doi: 10.1111/j.1095-8312.2005.00468.x.

Christian Lovato, Umberto Castellani, Carlo Zancanaro, and Andrea Giachetti. Automatic labelling of anatomical landmarks on 3D body scans. *Graphical Models*, 76(6):648–657, 2014. ISSN 15240703. doi: 10.1016/j.gmod.2014.07.001.

Jonathan Tompson, Arjun Jain, Yann LeCun, and Christoph Bregler. Joint Training of a Convolutional Network and a Graphical Model for Human Pose Estimation. *Advances in neural information processing systems*, pages 1799—–1807, 2014. ISSN 10495258.

Andrei State, Gentaro Hirota, David T Chen, William F Garrett, and Mark a Livingston. Superior augmented reality registration by integrating landmark tracking and magnetic tracking. *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques SIGGRAPH 96*, 30(Annual Conference Series):429–438, 1996. ISSN 0097-8930. doi: 10.1145/237170.237282. URL http://portal.acm.org/citation.cfm?doid=237170.237282.

V Kitanovski and E Izquierdo. 3D Tracking of Facial Features for Augmented Reality Applications. *The 12th International Workshop on Image Analysis for Multimedia Interactive Services - WIAMIS 2011*, 2011. ISSN 21585881. URL uuid:4a107a5a-a5a1-4d06-af4f-5c8da20c709b.

Caio Cesar Silva de Cerqueira, Virginia Ramallo, Tabita Hunemeier, Soledad de Azevedo, Mirsha Quinto Sanchez, Carolina Paschetta, Celia Cintas, Marina González, Lavínia Schüler-Faccini, Maria Cátira Bortolini, Rol, and o González-José. Predicting Physical Features and Diseases by DNA Analysis: Current Advances and Future Challenges. *Journal of Forensic Research*, 7(4):–, 2016. ISSN 21577145. doi: 10.4172/2157-7145.1000336. URL http://www.omicsonline.org/open-access/predicting-physical-features-and-diseases-by-dna-analysis-current-advancesand-future-challenges-2157-7145-1000336.php?aid=79947.

Celia Cintas, Pablo Navarro, Virginia Ramallo, Bruno Pazos, Claudio Delrieux, Anahí Ruderman, Carolina Paschetta, Soledad De Azevedo, and Rolando González-José. Posicionamiento automático de landmarks corporales 3d mediante morfometría geométrica y redes neuronales: aplicaciones bioantropológicas. In *Actas de las XIII Jornadas Nacionales de Antropología Biológica*, 2017.

Celia Cintas, Claudio Delrieux, Mirsha Quinto-Sánchez, Carolina Paschetta, Virginia Ramallo, Caio Cesar De Cerqueira, Soledad De Azevedo, and Rolando González-José. Posicionamiento automático de landmarks anatómicos 2d mediante morfometría geométrica y deep learning: aplicaciones bioantropológicas. In *Actas del XIV Congreso Asociación Latinoamericana de Antropología Biológica*. Asociación Latinoamericana de Antropología Biológica, 2016b.

Celia Cintas, Mirsha Quinto-Sánchez, Claudio Delrieux, and Rolando González-José. Automatic landmarking app for bioanthropology. In *Proceedings of IV EURO WG Conference on Operational Research in Computational Biology, Bioinformatics and Medicine.* EURO working group on Computational Biology, Bioinformatics and Medicine, 2014a.

Celia Cintas, Mirsha Quinto-Sánchez, Claudio Delrieux, and Rolando González-José. Python in the world of biological anthropology. In *EuroPython*, Berlin, Alemania, 2014b.

Celia Cintas, Mirsha Quinto-Sánchez, Gloria Bianchi, Nahuel Defossé, Claudio Delrieux, and Rolando González-José. Aplicación bioantropológica para manipulación de landmarks faciales en 2d. In *Actas en 16° Edición del Workshop de Investigadores en Ciencias de la Computación*, 2014c.

Mirsha Quinto-Sánchez, Celia Cintas, Carolina Paschetta, Virginia Ramallo, Caio Cesar De Cerqueira, Soledad De Azevedo, and Rolando González-José. Asimetría fluctuante facial y su relación con índices socioeconómicos. In *Actas del XIII Congreso de la Asociación Latinoamericana de Antropología Biológica*. Asociación Latinoamericana de Antropología Biológica, 2014.

Celia Cintas, Gloria Bianchi, Nahuel Defossé, Claudio Delrieux, Mirsha Quinto-Sánchez, and Rolando González-José. Popeye: Bioanthropology app for the automatic positioning of anatomical landmarks in 2d and 3d. In *Actas del 4to. Congreso Argentino de*

Bioinformática y Biología Computacional (4CAB2C) y 4ta. Conferencia Internacional de la Sociedad Iberoamericana de Bioinformática (SolBio), 2013a.

Celia Cintas, Claudio Delrieux, Mirsha Quinto-Sánchez, and Rolando González-José. Posicionamiento automático de landmarks anatómicos en 2 y 3d: aplicaciones bioantropológicas. In *Actas de las XI Jornadas Nacionales de Antropología Biológica*, 2013b.

Celia Cintas, Claudio Delrieux, and Rolando González-José. Posicionamiento automático de landmarks anatómicos en ojos. In *Actas del Congreso Argentino de Ciencias de la Computación (CACIC)*, 2012.

Stanislav Katina, Alena Šefčáková, Jana Velemínská, Jaroslav Brůžek, and Petr Velemínský. A geometric approach to cranial sexual dimorphism in fossil skulls from Předmostí (Upper Palaeolithic, Czech Republic). *J. Nat. Mus., Nat. Hist. Ser.*, 173(Svoboda 2000):133–144, 2004.

J Shi, A Samal, and D Marx. How effective are landmarks and their geometry for face recognition? *Computer Vision and Image Understanding*, 102(2):117–133, 2006. ISSN 1077-3142. doi: http://dx.doi.org/10.1016/j.cviu.2005.10.002. URL http://www.sciencedirect.com/science/article/pii/S1077314205001761.

Shi Jiazheng, Ashok Samal, and David Marx. Face recognition using landmark-based bidimensional regression. *Proceedings - IEEE International Conference on Data Mining, ICDM*, pages 765–768, 2005. ISSN 15504786. doi: 10.1109/ICDM.2005.61.

G. M. Beumer, Q. Tao, A. M. Bazen, and R. N J Veldhuis. A landmark paper in face recognition. *FGR 2006: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, 2006:73–78, 2006. doi: 10.1109/FGR.2006.10.

S. Pflug and C. Busch. Ear biometrics: a survey of detection, feature extraction and recognition methods. *IET Biometrics*, 1(2):114, 2012. ISSN 20474938. doi: 10.1049/iet-bmt.2011.0003.

Ricardo Teles Freitas, Kelson Romulo Teixeira Aires, and Victor Eulalio Sousa Campelo. Locating Facial Landmarks towards Plastic Surgery. *Brazilian Symposium of Computer*

Graphic and Image Processing, 2015-October:219–225, 2015. ISSN 15301834. doi: 10.1109/SIBGRAPI.2015.40.

T. F. Cootes and C. J. Taylor. Active Shape Models - 'smart snakes'. In *Procedings of the British Machine Vision Conference 1992*, pages 28.1–28.10, 1992. ISBN 3-540-19777-X. doi: 10.5244/C.6.28. URL http://www.bmva.org/bmvc/1992/bmvc-92-028.html.

Timothy F Cootes, Andrew Hill, Christopher J Taylor, and Jane Haslam. Use of active shape models for locating structures in medical images. *Image and vision computing*, 12(6):355–365, 1994.

A. Hill, T. F. Cootes, and C. J. Taylor. Active shape models and the shape approximation problem. *Image and Vision Computing*, 14(8 SPEC. ISS.):601–607, 1996. ISSN 02628856. doi: 10.1016/0262-8856(96)01097-9.

D. Cristinacce and T. F. Cootes. Boosted Regression Active Shape Models. *Procedings of the British Machine Vision Conference 2007*, pages 79.1–79.10, 2007. ISSN 15457885. doi: 10.5244/C.21.79. URL http://www.bmva.org/bmvc/2007/papers/paper-131.html.

T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active Appearance Models. *Proc. European Conference on Computer Vision (ICCV)*, 2:484–498, 1998. ISSN 0162-8828. doi: 10.1109/34.927467.

G.J. Edwards, C.J. Taylor, and T.F. Cootes. Interpreting face images using active appearance models. *Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition*, pages 300–305, 1998. ISSN 01628828. doi: 10.1109/AFGR.1998.670965. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=670965.

T F Cootes, G Edwards, and C J Taylor. Comparing Active Shape Models with Active Appearance Models. 1999. URL https://pdfs.semanticscholar.org/1da2/f22f46de0726cacfbe894946ea72032e8fbc.pdf.

Zisheng Li, Jun-ichi Imai, and Masahide Kaneko. Facial feature localization using statistical models and SIFT descriptors. *RO-MAN 2009 - The 18th IEEE International*

Symposium on Robot and Human Interactive Communication, pages 961–966, 2009. doi: 10.1109/ROMAN.2009.5326323. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5326323.

Ajita Rattani, D. R. Kisku, Andrea Lagorio, and Massimo Tistarelli. Facial template synthesis based on SIFT features. *2007 IEEE Workshop on Automatic Identification Advanced Technologies - Proceedings*, pages 69–73, 2007. doi: 10.1109/AUTOID.2007.380595.

David G. Lowe. Object recognition from local scale-invariant features. *Proceedings of the Seventh IEEE International Conference on Computer Vision*, 2([8):1150–1157, 1999. ISSN 0-7695-0164-8. doi: 10.1109/ICCV.1999.790410. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=790410.

David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004. ISSN 09205691. doi: 10.1023/B:VISI.0000029664.99615.94.

Jan J. Koenderink. The Structure of Images. *Biological Cybernetics*, 425:139–145, 1989.

T Lindeberg. Principles for automatic scale selection. *Handbook on Computer Vision and Applications*, pages 239–274, 1999. URL http://www.doc.ic.ac.uk/{~}xh1/Referece/Scale-Space-Theory/Principles-for-automatic-scale-selection.pdf{%}5Cnpapers2://publication/uuid/79C341F9-8C77-4F1D-90CE-6B9F80F95AE4.

Yoav E Freund Robert Schapire. A Short Introduction to Boosting. *Journal of Japanese Society for Artificial Intelligence*, 14(5):771–780, 1999. URL www.research.att.com/.

C P Papageorgiou, M Oren, and T Poggio. A general framework for object detection, 1998.

M L Zelditch, D L Swiderski, Hd David Sheets, and W L Fink. Geometric morphometrics for biologists. *Elsevier*, 59(3):457, 2004. ISSN 00255645. doi: 10.1016/B978-0-12-386903-6.00001-0. URL http:

//www.sciencedirect.com/science/article/B848M-4MWYDVJ-7/2/2ec7a8306777645c5716a53774ec699c$\delimiter"026E30F$nhttp://books.google.com/books?id=LKCVAGn8vkoC{&}pgis=1.

Philipp Mitteroecker and Philipp Gunz. Advances in Geometric Morphometrics, 2009. ISSN 0071-3260.

Kim van der Linde and David Houle. Inferring the nature of allometry from geometric data. *Evolutionary Biology*, 36(3):311–322, 2009. ISSN 00713260. doi: 10.1007/s11692-009-9061-z.

Una Strand Vidarsdottir, Paul O'Higgins, and Chris Stringer. A geometric morphometric study of regional differences in the ontogeny of the modern human facial skeleton+. *Journal of Anatomy*, 201(3):211–229, 2002. ISSN 0021-8782. doi: 10.1046/j.1469-7580.2002.00092.x. URL http://www.scopus.com/inward/record.url?eid=2-s2.0-0036737886{&}partnerID=tZOtx3y1.

Daniel E. Lieberman, Julian Carlo, Marcia Ponce de León, and Christoph P.E. Zollikofer. A geometric morphometric analysis of heterochrony in the cranium of chimpanzees and bonobos. *Journal of Human Evolution*, 52(6):647–662, jun 2007. ISSN 00472484. doi: 10.1016/j.jhevol.2006.12.005. URL http://www.ncbi.nlm.nih.gov/pubmed/17298840http://linkinghub.elsevier.com/retrieve/pii/S0047248407000024.

Daniel E. Lieberman, Gail E. Krovitz, Franklin W. Yates, Maureen Devlin, and Marisa St. Claire. Effects of food processing on masticatory strain and craniofacial growth in a retrognathic face. *Journal of Human Evolution*, 46(6):655–677, 2004. ISSN 00472484. doi: 10.1016/j.jhevol.2004.03.005.

Carolina Paschetta, Soledad de Azevedo, Lucía Castillo, Neus Martínez-Abadías, Miquel Hernández, Daniel E. Lieberman, and Rolando González-José. The influence of masticatory loading on craniofacial morphology: A test case across technological transitions in the Ohio valley. *American Journal of Physical Anthropology*, 141(2):297–314, feb 2010. ISSN 00029483. doi: 10.1002/ajpa.21151. URL http://www.ncbi.nlm.nih.gov/pubmed/19902454http://doi.wiley.com/10.1002/ajpa.21151.

Christian Peter Klingenberg, Marta Barluenga, and Axel Meyer. Shape analysis of symmetric structures: quantifying variation among individuals and asymmetry. *Evolution; international journal of organic evolution*, 56(10):1909–20, oct 2002. ISSN 0014-3820. URL http://www.ncbi.nlm.nih.gov/pubmed/12449478.

Rolando González-José, Maria Cátira Bortolini, Fabrício R. Santos, and Sandro L. Bonatto. The peopling of America: Craniofacial shape variation on a continental scale and its interpretation from an interdisciplinary view. *American Journal of Physical Anthropology*, 137(2):175–187, 2008b. ISSN 00029483. doi: 10.1002/ajpa.20854.

S. I. Perez, J. Klaczko, G. Rocatti, and S. F. Dos Reis. Patterns of cranial shape diversification during the phylogenetic branching process of New World monkeys (Primates: Platyrrhini). *Journal of Evolutionary Biology*, 24(8): 1826–1835, aug 2011. ISSN 1010061X. doi: 10.1111/j.1420-9101.2011.02309.x. URL http://www.ncbi.nlm.nih.gov/pubmed/21615587http://doi.wiley.com/10.1111/j.1420-9101.2011.02309.x.

R R Sokal and F J Rohlf. *Biometry: The Principles and Practices of Statistics in Biological Research [Hardcover]*. 1995. ISBN 0716724111. doi: 10.2307/2331669. URL http://www.amazon.com/Biometry-Principles-Practices-Statistics-Biological/dp/0716724111.

Sewall Wright. *Evolution and the genetics of populations*. University of Chicago Press, 1984. ISBN 0226910415.

Douglas S Falconer and Trudy F C Mackay. *Introduction to Quantitative Genetics (4th Edition)*, volume 12. 1996. ISBN 9780582243026. URL http://www.amazon.com/Introduction-Quantitative-Genetics-Douglas-Falconer/dp/0582243025.

Fred L. Bookstein. A statistical method for biological shape comparisons. *Journal of Theoretical Biology*, 107(3):475–520, 1984. ISSN 10958541. doi: 10.1016/S0022-5193(84)80104-6.

Daythal Lee Kendall. A syntactic analysis of takelma texts, 1977.

J. C. Gower. Generalized procrustes analysis. *Psychometrika*, 40(1):33–51, 1975. ISSN 00333123. doi: 10.1007/BF02291478.

Jos M F Ten Berge. Orthogonal procrustes rotation for two or more matrices. *Psychometrika*, 42(2):267–276, 1977. ISSN 00333123. doi: 10.1007/BF02294053.

S. P. Langron and A. J. Collins. Perturbation theory for generalized procrustes analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, 47(2):277–284, 1985. ISSN 00359246. URL http://www.jstor.org/stable/2345571.

F. James Rohlf and Dennis E. Slice. Extensions of the Procustes Methods for the Optimal Superimposition of Landmarks. *Systematic Zoology*, 39(l):40–59, 1990. ISSN 00397989. doi: 10.2307/2992207.

Paul O'Higgins and Ian L. Dryden. Sexual dimorphism in hominoids: further studies of craniofacial shape differences in Pan, Gorilla and Pongo. *Journal of Human Evolution*, 24(3):183–205, mar 1993. ISSN 00472484. doi: 10.1006/jhev.1993.1014. URL http://linkinghub.elsevier.com/retrieve/pii/S0047248483710146.

Fred L Bookstein. BIOMETRICS, BIOMATHEMATICS AND THE MORPHOMETRIC SYNTHESIS. *Bulletin of Marhemancal Biology*, 58(2):313–365, 1996. URL http://www.femininebeauty.info/i/bookstein.morphometrics.pdf.

C. R. Anderson. *Object recognition using statistical shape analysis*. PhD thesis, University of Leeds, 1997.

T F Cootes, C J Taylor, D H Cooper, and J Graham. COMPUTER VISION AND IMAGE UNDERSTANDING Active Shape Models-Their Training and Application. 61(1):38–59, 1995. URL http://www.vis.uni-stuttgart.de/plain/vdl/vdl{_}upload/148{_}24{_}cviu95.pdf.

David H Cooper, Christopher J Taylor, Jim Graham, and Tim F Cootes. Locating Overlapping Flexible Shapes Using Geometrical Constraints. *Bmvc*, pages 185–192, 1991. doi: 10.5244/C.5.24.

Claudia Lindner, Shankar Thiagarajah, J. Mark Wilkinson, Gillian A. Wallis, Tim F. Cootest, Claudia Lindner, Claudia Lindner, and arcOGEN Consortium. Accurate bone

segmentation in 2D radiographs using fully automatic shape model matching based on regression-voting. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, 16(Pt 2):181–189, 2013. ISSN 16113349. doi: 10.1007/978-3-642-40763-5{_}23.

Alejandro Valladares, Germán Manríquez, and Benjamín A. Suárez-Isla. Shell shape variation in populations of Mytilus chilensis (Hupe 1854) from southern Chile: A geometric morphometric approach. *Marine Biology*, 157(12):2731–2738, 2010. ISSN 00253162. doi: 10.1007/s00227-010-1532-3.

Fred L. Bookstein. Shape and the Information in Medical Images: A Decade of the Morphometric Synthesis. *Computer Vision and Image Understanding*, 66(2):97–118, 1997. ISSN 10773142. doi: 10.1006/cviu.1997.0607. URL http://www.sciencedirect.com/science/article/pii/S107731429790607X.

A. Valentin, J.-M. Sévigny, and J.-P. Chanut. Geometric morphometrics reveals body shape differences between sympatric redfish Sebastes mentella, Sebastes fasdatus and their hybrids in the Gulf of St Lawrence. *Journal of Fish Biology*, 60(4): 857–875, 2002. ISSN 1095-8649. doi: 10.1111/j.1095-8649.2002.tb02414.x. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1095-8649.2002.tb02414.x/abstract$\delimiter"026E30F$nhttp://onlinelibrary.wiley.com/store/10.1111/j.1095-8649.2002.tb02414.x/asset/j.1095-8649.2002.tb02414.x.pdf?v=1{&}t=ikikqsr5{&}s=b58ce58f0631236294b8cf92e43424b44fdec01a.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012a. URL http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

Clement Farabet, Camille Couprie, Laurent Najman, and Yann Lecun. Learning hierarchical features for scene labeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1915–1929, 2013. ISSN 01628828. doi: 10.1109/TPAMI.2012.231.

Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 07-12-June-2015, pages 1–9, 2015. ISBN 9781467369640. doi: 10.1109/CVPR.2015.7298594.

Tomáš Mikolov, Anoop Deoras, Daniel Povey, Lukáš Burget, and Jan Černocký. Strategies for training large scale neural network language models. In *2011 IEEE Workshop on Automatic Speech Recognition and Understanding, ASRU 2011, Proceedings*, pages 196–201, 2011. ISBN 9781467303675. doi: 10.1109/ASRU.2011.6163930.

Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R. Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv: 1207.0580*, pages 1–18, 2012. doi: arXiv:1207.0580. URL http://arxiv.org/abs/1207.0580.

Peter Sadowski, Pierre Baldi, and Daniel Whiteson. Searching for Higgs Boson Decay Modes with Deep Learning. *Advances in Nueral Information Processing Systems 27 (Proceedings of NIPS)*, pages 1–9, 2014. ISSN 10495258.

Moritz Helmstaedter, Kevin L Briggman, Srinivas C Turaga, Viren Jain, H Sebastian Seung, and Winfried Denk. Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature*, 500(7461):168–174, 2013. ISSN 0028-0836. doi: 10.1038/nature12346. URL http://www.ncbi.nlm.nih.gov/pubmed/23925239$\delimiter"026E30F$nhttp://www.nature.com/doifinder/10.1038/nature12346.

Michael K K Leung, Hui Yuan Xiong, Leo J. Lee, and Brendan J. Frey. Deep learning of the tissue-regulated splicing code. *Bioinformatics*, 30(12), 2014. ISSN 14602059. doi: 10.1093/bioinformatics/btu277.

Hui Y. Xiong, Babak Alipanahi, Leo J. Lee, Hannes Bretschneider, Daniele Merico, Ryan K. C. Yuen, Yimin Hua, Serge Gueroussov, Hamed S. Najafabadi, Timothy R. Hughes, Quaid Morris, Yoseph Barash, Adrian R. Krainer, Nebojsa Jojic, Stephen W. Scherer, Benjamin J. Blencowe, and Brendan J. Frey. The human splicing

code reveals new insights into the genetic determinants of disease. *Science*, 347 (6218):1254806, 2015. ISSN 1095-9203. doi: 10.1126/science.1254806. URL http://www.sciencemag.org.libproxy.mit.edu/content/347/6218/1254806.full$\delimiter"026E30F$nhttp://www.sciencemag.org/cgi/doi/10.1126/science.1254806$\delimiter"026E30F$nhttp://www.ncbi.nlm.nih.gov/pubmed/25525159$\delimiter"026E30F$nhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4362528.

Kunihiko Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36 (4):193–202, 1980. ISSN 03401200. doi: 10.1007/BF00344251.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2323, 1998. ISSN 00189219. doi: 10.1109/5.726791.

Alexander Toshev and Christian Szegedy. DeepPose: Human Pose Estimation via Deep Neural Networks. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1653–1660, 2014. ISBN 9781479951178. doi: 10.1109/CVPR.2014.214. URL http://arxiv.org/abs/1312.4659.

Alex Krizhevsky, Ilya Sutskever, and Hinton Geoffrey E. ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems 25 (NIPS2012)*, pages 1–9, 2012b. ISSN 10495258. doi: 10.1109/5.726791. URL https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

S. Dieleman, K. W. Willett, and J. Dambre. Rotation-invariant convolutional neural networks for galaxy morphology prediction. *Monthly Notices of the Royal Astronomical Society*, 450(2):1441–1459, 2015a. ISSN 0035-8711. doi: 10.1093/mnras/stv632. URL http://arxiv.org/abs/1503.07077.

Ilya Sutskever. Training Recurrent neural Networks. *PhD thesis*, page 101, 2013.

Dong C. Liu and Jorge Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45(1-3):503–528, aug 1989. ISSN 0025-5610. doi: 10.1007/BF01589116. URL http://link.springer.com/10.1007/BF01589116.

Léon Bottou. Large-Scale Machine Learning with Stochastic Gradient Descent. 2010. URL http://leon.bottou.org/publications/pdf/compstat-2010.pdf.

B.T. Polyak. Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics*, 4(5):1–17, 1964. ISSN 00415553. doi: 10.1016/0041-5553(64)90137-5.

Yurii Nesterov. A Method of Solving A Convex Programming Problem With Convergence rate O(1/k^2), 1983. URL http://www.core.ucl.ac.be/{~}nesterov/Research/Papers/DAN83.pdf.

David E Rumelhart, Geoffrey E Hinton, and R J Williams. Learning Internal Representations by Error Propagation. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, 1:318–362, 1986. ISSN 1-55860-013-2. doi: 10.1016/B978-1-4832-1446-7.50035-2.

Yann Le Cun. Learning Process in an Asymmetric Threshold Network, 1986. URL http://link.springer.com/chapter/10.1007/978-3-642-82657-3{_}24.

M J Kearns and U V Vazirani. *An Introduction to Computational Learning Theory*, volume 8. MIT Press, 1994. ISBN 0262111934. URL http://dl.acm.org/citation.cfm?id=200548.

G Valiant. A Theory of the Learnable. 1984. URL http://web.mit.edu/6.435/www/Valiant84.pdf.

Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014. URL http://jmlr.org/papers/v15/srivastava14a.html.

L Prechelt. Early stopping–But when? In *Neural Networks: Tricks of the Trade*, pages 55–69. 1998a. ISBN 3-540-65311-2.

Lutz Prechelt. Automatic early stopping using cross validation: Quantifying the criteria. *Neural Networks*, 11(4):761–767, 1998b. ISSN 08936080. doi: 10.1016/S0893-6080(98)00010-0.

Lecun Yann. *Efficient backprop*, volume 53. 1998. ISBN 9788578110796. doi: 10.1017/CBO9781107415324.004.

Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 9:249–256, 2010. ISSN 15324435. doi: 10.1.1.207.2059. URL http://machinelearning.wustl.edu/mlpapers/paper{_}files/AISTATS2010{_}GlorotB10.pdf.

Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation Applied to Handwritten Zip Code Recognition, 1989. ISSN 0899-7667.

Chris Donahue, Zachary C. Lipton, and Julian McAuley. Dance Dance Convolution. mar 2017. URL http://arxiv.org/abs/1703.06891.

Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359, 2008. ISSN 10773142. doi: 10.1016/j.cviu.2007.09.014.

Adam Coates, Ann Arbor, and Andrew Y Ng. An Analysis of Single-Layer Networks in Unsupervised Feature Learning. *Aistats 2011*, pages 215–223, 2011. ISSN <null>. doi: 10.1109/ICDAR.2011.95.

Kevin Jarrett, Koray Kavukcuoglu, Marc'Aurelio Ranzato, and Yann LeCun. What is the best multi-stage architecture for object recognition? In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2146–2153, 2009. ISBN 9781424444205. doi: 10.1109/ICCV.2009.5459469.

Andrew M Saxe, Pang Wei Koh, Zhenghao Chen, Maneesh Bhand, Bipin Suresh, and Andrew Y Ng. On Random Weights and Unsupervised Feature Learning. *Learning*,

(2009):1–9, 2011. URL http://ai.stanford.edu/{~}ang/papers/nipsdlufl10-RandomWeights.pdf.

Nicolas Pinto, Zak Stone, Todd Zickler, and David Cox. Scaling up biologically-inspired computer vision: A case study in unconstrained face recognition on facebook. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2011. ISBN 9781457705298. doi: 10.1109/CVPRW.2011.5981788.

David Cox and Nicolas Pinto. Beyond simple features: A large-scale feature search approach to unconstrained face recognition. In *2011 IEEE International Conference on Automatic Face and Gesture Recognition and Workshops, FG 2011*, pages 8–15, 2011. ISBN 9781424491407. doi: 10.1109/FG.2011.5771385.

Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Y Ng. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. *Proceedings of the 26th Annual International Conference on Machine Learning ICML 09*, 2008:1–8, 2009. ISSN 02643294. doi: 10.1145/1553374.1553453. URL http://portal.acm.org/citation.cfm?doid=1553374.1553453.

Y-Lan Boureau, Jean Ponce, and Yann Lecun. A Theoretical Analysis of Feature Pooling in Visual Recognition. *Proceedings of the 27th International Conference on Machine Learning (2010)*, pages 111–118, 2010. doi: citeulike-article-id:8496352. URL http://www.ece.duke.edu/{~}lcarin/icml2010b.pdf.

Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. dec 2013. URL http://arxiv.org/abs/1312.6034.

R Azaria, N Adler, R Silfen, D Regev, and D J Hauben. Morphometry of the adult human earlobe: a study of 547 subjects and clinical application. *Plastic and reconstructive surgery*, 111(7):2398–402; discussion 2403–4, jun 2003. ISSN 0032-1052. doi: 10.1097/01.PRS.0000060995.99380.DE. URL http://www.ncbi.nlm.nih.gov/pubmed/12794488.

Chiarella Sforza, Gaia Grandi, Miriam Binelli, Davide G Tommasi, Riccardo Rosati, and Virgilio F Ferrario. Age- and sex-related changes in the normal human ear. *Forensic*

science international, 187(1-3):110.e1–7, may 2009. ISSN 1872-6283. doi: 10.1016/j.forsciint.2009.02.019. URL http://www.ncbi.nlm.nih.gov/pubmed/19356871.

T.E. Oliphant. Python for Scientific Computing. *Computing in Science & Engineering*, 9, 2007. ISSN 1521-9615. doi: 10.1109/MCSE.2007.58.

Stéfan van der Walt, S. Chris Colbert, and Gaël Varoquaux. The numpy array: a structure for efficient numerical computation. *CoRR*, abs/1102.1523, 2011. URL http://arxiv.org/abs/1102.1523.

Sander Dieleman, Jan Schlüter, Colin Raffel, Eben Olson, Søren Kaae Sø nderby, Daniel Nouri, Daniel Maturana, Martin Thoma, Eric Battenberg, Jack Kelly, Jeffrey De Fauw, Michael Heilman, diogo149, Brian McFee, Hendrik Weideman, takacsg84, peterderivaz, Jon, instagibbs, Dr. Kashif Rasul, CongLiu, Britefury, and Jonas Degrave. Lasagne: First release., Agu 2015b. URL http://dx.doi.org/10.5281/zenodo.27878.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Pierre Geurts, Damien Ernst, and Louis Wehenkel. Extremely randomized trees. *Machine learning*, 63(1):3–42, 2006.

Alfred V. Iannarelli. *Ear Identification*, volume 1. Paramont Publishing Company, 1989. ISBN 0962317802. URL https://books.google.com/books?id=jgPkAAAACAAJ{&}pgis=1.

H. Chen and B. Bhanu. Shape Model-Based 3D Ear Detection from Side Face Range Images. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, volume 3, pages 122–122. IEEE, 2005. ISBN 0-7695-2372-2. doi: 10.1109/CVPR.2005.525. URL http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=1565434.

H. Chen and B. Bhanu. Contour matching for 3D ear recognition. In *Proceedings - Seventh*

IEEE Workshop on Applications of Computer Vision, WACV 2005, pages 123–128, 2007. ISBN 0769522718. doi: 10.1109/ACVMOT.2005.38.

Sepehr Attarchi, Karim Faez, and Aref Rafiei. *Advanced Concepts for Intelligent Vision Systems*, volume 5259 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, Berlin, Heidelberg, oct 2008. ISBN 978-3-540-88457-6. doi: 10.1007/978-3-540-88458-3. URL http://dl.acm.org/citation.cfm?id=1462298.1462399.

Saeeduddin Ansari and Phalguni Gupta. Localization of ear using outer helix curve of the ear. *International conferenceon computing:the theoryand applications*, pages 688–692, 2007. doi: 10.1109/ICCTA.2007.82.

Surya Prakash and Phalguni Gupta. An efficient ear localization technique. *Image and Vision Computing*, 30(1):38–50, 2012. ISSN 02628856. doi: 10.1016/j.imavis.2011.11.005.

Ping Yan and Kevin W Bowyer. Biometric recognition using 3D ear shape. *IEEE transactions on pattern analysis and machine intelligence*, 29(8):1297–308, 2007. ISSN 0162-8828. doi: 10.1109/TPAMI.2007.1067. URL http://www.ncbi.nlm.nih.gov/pubmed/17568136.

Ayman Abaza, Christina Hebert, and Mary Ann F Harrison. Fast learning ear detection for real-time surveillance. *IEEE 4th International Conference on Biometrics: Theory, Applications and Systems, BTAS 2010*, 2010. doi: 10.1109/BTAS.2010.5634486.

S. M S Islam, M. Bennamoun, and R. Davies. Fast and fully automatic ear detection using cascaded adaboost. *2008 IEEE Workshop on Applications of Computer Vision, WACV*, 2008. ISSN 1550-5790. doi: 10.1109/WACV.2008.4544023.

Anika Pflug, Andreas Winterstein, and Christoph Busch. Robust localization of ears by feature level fusion and context information. In *Biometrics (ICB), 2013 International Conference on*, pages 1–8. IEEE, 2013.

Li Yuan, Wei Liu, and Yang Li. Non-negative dictionary based sparse representation classification for ear recognition with occlusion. *Neurocomputing*, 171:540–550, 2016.

Ajay Kumar and Tak-Shing T Chan. Robust ear identification using sparse representation of local texture descriptors. *Pattern recognition*, 46(1):73–85, 2013.

Thomas G. Dietterich. An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees. *Machine Learning*, 40: 139–157, 2000. ISSN 0885-6125. doi: 10.1023/A:1007607513941. URL http://en.scientificcommons.org/42637098$\delimiter"026E30F$nuuid/7906280C-AEF8-405A-9A94-6BAA1DDAED1E.

Leo Breiman. Random Forests. *Machine Learning*, 45(1):5–32, 2001. ISSN 08856125. doi: 10.1023/A:1010933404324. URL http://link.springer.com/10.1023/A:1010933404324.

Louis Wehenkel, Damien Ernst, and Pierre Geurts. Ensembles of extremely randomized trees and some generic applications. In *Robust Methods for Power System State Estimation and Load Forecasting*, 2006.

V. Bruce, A. M. Burton, E. Hanna, P. Healey, O. Mason, A. Coombes, R. Fright, and A. Linney. Sex discrimination: how do we tell the difference between male and female faces? *Perception*, 22(2):131–152, 1993. ISSN 03010066. doi: 10.1068/p220131.

Choon Boon Ng, Yong Haur Tay, and Bok Min Goi. A review of facial gender recognition. *Pattern Analysis and Applications*, 18(4):739–755, 2015. ISSN 14337541. doi: 10.1007/s10044-015-0499-6. URL "http://dx.doi.org/10.1007/s10044-015-0499-6.

Erno Mäkinen and Roope Raisamo. An experimental comparison of gender classification methods. *Pattern Recognition Letters*, 29(10):1544–1556, 2008. ISSN 01678655. doi: 10.1016/j.patrec.2008.03.016.

Tracey Caldwell. Vending machines recommend based on face recognition. *Biometric Technology Today*, 2011(1):12, 2011. ISSN 0969-4765. doi: https://doi.org/10.1016/S0969-4765(11)70018-2. URL http://www.sciencedirect.com/science/article/pii/S0969476511700182.

Srinivas Gutta, Jeffrey R.J. Huang, P. Jonathon, and Harry Wechsler. Mixture of experts

for classification of gender, ethnic origin, and pose of human faces. *IEEE Transactions on Neural Networks*, 11(4):948–960, 2000. ISSN 10459227. doi: 10.1109/72.857774.

Amit Jain, Jeffrey Huang, and Shiaofen Fang. Gender identification using frontal facial images. In *IEEE International Conference on Multimedia and Expo, ICME 2005*, volume 2005, pages 1082–1085, 2005. ISBN 0780393325. doi: 10.1109/ICME.2005.1521613.

Fok Hing Chi Tivive and Abdesselam Bouzerdoum. A gender recognition system using shunting inhibitory convolutional neural network for gender classification. In *Proceedings - International Conference on Pattern Recognition*, volume 4, pages 421–424, 2006. ISBN 0-7695-2521-0. doi: 10.1109/ICPR.2006.173.

Luís A. Alexandre. Gender recognition: A multiscale decision fusion approach. *Pattern Recognition Letters*, 31(11):1422–1427, 2010. ISSN 01678655. doi: 10.1016/j.patrec.2010.02.010.

M. Hussain, I. Ullah, H.A. Aboalsamh, G. Muhammad, G. Bebis, and A.M. Mirza. Gender recognition from face images with dyadic wavelet transform and local binary pattern. *International Journal on Artificial Intelligence Tools*, 22(6), 2013. ISSN 17936349 02182130. doi: 10.1142/S021821301360018X.

Cha Zhang and Zhengyou Zhang. A Survey of Recent Advances in Face Detection. *Microsoft Research*, (June):17, 2010. doi: 10.1.1.167.5270. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.167.5270{&}amp;rep=rep1{&}amp;type=pdf.

Chiraz BenAbdelkader and Paul Griffin. A Local Region-based Approach to Gender Classification From Face Images. *Conference on Computer Vision and Pattern Recognition*, 3:52–52, 2005. ISSN 1063-6919. doi: 10.1109/CVPR.2005.388.

Bing Li, Xiao Chen Lian, and Bao Liang Lu. Gender classification by combining clothing, hair and facial component classifiers. *Neurocomputing*, 76(1):18–27, 2012. ISSN 09252312. doi: 10.1016/j.neucom.2011.01.028.

Roberto Brunelli and Tomaso Poggio. Face Recognition: Features versus Templates. *IEEE*

Transactions on Pattern Analysis and Machine Intelligence, 15(10):1042–1052, 1993. ISSN 01628828. doi: 10.1109/34.254061.

Jean Marc Fellous. Gender discrimination and prediction on the basis of facial metric information. *Vision Research*, 37(14):1961–1973, 1997. ISSN 00426989. doi: 10.1016/S0042-6989(97)00010-2.

Shumeet Baluja and Henry A. Rowley. Boosting sex identification performance, 2007. ISSN 09205691.

H. Abdi, D. Valentin, B. Edelman, and A. J. O'Toole. More about the difference between men and women: evidence from linear neural networks and the principal-component approach. *Perception*, 24(5):539–562, 1995. ISSN 03010066. doi: 10.1068/p240539.

Timo Ojala, Matti Pietikäinen, and Topi Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002. ISSN 01628828. doi: 10.1109/TPAMI.2002.1017623.

Francis Galton. Numeralised profiles for classification and recognition. *Nature*, 83:127–130, 1910.

Leon D Harmon and Willard F Hunt. Automatic recognition of human face profiles. *Computer Graphics and Image Processing*, 6(2):135–156, 1977.

Bir Bhanu and Xiaoli Zhou. Face recognition from face profile using dynamic time warping. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 4, pages 499–502. IEEE, 2004.

Xiaoli Zhou and Bir Bhanu. Human recognition based on face profiles in video. In *Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, pages 15–15. IEEE, 2005.

Ioannis A Kakadiaris, H Abdelmunim, W Yang, and Theoharis Theoharis. Profile-based face recognition. In *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*, pages 1–8. IEEE, 2008.

Zdravko Lipoščak and Sven Loncaric. A scale-space approach to face recognition from profiles. In *Computer Analysis of Images and Patterns*, pages 243–250. Springer, 1999.

J. Velemínská, L. Bigoni, V. Krajíček, J. Borský, D. Šmahelová, V. Cagáňová, and M. Peterka. Surface facial modelling and allometry in relation to sexual dimorphism. *HOMO-Journal of Comparative Human Biology*, 63(2):81–93, 2012. ISSN 0018442X. doi: 10.1016/j.jchb.2012.02.002.

Matthew J. Kesterke, Zachary D. Raffensperger, Carrie L. Heike, Michael L. Cunningham, Jacqueline T. Hecht, Chung How Kau, Nichole L. Nidey, Lina M. Moreno, George L. Wehby, Mary L. Marazita, and Seth M. Weinberg. Using the 3D Facial Norms Database to investigate craniofacial sexual dimorphism in healthy children, adolescents, and adults. *Biology of Sex Differences*, 7(1):23, 2016. ISSN 2042-6410. doi: 10.1186/s13293-016-0076-8. URL http://bsd.biomedcentral.com/articles/10.1186/s13293-016-0076-8.

Hyunsook Han and Yunja Nam. Automatic body landmark identification for various body figures. *International Journal of Industrial Ergonomics*, 41(6):592–606, 2011. ISSN 01698141. doi: 10.1016/j.ergon.2011.07.002.

Byoung-Keon Park, Julie C. Lumeng, Carey N. Lumeng, Sheila M. Ebert, and Matthew P. Reed. Child body shape measurement using depth cameras and a statistical body shape model. *Ergonomics*, 58(2):301–9, 2015. ISSN 1366-5847. doi: 10.1080/00140139.2014.965754. URL http://www.ncbi.nlm.nih.gov/pubmed/25323820.

Steven Paquette. 3D scanning in apparel design and human engineering. *IEEE Computer Graphics and Applications*, 16(5):11–15, 1996. ISSN 02721716. doi: 10.1109/38.536269.

Nicola D'Apuzzo. 3D body scanning technology for fashion and apparel industry. *Spine*, 6491:64910O–64910O–12, 2007. ISSN 0277786X. doi: 10.1117/12.703785. URL http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1298458.

Zouhour Ben Azouz, Marc Rioux, Chang Shu, and Richard Lepage. Characterizing human

shape variation using 3D anthropometric data. *The Visual Computer*, 22(5):302–314, 2006. ISSN 0178-2789. doi: 10.1007/s00371-006-0006-6.

Jonathan C K Wells, Janet Stocks, Rachel Bonner, Emma Raywood, Sarah Legg, Simon Lee, Philip Treleaven, and Sooky Lum. Acceptability, Precision and Accuracy of 3D Photonic Scanning for Measurement of Body Shape in a Multi-Ethnic Sample of Children Aged 5-11 Years: The SLIC Study. *PLOS ONE*, 10(4):1–15, 2015. doi: 10.1371/journal.pone.0124193. URL https://doi.org/10.1371/journal.pone.0124193.

J C K Wells, A Ruto, and P Treleaven. Whole-body three-dimensional photonic scanning: a new technique for obesity research and clinical practice. *International Journal of Obesity*, 32(2):232–238, 2008. ISSN 0307-0565. doi: 10.1038/sj.ijo.0803727. URL http://www.nature.com/doifinder/10.1038/sj.ijo.0803727.

Philip Treleaven and Jonathan Wells. 3D body scanning and healthcare applications. *Computer*, 40(7):28–34, 2007. ISSN 00189162. doi: 10.1109/MC.2007.225.

David A Hirshberg, Matthew Loper, Eric Rachlin, Aggeliki Tsoli, Alexander Weiss, Brian Corner, and Michael J Black. Evaluating the Automated Alignment of 3D Human Body Scans. 2011. URL http://cs.brown.edu/{~}aweiss/papers/hirshberg3dbst11.pdf.

N Halko, P G Martinsson, and J A Tropp. Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions. *SIAM Review*, 53(2):217—-288, 2011. ISSN 0036-1445. doi: 10.1137/090771806. URL http://dx.doi.org/10.1137/090771806.

Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. *arXiv preprint arXiv:1612.00593*, 2016.

Callum Faris. Scott-Brown's Otorhinolaryngology, Head and Neck Surgery, 7th edn, oct 2011. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3604940/.

Timothy C Cox, Esra D Camci, Siddharth Vora, Daniela V Luquetti, and Eric E Turner. The genetics of auricular development and malformation: new findings in mo-

del systems driving future directions for microtia research. *European journal of medical genetics*, 57(8):394–401, aug 2014. ISSN 1878-0849. doi: 10.1016/j.ejmg.2014.05.003. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4143470{&}tool=pmcentrez{&}rendertype=abstract.

Bowen Baker, Otkrist Gupta, Nikhil Naik, and Ramesh Raskar. Designing Neural Network Architectures using Reinforcement Learning. *arXiv preprint*, pages 1–16, 2016. URL http://arxiv.org/abs/1611.02167.

Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando De Freitas. Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175, 2016. ISSN 00189219. doi: 10.1109/JPROC.2015.2494218.

Nicolas Pinto, David Doukhan, James J. DiCarlo, and David D. Cox. A high-throughput screening approach to discovering good forms of biologically inspired visual representation. *PLoS Computational Biology*, 5(11), 2009. ISSN 1553734X. doi: 10.1371/journal.pcbi.1000579.

Jorge A. Gómez-Valdés, Mirsha Quinto-Sánchez, Antinea Menéndez Garmendia, Jana Veleminska, Gabriela Sánchez-Mejorada, and Jaroslav Bruzek. Comparison of methods to determine sex by evaluating the greater sciatic notch: Visual, angular and geometric morphometrics. *Forensic Science International*, 221(1-3), 2012. ISSN 03790738. doi: 10.1016/j.forsciint.2012.04.027.

Soledad De Azevedo, Ariadna Nocera, Carolina Paschetta, Lucía Castillo, Marina González, and Rolando González-José. Evaluating microevolutionary models for the early settlement of the New World: The importance of recurrent gene flow with Asia. *American Journal of Physical Anthropology*, 146(4):539–552, 2011. ISSN 00029483. doi: 10.1002/ajpa.21564.

Theano Development Team. Theano: A Python framework for fast computation of mathematical expressions. *arXiv e-prints*, abs/1605.02688, May 2016. URL http://arxiv.org/abs/1605.02688.

Daniel Nouri. nolearn: scikit-learn compatible neural network library. 2014.

Matthew D. Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. *CoRR*, abs/1311.2901, 2013. URL http://arxiv.org/abs/1311.2901.