

TESIS DE DOCTORADO
EN INGENIERÍA
MENCIÓN EN PROCESAMIENTO DE
SEÑALES E IMÁGENES

Título:

“Algoritmos de alta eficiencia para el fenotipado automático: aplicaciones en reconstrucción morfométrica 3D, medicina y ciencia forense”

Autor: Navarro J. Pablo

Director de Tesis: Delrieux Claudio y González-José Rolando

Buenos Aires - 2021

Prefacio

Esta Tesis se presenta como parte de los requisitos para optar al grado Académico de Doctor en Ingeniería – Mención en Procesamiento de Señales e Imágenes de la UTN Facultad Regional Buenos Aires 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 2016 y el 01 de Abril de 2021, bajo la dirección del Dr. Claudio Delrieux y Co-dirección del Dr. Rolando González-José.

A mis padres, Carlos y Alicia.

Agradecimientos

En primer lugar quiero agradecer a Claudio y Rolo, quienes con sus conocimientos y apoyo me guiaron a lo largo de todo el doctorado. Especial mención a Celia, estoy convencido de que este trabajo no hubiese llegado a ser lo que es sin su apoyo incondicional. También quiero agradecer al GIBEH por brindarme todos los recursos y herramientas que fueron necesarios para llevar a cabo el proceso de investigación. Por último, quiero agradecer a mis amigos (Leo, Bruno, Emiliano, Samuel, Nahuel, Diego y Virginia) y a mi familia, por apoyarme en todo momento. En especial, quiero hacer mención a Paula, mi compañera de vida, que siempre estuvo ahí para darme palabras de apoyo.

Muchas gracias a todos.

Abreviaturas

CNN Convolutional Neural Networks

CV Validación Cruzada

DL Deep Learning

ERT Extremely Randomized Trees

ETC Extra Trees Classifier

IA Inteligencia Artificial

ICC Índice Cintura Cadera

ICT Índice de Cintura Talla

IMC Índice de Masa Corporal

ML Machine Learning

MVCNN Redes Neuronales Convolucionales de Múltiples Vistas

OMS Organización Mundial de la Salud

PCA Análisis de Componentes Principales

SFM Structure from Motion

SGD Stochastic Gradient Descent

SVC Support Vector Classifier

SVM Support Vector Machine

VC Visión Computacional

Índice general

1. Introducción	16
1.1. Contexto	18
1.2. Objetivos	20
1.3. Contribuciones	20
1.3.1. Publicaciones referidas a esta tesis en revistas internacionales	21
1.3.2. Actas en conferencias nacionales e internacionales	22
1.3.3. Otras publicaciones y presentaciones en congresos y revistas.	24
1.4. Aplicaciones	26
1.5. Estructura de la Tesis	26
2. Marco Teórico	27
2.1. Visión Computacional	27
2.2. Adquisición de datos	29
2.2.1. Imágenes 2D	29
2.2.2. Nube de Puntos y Mallas 3D	30
2.2.3. Técnicas de escaneo 3D	32

2.2.4. Fotogrametría	36
2.3. Pre-procesamiento	38
2.4. Inteligencia Artificial	41
2.4.1. Machine Learning	42
2.4.2. Deep Learning	49
3. Colecta de datos	55
3.1. Proyecto Patagonia 3D Lab	55
3.2. Campaña de Voluntarios	57
3.3. Protocolo de recolección de datos	58
3.4. Datos Artificiales	60
3.5. Configuración de landmarks	60
3.6. Proyecto Raíces	62
4. Reconocimiento automático de sexo biológico utilizando capturas 3D	63
4.1. Introducción	63
4.2. Materiales y Métodos	65
4.2.1. Descripción de la muestra	65
4.2.2. Extracción de características y entrenamiento	65
4.3. Experimentos	68
4.3.1. Selección y optimización de hiper parámetros	68
4.3.2. Selección de datos de Entrenamiento.	69
4.3.3. Tiempo de entrenamiento y predicción	71

4.4. Resultados	71
4.5. Conclusiones	72
5. Extracción automática de características de la forma corporal para el estudio de obesidad	74
5.1. Introducción	74
5.2. Método	76
5.2.1. Descripción de la muestra	76
5.2.2. Extracción automática de características de la forma corporal	77
5.3. Resultados	79
5.3.1. Medición de error y adquisición de características	81
5.4. Conclusiones	82
6. Descriptores Multi-vista basado en aprendizaje profundo	84
6.1. Introducción	84
6.2. Materiales y Métodos	85
6.2.1. Entrenamiento de la red	87
6.2.2. Construcción del conjunto de datos	91
6.3. Diseño experimental	94
6.3.1. Configuración de la red de categoría específica	94
6.3.2. Configuración de la red multicategoría	97
6.4. Resultados	98
6.4.1. Imágenes dibujadas a mano	98

6.4.2. Estudio de usuario	99
6.4.3. Pruebas con figuras humanas	101
6.5. Conclusión y trabajo futuro	102
7. Conclusiones Generales	104
7.1. Conclusiones	104
7.2. Trabajo Futuro	106
Anexos	109
Anexo A. Cartilla de Información	110
Anexo B. Comité de ética	111
Bibliografía	112

Índice de figuras

2.1. Esquema de las etapas que componen la Visión Computacional (VC).	29
2.2. Ejemplo de modelos creados a partir del software Pose 3D, utilizando “Apparent Ridges algorithm” (Judd et al., 2007) para generar una imagen con estilo de dibujo sobre el modelo 3D.	30
2.3. Nube de puntos de la Catedral de Murcia. Sensor Leica RTC360 (Dronica).	32
2.4. Malla 3D. Sensor Structure. 99,000 puntos (aprox 199,000 triángulos).	33
2.5. <i>Structure Sensor</i> . Escáner basado en triangulación láser (ItSeez3D).	34
2.6. Escaneo de rostro con <i>GoScan20</i> . Escáner basado en luz estructurada. Foto capturada durante la colecta de datos Raíces en Puerto Madryn.	35
2.7. P y Q visto de dos cámaras C_1, C_2 . Los puntos 2D $P'1, P'2, Q'1$ y $Q'2$ son los puntos de interés visto en dos imágenes. P y Q es proyección 3D calculada.	37
2.8. Esquema de cómo se compone la Inteligencia Artificial (IA).	41
2.9. Clasificación, Regresión y Clusterización.	44
2.10. Funcionamiento del algoritmo de Support Vector Machine (SVM).	46
2.11. Visualización en forma de Árbol. Diferentes valores del índice cintura-cadera junto con el sexo biológico determinar la probabilidad de tener o no riesgo cardiovascular (Elsayed et al., 2008).	47

2.12. Comportamiento de la métrica entropía. El árbol de la izquierda muestra un caso ideal, donde las muestras generan conjuntos homogéneos. El árbol de la derecha es un caso con <i>incertidumbre</i> = 1, la selección de la clase es igual a seleccionar una clase al azar.	48
2.13. Arquitectura de capas en redes neuronales.	52
3.1. Campaña de convocatoria en redes sociales.	57
3.2. Protocolo de captura de datos en 360. Vista superior.	59
3.3. Protocolo de medición.	60
3.4. Configuración de <i>landmarks</i> corporales 3D sobre un modelo artificial.	61
4.1. Modelos artificiales del dataset (Navarro et al., 2018).	66
4.2. Modelos reales del dataset (Navarro et al., 2018).	67
4.3. <i>F1-score</i> usando diferentes cantidades de componentes para dos clasificadores (Navarro et al., 2018).	68
4.4. Matriz de confusión obtenida con el conjunto de prueba (Navarro et al., 2018).	71
4.5. Matriz de confusión obtenida agregando el conjunto de datos reales (Navarro et al., 2018).	72
5.1. Ejemplo de imagen 3D del conjunto de datos, junto con la configuración de <i>landmarks</i> usada para la extracción de las características de la forma corporal. Los <i>landmarks</i> 1 – 2 corresponde a la articulación del hombro (glenohumeral) y los <i>landmarks</i> 3 – 4 a los punto donde articula la cabeza del fémur en la cadera (Navarro et al., 2019a)	75

5.2. Localización de los <i>semi-landmarks</i> s_0 y s_1 en Naranja. Los 500 puntos ajustados por el criterio de mínimo <i>BendingEnergy</i> en verde. El área abdominal descrita en distintas vistas (2D, 3D y vista transversal.) generados con 250 segmentos. (Navarro et al., 2019a)	78
5.3. Análisis de componentes principales. Gráfico de dispersión y variación de forma y distribución de la grasa abdominal con el descriptor propuesto. Elipses coloreadas basadas en Índice de Masa Corporal (IMC), Índice de Cintura Talla (ICT), Índice Cintura Cadera (ICC) estados nutricionales. Los cambios de forma representan el cambio de forma promedio observado en los extremos de los ejes. (Navarro et al., 2019a)	80
6.1. Dibujo de humanos con diferentes formas corporales.	85
6.2. Vista general del <i>pipeline</i> propuesto. Izquierda: Dado un conjunto de datos precompilados de modelos 3D con datos de correspondencias, se generan automáticamente dibujos a diferentes escalas y posiciones utilizando un motor de renderizado no fotorrealista de última generación. Derecha: se utilizan estos dibujos como entradas para la red neuronal convolucional a fin de aprender los descriptores de múltiples vistas. Además se observa cómo los puntos coincidentes se asignan juntos independientemente del ángulo de proyección. Se utiliza una arquitectura de múltiples vistas conjuntamente con capas lineales para reducir el tamaño del descriptor y las capas de <i>MaxPooling</i> que agregan información importante a las vistas ampliadas. (Navarro et al., 2019b)	86
6.3. Visualización de imágenes del conjunto de datos. Dado un par de modelos 3D densamente registrados (izquierda), se realizaron renderizaciones desde tres ángulos (columnas) predefinidos y tres <i>zooms</i> (filas) diferentes. Cada imagen se centra en el punto de interés correspondiente al punto coincidente (indicado con un contorno de color).	91

6.4. (a) Correspondencias por pares en cuatro categorías de objetos. Los puntos que se corresponden tienen el mismo color. (b) Visualización de correspondencias densas utilizando el descriptor aquí presentado en diferentes categorías. Los píxeles correspondientes se indican en colores similares. Se utilizó una máscara binaria para excluir píxeles fuera de la región del boceto. A pesar de las diferencias extremas en la geometría y las posiciones de la cámara, el descriptor logra registrar correctamente los bocetos. (Navarro et al., 2019b) 95

6.5. Izquierda: Gráficos de CMC para los descriptores evaluados en los conjuntos de datos de prueba para cada categoría. El eje Y representa el porcentaje de coincidencias recuperadas debajo de la posición indicada en X. Derecha: curvas de precisión de correspondencia donde el eje X muestra un error de distancia euclidiana normalizado. El eje Y representa el porcentaje de coincidencias recuperadas por debajo del margen de error indicado en X. (Navarro et al., 2019b) 96

6.6. Mapa de calor con las correspondencias anotadas por los usuarios. En cada par, la imagen de la izquierda muestra los puntos de referencia que debían ser correspondidos por el usuario, la imagen de la derecha representa un mapa de calor con las coincidencia de los usuarios más fuertes. Los símbolos indican las correspondencias para cada método evaluada. 100

6.7. Resultados preliminares de la red en cuerpos humanos. Cada fila determina la correspondencia densa entre tres imágenes. Los puntos que se corresponden cuentan con el mismo color. 102

B.1. Aval del comité de ética. 111

Índice de tablas

2.1. Descripción de los <i>dataset</i> 2D más conocidos.	31
3.1. Definición anatómica de los 15 <i>landmarks</i>	61
4.1. Support Vector Classifier (SVC), promedio por corrida (Navarro et al., 2018).	70
4.2. Extra Trees Classifier (ETC), promedio por corrida (Navarro et al., 2018).	70
4.3. Tiempos de entrenamiento y predicción (Promedio de 1000 corridas) (Navarro et al., 2018).	71
5.1. Fila Superior: promedio y desviación estándar de las medidas lineales. Filas Inferiores: Tamaños de la muestra por sexo y estado nutricional. Índice de Masa Corporal (IMC) puntos de corte de acuerdo a la Organización Mundial de la Salud: <i>Obeso</i> ≥ 30 , <i>Sobrepeso</i> ≥ 25 <i>Saludable</i> $\geq 18,50 - 24,99$ WHO (2018). Índice de Cintura Talla (ICT) con punto de corte en 0.5 para ambos géneros. Browning et al. (2010) Índice Cintura Cadera (ICC), valores sobre 0,85 y 0,9 son considerados no saludables para mujeres y hombres, respectivamente. Navarro et al. (2019a).	77
5.2. Regresión Lineal PC1 y PC2 contra índices clásicos (IMC, ICT, ICC) (Navarro et al., 2019a)	82

6.1. Resultados de la métrica *Cumulative Match Characteristic (CMC)* medidos en 4 métodos. CMC es reportado sobre 100 correspondencias. Aquí se presentan los resultados para la red entrenada con cada categoría y para la red multicategoría. (Navarro et al., 2019b) 98

6.2. Estudio de métricas cuantitativas. Los valores en la primera fila representan la probabilidad de que cada descriptor probado forme parte de las estadísticas del usuario, modeladas con una distribución gaussiana en 2D sobre los datos del estudio. En la segunda fila se detalla los valores de Error cuadrático medio normalizados (RMSE) resultado del rendimiento de los descriptores con respecto a los datos del usuario. 100

Resumen

El uso de escáneres 3D para capturar información sobre la forma del cuerpo humano está creciendo rápidamente y no solo se limita a estudios ergonómicos o de diseño, sino que también resulta de interés para aplicaciones relacionadas con la salud. Para ello, contar con herramientas que nos permitan obtener automáticamente representaciones de características relevantes es un paso crucial. Éstas representan una característica esencial durante la medición e interpretación de los datos, tanto para realizar diversos análisis (por ejemplo el somatotipo o estudio de porcentaje y distribución del tejido adiposo), como para el diagnóstico apropiado de condiciones médicas (evaluación del sobrepeso, detección de malformaciones, etc.). La presente tesis se centra en el estudio de la forma corporal, y cómo utilizando técnicas de *Machine Learning* y *Deep Learning*, se pueden encontrar de manera automática mejores representaciones geométricas a fenotipos estudiados en el campo de la antropología. Se presenta el desarrollo y validación de distintos tipos de descriptores junto con el procesamiento para extraerlos y utilizarlos en la práctica. Los resultados indican que los indicadores de forma no sólo predicen bien los marcadores clásicos, sino que permiten la aplicación de más indicadores que reflejan con mayor precisión el fenotipo en estudio.

Capítulo 1

Introducción

Cada ser vivo cuenta con información hereditaria (genotipo) que da origen, junto con los factores ambientales, a las características particulares que definen sus propiedades observables (como la morfología, el color y forma del pelo, ojos, etc.). Esta expresión del genotipo en función del ambiente se denomina fenotipo, algunas de estas expresiones pueden observarse físicamente en un individuo (por ejemplo el color de ojos o la forma de la boca, nariz, etc.) o pueden manifestarse de forma no visible, como la predisposición a tener ciertas enfermedades.

Este conjunto de genotipos tiene una compleja correspondencia con un grupo de fenotipos. A esta relación se la se conoce como el mapa genotipo-fenotipo ([Pigliucci, 2010](#)). El genotipo es el factor de influencia más grande en el desarrollo del fenotipo, pero no es el único. Un ejemplo claro son los gemelos humanos monocigóticos. Sus genomas son idénticos, pero no tienen el mismo fenotipo.

En humanos, la herencia de muchos fenotipos asociados a la acumulación de grasa, la masa libre de grasa y otras características físicas (como el peso y la altura) han sido estimadas en distintas poblaciones y diferentes grupos de edad. Estas observaciones fueron consistentes con la contribución de factores genéticos y la variación de estos rasgos físicos ([Loos and Bouchard, 2003](#)). Las herencias fenotípicas pueden predisponer a ciertas poblaciones a distintas patologías como la obesidad y el sobrepeso. Estudiar estos fenotipos dentro de una población específica puede ser de gran importancia para la prevención,

diagnóstico y tratamiento de diferentes afecciones. Estos estudios son una tarea compleja que requiere de trabajo manual, los equipos de investigación relacionados no acuden a la utilización masiva de todas las variables cuantitativas de rasgos fenotípicos, ya que derivan en un proceso lento y que origina grandes errores (Segev et al., 2010, Kamoen et al., 2001).

En la actualidad, debido al avance de la tecnología, se puede encontrar distintos tipos de dispositivos de captura de datos para estudiar el fenotipo humano. La que se utilice depende de las características que se quieran relevar. Los tipos más usados abarcan imágenes 2D/3D, vídeos o datos antropométricos (por ejemplo: medidas, peso, balanza de impedancia, etc.). El desarrollo de nuevos dispositivos de captura permiten recolectar grandes volúmenes de datos, así como aumentar el tamaño cruzando distintos tipos de información logrando grandes biobancos. Con estas bases de datos disponibles, es necesario pensar en el desarrollo de nuevas técnicas computacionales que permitan extraer automáticamente información relevante. Esta información, sin lugar a dudas, permite nuevas investigaciones que van desde la práctica clínica hasta estudios poblacionales. Esto a su vez, contando con otras fuentes de información ambiental, y aplicando técnicas de grandes datos (Big Data) tiene el potencial de extraer y analizar información demográfica valiosa que permitiría, entre otras cosas, generar indicadores relevantes para la determinación de políticas públicas. Actualmente existen múltiples trabajos donde técnicas innovadoras de *Machine Learning (ML)* y *Deep Learning (DL)* son aplicadas a problemáticas en salud, en las cuales podemos destacar modelos para secuenciación del genoma (Zhou et al., 2018, Alipanahi et al., 2015), detección de tumores en tomografías (Jin et al., 2018, Dong et al., 2017) y predicción del estado de un paciente mediante su historial clínico (Miotto et al., 2016).

La presente tesis se centra en el estudio de un fenotipo en particular: la forma corporal (*Body Shape*), y cómo utilizando técnicas de ML y DL, se pueden encontrar de manera automática mejores representaciones geométricas a fenotipos estudiados en el campo de la antropología. Se presenta el desarrollo y validación de distintos tipos de descriptores junto con el procesamiento para extraerlos y utilizarlos en la práctica. Dentro de esta idea, se propone la utilización de estos descriptores en la medicina clínica como herramienta para el diagnóstico y seguimiento de afecciones como la obesidad y el sobrepeso. Además se

introduce el estado del arte de las principales tecnologías que hoy en día son utilizadas para capturar y procesar información.

1.1. Contexto

De acuerdo con la Organización Mundial de la Salud (OMS), obesidad y sobrepeso constituyen el principal problema de salud en el siglo XXI. En Argentina, conforme con la última encuesta nacional de factores de riesgo (Begué et al., 2013), seis de cada diez adultos, y tres de cada diez adolescentes presentan problemas de sobrepeso u obesidad. Los porcentajes de obesidad están aumentando en todo el mundo, pero especialmente en Latinoamérica (Galante et al., 2016). La obesidad y particularmente la adiposidad central están asociadas con diversas patologías crónicas, como las enfermedades cardiovasculares, síndrome metabólico o cáncer, entre otras. Los métodos de detección temprana y los valores de corte precisos son un problema importante en muchos estudios de población (Ashwell et al., 2012, 2014, Krakauer and Krakauer, 2012, Fagherazzi et al., 2013, Romero-Corral et al., 2006, Löffler-Wirth et al., 2016, Lanas et al., 2017). En la práctica clínica, las medidas antropométricas son importantes para evaluar la evolución del desarrollo y la predisposición a enfermedades. Además, dichas medidas se usan ampliamente durante las fases de diagnóstico, tratamiento y seguimiento de la obesidad y las patologías relacionadas. Estas medidas están basadas en indicadores clásicos como la relación entre peso y altura (Keys et al., 1972), circunferencia de cintura y cadera e índices derivados (Ashwell et al., 2012, Larsson et al., 1984, Lapidus et al., 1984). En algunos casos, se utilizan evaluaciones antropométricas más complejas, para obtener un marco general de proporciones corporales y distribución de grasa. Ejemplos recientes pueden verse en Sun et al. (2017), combinando resonancias magnéticas e imágenes 3D de todo el cuerpo.

Un aspecto importante a tener en cuenta es que una encuesta clásica de antropométrica exhaustiva lleva mucho tiempo, los dispositivos necesarios son costosos y los operadores requieren capacitación específica para incorporar dichos datos en la práctica clínica. El avance de la tecnología en la captura de datos complejos debe acompañarse con algoritmos y procedimientos que permitan extraer, automáticamente, información que aporte

nuevos indicios para entender su comportamiento, así como también, promover el desarrollo de nuevas aplicaciones que asistan en el ámbito clínico como nuevas formas de diagnóstico y tratamiento. Los escáneres de cuerpo entero son herramientas de medición efectivas, que permiten una cuantificación rápida, precisa y repetible de las dimensiones del cuerpo humano (Bretschneider et al., 2009, Pepper et al., 2010, Jaeschke et al., 2015). Además, la naturaleza digital de los datos facilita el archivo rápido y la recuperación de escaneos sucesivos (Wells et al., 2007). Debido a estas ventajas, el escáner corporal 3D es ideal para examinar individuos en encuestas epidemiológicas a gran escala (Löffler-Wirth et al., 2016, Jaeschke et al., 2015) por lo que se esperan mejoras en el diagnóstico de muchas enfermedades, reemplazando la dependencia actual de índices corporales simples con datos primarios más sofisticados (Wells et al., 2008). Existen diversos escáneres y sistemas aplicables al fenotipo corporal, pero debido a su costo no son de uso masivo o accesibles. Además los *software* desarrollados en base a estos escáneres están construidos bajo licencias privativas, por lo cual es muy difícil realizar modificaciones. Cabe considerar, por otra parte, que estos aplicativos están desarrollados y probados sobre poblaciones distintas a las Latinoamericanas.

La necesidad de tener datos sobre la población en la cual se van a desarrollar y validar las metodologías computacionales es una de las tareas más importantes y necesarias para su buen funcionamiento. Son muy pocos los estudios poblacionales aplicados sobre Argentina y países Latinoamericanos, lo cual presenta un problema para diseñar aplicaciones médicas para ser utilizadas en la región (Ruderman et al., 2019). La recolección de datos para estudios de población en humanos es realmente costosa debido a que es necesario definir protocolos de colecta sobre la población a muestrear (que cuente con la debida aprobación de un comité de bioética independiente), definir campañas publicitarias para generar una convocatoria masiva, contar con equipamiento y personal capacitado, etc.

Esta tesis se desarrolló como parte del proyecto PoblAr (Dopazo et al., 2019), cuyo objetivo principal es construir una plataforma multicéntrica capaz de construir y poner a disposición una base de datos integradora de información genómica, fenotípica y socio-cultural de referencia para Argentina. Dicha plataforma brinda información para comprender cómo un conjunto grande de genes interactúan entre sí y con diversos factores ambientales, para

dar como resultado un determinado fenotipo (por ejemplo: una enfermedad o un aumento en la susceptibilidad de padecerla). Estas particularidades genómicas y ambientales, así como sus tipos de interacción, son específicas de cada población, de modo que la construcción de estos mapas genotipo/fenotipo/ambiente también debe ser específica.

1.2. Objetivos

El objetivo principal de esta tesis es plantear el desarrollo de algoritmos y métodos para hacer factible y eficiente la extracción de características sobre un conjunto de datos orientado al cuerpo humano a grandes escalas. Se investigaron técnicas de ML y DL diseñadas para analizar datos, buscando determinar aquellas características discriminantes a determinados rasgos fenotípicos asociados a la obesidad y el sobrepeso. Entender cómo identificar estos rasgos sobre datos de la población argentina es importante para contribuir con el diagnóstico y tratamiento de estas afecciones. Para obtener los resultados esperados, los objetivos particulares propuestos son los siguientes:

- Investigar y desarrollar algoritmos de clasificación con bases de datos fenotípicas.
- Desarrollar metodologías de prueba que permitan la predicción del estado nutricional a partir de información obtenida de escaneos corporales.
- Diseñar e implementar descriptores basados en ML para representar de forma automáticamente y efectiva variaciones de forma en el cuerpo humano.
- Buscar correlaciones de información poblacional que permita inferir tendencias de interés sanitario.

1.3. Contribuciones

En primer lugar se presenta un método para obtener un descriptor a partir de un conjunto de modelos 3D que permite clasificar el sexo biológico (Navarro et al., 2018). Esta tarea es importante debido a que los diagnósticos clásicos utilizados en la práctica

necesitan de esta distinción. Además se complementa el estudio con un análisis comparativo de distintos clasificadores utilizando ML.

En segundo lugar se presenta un procedimiento de extracción automática de características para el análisis de la forma corporal. Este descriptor permite cuantificar adecuadamente el estado nutricional de una persona utilizando técnicas de segmentación en 3D. Los indicadores de forma generados a partir de este nuevo enfoque se correlacionan con el comportamiento de los marcadores clásicos utilizados actualmente en la práctica clínica: Índice de Masa Corporal (IMC), Índice Cintura Cadera (ICC) y Índice Cintura Talla (ICT) (Navarro et al., 2019a).

Por último se desarrolla un descriptor multi-vista basado en aprendizaje profundo, el cual permite realizar correspondencias densa entre dos imágenes con distintas vistas. El descriptor fue creado utilizando imágenes de una base de datos objetos genéricos y luego entrenado sobre imágenes de personas reales. Adicionalmente, algunos resultados fueron publicados en periódicos científicos y congresos, mientras que otros resultados se encuentran en preparación.

1.3.1. Publicaciones referidas a esta tesis en revistas internacionales

- **NAVARRO, P.**; RAMALLO, V.; CINTAS, C.; RUDERMAN, A.; DE AZEVEDO, S.; PASCHETTA, C.; PÉREZ, O.; PAZOS, B.; DELRIEUX, C.; GONZÁLEZ JOSÉ, R.
Body shape: Implications in the study of obesity and related traits
AMERICAN JOURNAL OF HUMAN BIOLOGY; Lugar: New York; Año: 2019
[Navarro et al. \(2019a\)](#)
- **NAVARRO, P.**; PAZOS, B.A.; CINTAS, C.; RAMALLO, V; GONZÁLEZ-JOSÉ, R.; DELRIEUX, C
Gender Recognition using 3D Human Body Scans
2018 IEEE Biennial Congress of Argentina (ARGENCON)
Lugar: San Miguel de Tucumán, Argentina, Argentina; Año: 2019;
[Navarro et al. \(2018\)](#)

- RUDERMAN, A.; PÉREZ, L. O.; ADHIKARI, K.; **NAVARRO, P.**; RAMALLO, V.; GALLO, C.; POLETTI, G.; BEDOYA, G.; BORTOLINI, M. C.; ACUÑA-ALONZO, V.; CANIZALES-QUINTEROS, S.; ROTHHAMMER, F.; RUIZ-LINARES, A.; GONZÁLEZ-JOSÉ, R.
 Obesity, genomic ancestry, and socioeconomic variables in Latin American mestizos
 AMERICAN JOURNAL OF HUMAN BIOLOGY; Año: 2019 vol. 31
[Ruderman et al. \(2019\)](#)

- TRUJILLO-JIMÉNEZ, M. A.; **NAVARRO, P.**; PAZOS, B.; MORALES, L.; RAMALLO, V.; PASCHETTA, C.; DE AZEVEDO, S.; RUDERMAN, A.; PÉREZ, O.; DELRIEUX, C.; GONZALEZ-JOSÉ, R.
 body2vec: 3D Point Cloud Reconstruction for Precise Anthropometry with Handheld Devices
 Journal of Imaging; Año: 2020 vol. 6
[Trujillo-Jiménez et al. \(2020\)](#)

1.3.2. Actas en conferencias nacionales e internacionales

- DE AZEVEDO, S.; CINTAS, C.; GONZÁLEZ, M.F.; RAMALLO, V.; QUINTO-SÁNCHEZ, M.; MARQUEZ, F.; HUNEMEIER, T; PASCHETTA, C.; RUDERMAN, A.; **NAVARRO, P**; PAZOS, B.A.; SILVA DE CERQUEIRA, C.C.; VERAN, O.; RAMÍREZ-ROZZI, F.; CASTRO, H.G.; PAZ, R.R.; GONZÁLEZ-JOSÉ, R.
 Covariation patterns and respiratory simulation in modern human and Neanderthal noses
 The 87th Annual Meeting of the American Association of Physical Anthropologists (2018)
 Lugar: Austin, Texas; Año: 2018;

- **NAVARRO, P.**; CINTAS, C.; RAMALLO, V.; DE AZEVEDO, S.; MORALES, L.; TRUJILLO, A.; PAZOS, B.A.; GONZÁLEZ-JOSÉ, R.; DELRIEUX, C.
 Escaneo corporal 3D de bajo costo para monitoreo y seguimiento remotos de sobrepeso

- 47 JAIIO- CAIS 2018 Congreso Argentino de Informática y Salud
Lugar: Buenos Aires; Año: 2018;
- **NAVARRO, P**; RAMALLO, V.; CINTAS, C.; PASCHETTA, C.; DE AZEVEDO, S.; RUDERMAN, A.; PEREZ, O; PAZOS, B.A.; DELRIEUX, C; GONZÁLEZ-JOSÉ, R.
Estrategias informáticas alternativas en el análisis de la forma corporal
XIII Jornadas Nacionales de Antropología Biológica
Lugar: Necochea-Quequén; Año: 2017;
 - **NAVARRO, P**; PAZOS, B.A.; CINTAS, C.; PEREZ, O; RAMALLO, V.; PASCHETTA, C.; RUDERMAN, A.; DE AZEVEDO, S.; GONZÁLEZ-JOSÉ, R.; DELRIEUX, C
Python en la Antropología Biológica
Pydata San Luis 2017
Lugar: San Luis; Año: 2017;
 - PAZOS, B.A.; **NAVARRO, P**; CINTAS, C.; RUDERMAN, A.; RAMALLO, V.; GONZÁLEZ-JOSÉ, R.; DELRIEUX, C; PASCHETTA, C.; DE AZEVEDO, S.
Escaneo corporal 3D de bajo costo para monitoreo y seguimiento remotos de sobrepeso
Pydata San Luis 2017
Lugar: San Luis; Año: 2017;
 - CINTAS, C.; **NAVARRO, P**; RAMALLO, V.; PAZOS, B.A.; DELRIEUX, C; RUDERMAN, A.; PASCHETTA, C.; DE AZEVEDO, S.; GONZÁLEZ-JOSÉ, R.
Posicionamiento automático de landmarks corporales 3d mediante morfometría geométrica y redes neuronales: aplicaciones bioantropológicas
XIII Jornadas Nacionales de Antropología Biológica
Lugar: Necochea-Quequén; Año: 2017;
 - PASCHETTA, C.; RAMALLO, V.; DE AZEVEDO, S.; CINTAS, C.; RUDERMAN, A.; **NAVARRO, P**; PAZOS, B.A.; PEREZ, O.; RUIZ-LINARES, A.; POLETTI, G.; GALLO, C.; BORTOLINI, MC.; CANIZALES-QUINTERO, S.; ROTHAMMER, G.; BEDOYA, G.; GONZÁLEZ-JOSÉ, R.

Distancia entre la auto-percepción de ancestría y la ancestría genética en poblaciones mestizas latinoamericanas

XIII Jornadas Nacionales de Antropología Biológica

Lugar: Necochea-Quequén; Año: 2017;

- RUDERMAN, A.; PEREZ, O.; RAMALLO, V.; **NAVARRO, P.**; CINTAS, C.; PASCETTA, C.; DE AZEVEDO, S.; PAZOS, B.A.; RUIZ-LINARES, A.; POLETTI, G.; GALLO, C.; CANIZALES-QUINTERO, S.; ROTHAMMER, G.; BEDOYA, G.; GONZÁLEZ-JOSÉ, R.

Variación del estado nutricional en poblaciones latinoamericanas y su relación con variables socioculturales y genéticas

XIII Jornadas Nacionales de Antropología Biológica

Lugar: Necochea-Quequén; Año: 2017;

1.3.3. Otras publicaciones y presentaciones en congresos y revistas.

- **NAVARRO, P.**; CINTAS, C.; LUCENA, M.; FUERTES, J. M.; DELRIEUX, C.; MOLINOS, M.

Unsupervised techniques for shape analysis of Iberian ceramics based on Geometric Morphometrics

CAA 2019 Kraków

Lugar: Kraków; Año: 2019;

- CINTAS, C.; DELRIEUX, C; **NAVARRO, P.**; QUINTO-SÁNCHEZ, M.; PAZOS, B.A.; GONZÁLEZ-JOSÉ, R.

Automatic Ear Detection and Segmentation over Partially Occluded Profile Face Images

JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY; Lugar: La Plata; Año: 2019 vol. 19 p. 81 - 90

- CINTAS, C.; LUCENA, M.; FUERTES, J. M; DELRIEUX, CLAUDIO; **NAVARRO, P.**; GONZÁLEZ-JOSÉ, R.; MOLINOS, M.

Automatic feature extraction and classification of Iberian ceramics based on deep convolutional networks

JOURNAL OF CULTURAL HERITAGE; Año: 2019

- ALMONACID, S.; KLAGGES, M. R.; **NAVARRO, P.**; MORALES, L.; PAZOS, B.; PUIGBÓ, A. CONTRERAS; FIRMENICH, DIEGO
Mobile and Wearable Computing in Patagonian Wilderness
Springer; Año: 2019 vol. 1050 p. 137 - 154
- QUINTO-SÁNCHEZ, M.; MUÑOZ-MUÑOZ, F.; GOMEZ-VALDES, J.; CINTAS, C.; **NAVARRO, P.**; CERQUEIRA, C.; PASCHETTA, C.; DE AZEVEDO, S.; RAMALLO, V.; ACUÑA-ALONZO, V.; ADHIKARI, K.; FUENTES-GUAJARDO, M.; HÜNEMEIER, T.; EVERARDO, P.; DE AVILA, F.; JARAMILLO, C.; ARIAS, W.; GALLLO, C.; POLETTI, G.; BEDOYA, G.; BORTOLINI, M. C.; CANIZALES-QUINTEROS, S.; ROTHHAMMER, F.; ROSIQUE, J.; RUIZ-LINARES, A.; GONZALEZ-JOSE, R.
Developmental pathways inferred from modularity, morphological integration and fluctuating asymmetry patterns in the human face
Scientific Reports; Año: 2018 vol. 8
- DE AZEVEDO, S.; GONZÁLEZ, M.F.; CINTAS, C.; RAMALLO, V.; QUINTO-SÁNCHEZ, M.; MÁRQUEZ, F.; HÜNEMEIER, T.; PASCHETTA, C.; RUDERMAN, A.; **NAVARRO, P.**; PAZOS, B.A.; SILVA DE CERQUEIRA, C.C.; VELAN, O.; RAMÍREZ-ROZZI, F.; CALVO, N.; CASTRO, H.G.; PAZ, R.R.; GONZÁLEZ-JOSÉ, R.
Nasal airflow simulations suggest convergent adaptation in Neanderthals and modern humans
PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA; Año: 2017 vol. 114 p. 12442 - 12447

1.4. Aplicaciones

El uso de escáneres 3D para capturar información sobre la forma del cuerpo humano está creciendo rápidamente y no solo se limita a estudios ergonómicos o de diseño, sino que también resulta de interés para aplicaciones relacionadas con la salud. Definir protocolos para la captura y el análisis de estos datos representa una tarea importante. El desarrollo de aplicaciones médicas y su validación representa un problema en Latinoamérica, ya que son mínimas las aplicaciones desarrolladas para *población específica* en la zona.

Los modelos y descriptores desarrollados en esta tesis, permitirán crear aplicaciones basados en datos de población específica, que concederá información detallada sobre características de los pacientes, tanto para realizar diversos análisis (por ejemplo el somatotipo o estudio de porcentaje y distribución del tejido adiposo), como para el diagnóstico apropiado de condiciones médicas (evaluación del sobrepeso, detección de malformaciones, etc.).

1.5. Estructura de la Tesis

La presente tesis se divide en siete capítulos, el primero introduce la problemática y aborda los conceptos biológicos que encierran la solución. El segundo capítulo presenta el marco teórico para la resolución de problemas en visión computacional, e introduce los conceptos de inteligencia artificial junto con técnicas de *machine learning* y *deep learning*. El capítulo tres se centra en cómo fue la adquisición de los datos utilizados en el presente trabajo junto con los protocolos, estrategia de comunicación y características de la muestra. Los capítulo cuatro, cinco y seis presentan métodos que permiten extraer características utilizando distintas técnicas sobre imágenes para análisis de población. El capítulo siete culmina la investigación con conclusiones y trabajo futuro.

Capítulo 2

Marco Teórico

2.1. Visión Computacional

Los objetivos de la Visión Computacional (VC) pueden interpretarse de dos formas dependiendo el campo donde se los estudie. Desde el punto de vista de las ciencias biológicas, la VC tiene como objetivo crear modelos del sistema visual humano, los cuales permiten estudiar y comprender el comportamiento de estos complejos mecanismos ([Huang, 1996](#)). Por otro lado, la ingeniería nos plantea como objetivo construir sistemas autónomos que permitan automatizar tareas que requieran cognición visual. Estas están relacionadas con la extracción de información 3D a partir de datos 2D con variación en el tiempo, como por ejemplo comprender escenas dinámicas obtenidas en cámaras de vídeo. Los objetivos de la VC en ambos campos están estrechamente relacionados. Las propiedades y características del sistema visual humano inspiran a quienes diseñan sistemas de visión por computadora y, por otro lado, los algoritmos que comprenden estos últimos ofrecen información sobre cómo es el funcionamiento de este sistema ([Huang, 1996](#)). En esta tesis utilizaremos la definición más ingenieril, definiendo a la VC como un campo científico interdisciplinario que se ocupa de estudiar cómo las computadoras pueden obtener una comprensión de alto nivel a partir de vídeos o imágenes 2D o 3D.

Hoy en día las aplicaciones de la VC van desde la medicina hasta la industria e involucran también otras ramas de la ciencia. Por ejemplo, en salud existen varias aplicaciones que van

desde el estudio e identificación de enfermedades a partir de imágenes médicas (Islam and Zhang, 2017, Sharma et al., 2017, Yu-Jen Chen et al., 2015, Poplin et al., 2018a) hasta la reconstrucción de órganos o estructuras internas (Xiong et al., 2015, Benz et al., 2020) para su estudio. Por otra parte, en arqueología se utilizan técnicas de VC para asistir a los arqueólogos a la hora de clasificar e identificar (en tiempo y lugar) materiales encontrados en excavaciones (Cintas et al., 2020). En materia de seguridad, VC es utilizada para los análisis biométricos tales como el reconocimiento facial (Kumar et al., 2019, Parmar and Mehta, 2014), de retina, entre otros (Guo et al., 2018, Cintas et al., 2016) con propósitos de identificación individual y seguimiento.

En la última década, la VC se enriqueció significativamente mediante la utilización conjunta de técnicas de Inteligencia Artificial como *Machine Learning* y, en particular, *Deep Learning*. Si bien siguen utilizándose algoritmos basados en métodos tradicionales (la denominada *feature engineering* o ingeniería de atributos), estas técnicas tradicionales no pueden ser aplicadas a todos los problemas de VC, y para grandes volúmenes de datos el procedimiento puede ser ineficiente. En la figura 2.1 se puede ver el esquema de etapas que permiten resolver un problema de VC.

El uso de deep learning en VC se puede pensar como la composición de tres tareas principales: el acopio de datos, el pre-procesamiento y la técnica. En cuanto al proceso de recolectar datos, resulta muy difícil o imposible construir una aplicación de VC sin contar con una base de imágenes de las cuales extraer información. Para ello, debemos definir cómo es posible obtener imágenes 2D y 3D en la actualidad, con el fin de conformar un conjunto de datos comúnmente conocido como *dataset*. Veremos en secciones posteriores que la calidad y condición de las imágenes de entrada influyen de manera significativa en la calidad de los resultados obtenidos. El pre-procesamiento de datos puede ayudar a obtener mejores resultados ya que, en términos generales, reduce la complejidad del problema y el tiempo necesario para extraer información de utilidad necesaria para resolver un problema. Durante esta etapa, y dependiendo del escenario, es común utilizar técnicas de manipulación de imágenes como normalización, segmentación y reducción de dimensionalidad, entre otras. La técnica, finalmente, consiste en definir o reutilizar arquitecturas de aprendizaje profundo,

las cuales si la problemática lo permite podrían utilizarse pre-entrenadas adaptándolas al problema particular.

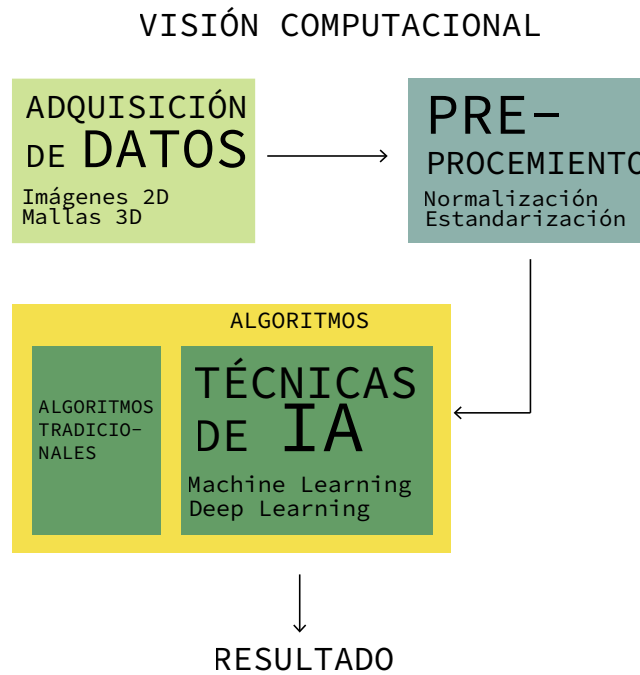


Figura 2.1: Esquema de las etapas que componen la VC.

2.2. Adquisición de datos

2.2.1. Imágenes 2D

Como resultado del avance constante de la tecnología, en la actualidad se pueden obtener imágenes de mejor resolución y calidad que hace pocos años atrás. A tal punto, que en ocasiones, las diferencias entre cámaras profesionales y cualquier dispositivo móvil es mínima. Los *smartphone* que se venden en el mercado pueden obtener imágenes superiores a 15 Mpx. Además de esto, las técnicas actuales de *rendering* 2D permiten reconstruir imágenes artificiales de objetos tridimensionales con un gran realismo (Ver figura 2.2).

La capacidad de cómputo actual no solo permite realizar procesamiento de imágenes más complejo, sino que también mucho más veloz. Los conjuntos de datos disponibles de manera gratuita en la web, por ejemplo *ImageNet*, superan el orden del millón de imágenes. El acopio masivo (ya sea imágenes reales o artificiales) permite resolver problemas de VC

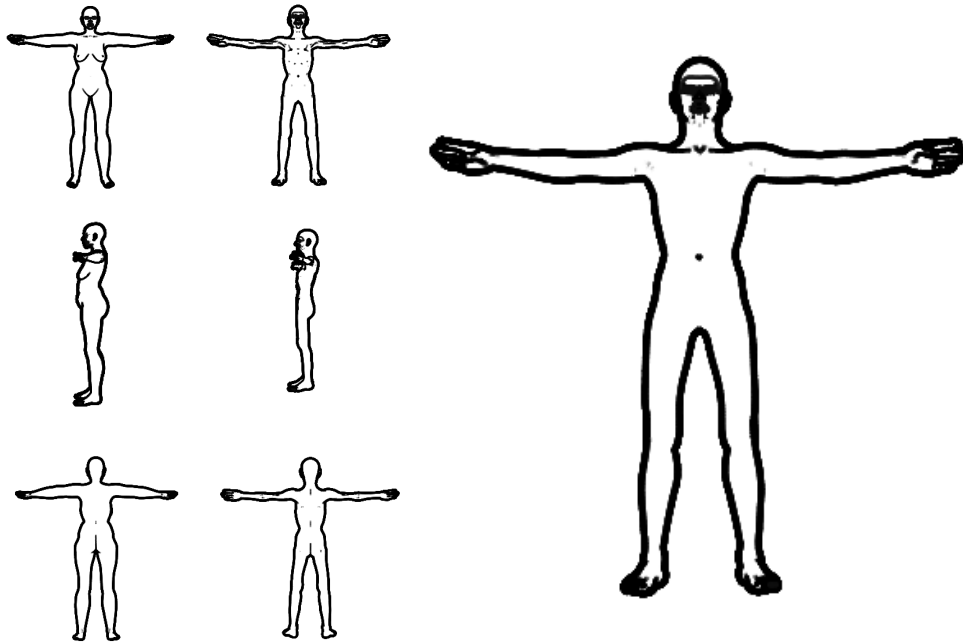


Figura 2.2: Ejemplo de modelos creados a partir del software Pose 3D, utilizando “Apparent Ridges algorithm” (Judd et al., 2007) para generar una imagen con estilo de dibujo sobre el modelo 3D.

cada vez más complejos. En la Tabla 2.1 se puede ver la descripción de los cuatro *datasets* más conocidos, los cuales nos brindan estructuras de datos para resolver distintos tipos de problemas como: clasificar y segmentar imágenes, detectar objetos y escenas, describir un contexto, etc.

2.2.2. Nube de Puntos y Mallas 3D

En diversos escenarios donde se busca implementar soluciones por medio de VC, es necesario capturar por completo la naturaleza geométrica de los objetos. Para lo cual necesitamos un tipo de dato que nos permita modelar el objeto en tres dimensiones. Es aquí en donde aparecen las nubes de puntos y mallas 3D. Una nube de puntos 3D es un conjunto de puntos posicionados en un espacio tridimensional. La cantidad de puntos que conforman la nube determina la calidad de la misma. Además de las coordenadas de cada punto (X, Y, Z , en este caso) puede contener un color asociado, típicamente representado por una terna de valores RGB (0–255). En la figura 2.3 puede verse una nube de puntos de la Catedral de Murcia capturado con un escáner profesional. Por otra parte se encuentran

<i>dataset</i>	Imágenes	Descripción
ImageNet (Deng et al., 2009)	+14M de imágenes. +20k categorías.	Clasificación de Imágenes. Anotaciones en texto de imágenes.
CIFAR-100 (Krizhevsky et al., 2009)	+60k imágenes. 100 clases. 20 superclases.	Clasificación de Imágenes. Clasificación en superclases.
COCO (Lin et al., 2014)	+330k imágenes. 1,5M objetos segmentados. 80 clases.	Clasificación de Imágenes. Reconocimiento del contexto. Segmentación.
Pascal VOC (Everingham et al., 2010)	+12mil imágenes. +20mil objetos segmentados. 20 clases.	Clasificación de Imágenes. Segmentación.

Tabla 2.1: Descripción de los *dataset* 2D más conocidos.

las Mallas 3D (*3D Mesh* en inglés) que además de estar compuestas por un conjunto de puntos llamados vértices, contiene una lista de caras conformadas por la unión de tres de estos. Las caras componen la unidad básica de un polígono tridimensional. En la figura 2.4 puede verse un ejemplo de malla 3D. Para las mallas y puntos 3D, el acopio masivo de datos es más complejo, ya que en la actualidad no es tan frecuente contar con un sensor 3D. Si bien algunos teléfonos modernos están incorporando este tipo de tecnología, aún no es tan accesible como una cámara de fotos. En este punto debemos pensar en técnicas que nos permita hacer más accesible este tipo de datos, para las cuales existen varias alternativas para modelar un objeto tridimensional dependiendo del dispositivo o de la técnica que utilicemos: escáneres comerciales, procesamiento 2D y modelado manual. La más trivial, es la de utilizar algún tipo de dispositivo comercial de escaneo 3D, si bien en el mercado hay varias opciones que ofrecen resultados de alta calidad, no son accesibles en cuanto a precio. Al mismo tiempo, el costo del dispositivo tiende a subir dependiendo del tamaño de los objetos que se requieran escanear (no es lo mismo un dispositivo para escanear una habitación completa, que una pieza mecánica de 15cm). Otra técnica consiste en utilizar procesamiento de imágenes en 2D, tales como la fotogrametría, que permitan reconstruir en 3D un objeto o escena real a partir de un conjunto de imágenes 2D. Estos métodos resultan más accesibles en cuanto a costo en comparación con un escáner comercial, en este caso la calidad de los resultados pasa a depender de la cantidad de imágenes y vídeos disponibles.



Figura 2.3: Nube de puntos de la Catedral de Murcia. Sensor Leica RTC360 ([Dronica](#)).

Por último, se encuentra el modelado manual, el cual se basa en esculpir manualmente en tres dimensiones los objetos o escenas reales utilizando un software dedicado para este fin. Este proceso es muy laborioso dado que requiere una extensa cantidad de mediciones de los objetos que se quieran replicar. Por motivos de practicidad y menor costo, se descartó la tercer alternativa y en los siguientes apartados se describirán con mayor detalle las primeras dos.

2.2.3. Técnicas de escaneo 3D

Como se mencionó anteriormente, la elección del tipo de dispositivo a utilizar está dada por las características del objeto que debe ser escaneado. La estructura, el material, el tamaño y la calidad de escaneo determinan la tecnología a utilizar, cada una de estas tienen sus desventajas, limitaciones y costos. En esta tesis se detalla el funcionamiento de las cuatro más empleadas:

Triangulación láser

Este método se basa en proyectar un rayo láser en la superficie a escanear y luego, valiéndose de un sensor, mide la deformación de los rayos emitidos. El láser a utilizar

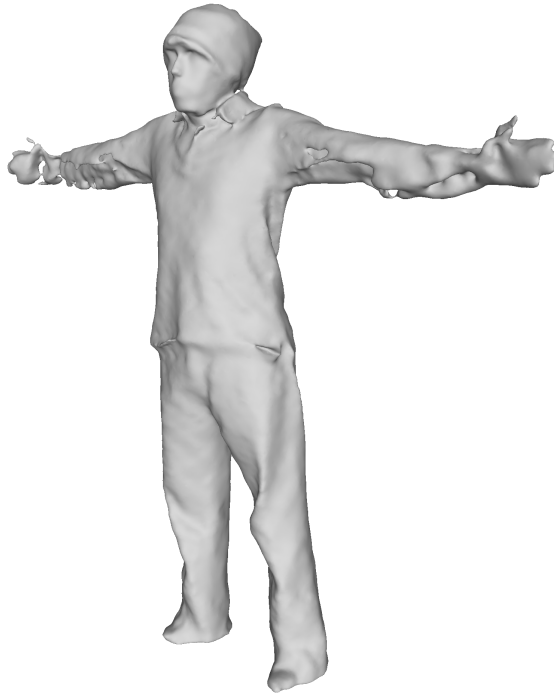


Figura 2.4: Malla 3D. Sensor Structure. 99,000 puntos (aprox 199,000 triángulos).

puede ser un único punto o una línea con varios punteros. La mecánica de estos tipos de escáneres es muy sencilla de explicar, se dispara uno o varios rayos sobre el objeto y se registra, mediante un sensor receptor, como es la trayectoria del haz de láser al reflejar sobre el objeto. El ángulo de desviación específico es calculado aplicando una triangulación sobre el cambio de trayectoria junto con el ángulo original y la distancia que hay entre el escáner y el objeto. Conociendo esta última distancia, el hardware de escaneo puede mapear la superficie en una nube de puntos, generando una representación tridimensional del objeto escaneado. La ventaja de esta tecnología es su resolución y precisión, mientras que la desventaja es que es sensible a las propiedades de la superficie que se escanea. Al utilizar el rebote del haz de luz, objetos con superficies muy brillantes o transparentes pueden deformar o alterar la trayectoria de dicho haz, lo que provoca una reconstrucción con fallas, malformaciones o imperfecciones. En la figura 2.5 se puede ver un ejemplo comercial de este tipo de escáner.



Figura 2.5: *Structure Sensor*. Escáner basado en triangulación láser ([ItSeez3D](#)).

Basado en luz estructurada

Esta tecnología busca comprender la geometría tridimensional de un objeto o escena, a partir de analizar el patrón de deformación de la luz emitida sobre el mismo. El dispositivo proyecta una serie de luces en forma de patrón (fijo o aleatorio) sobre un objeto, la cual es examinada mediante un sensor. Las deformaciones en los bordes del patrón permiten calcular indirectamente la distancia hasta la superficie del objeto y su forma. Las luces empleadas son de colores blanco o azul y están generadas utilizando varios tipos de proyectores, como la tecnología DLP (sigla en inglés de Procesamiento digital de luz). Como puede apreciarse, esta técnica se basa en un método de triangulación similar al del método anterior, pero utilizando patrones de luz fijos. Como ventajas se lista resolución, velocidad y posibilidad de escanear personas (si es que se le suma alguna protección para la vista). La principal desventaja es su sensibilidad con respecto a condiciones de iluminación, por lo que resulta difícil trabajar al aire libre o en lugares con iluminación muy pobre. En la figura 2.6 se puede ver en funcionamiento el escáner *GoScan20*, el cual se basa en luz estructurada.

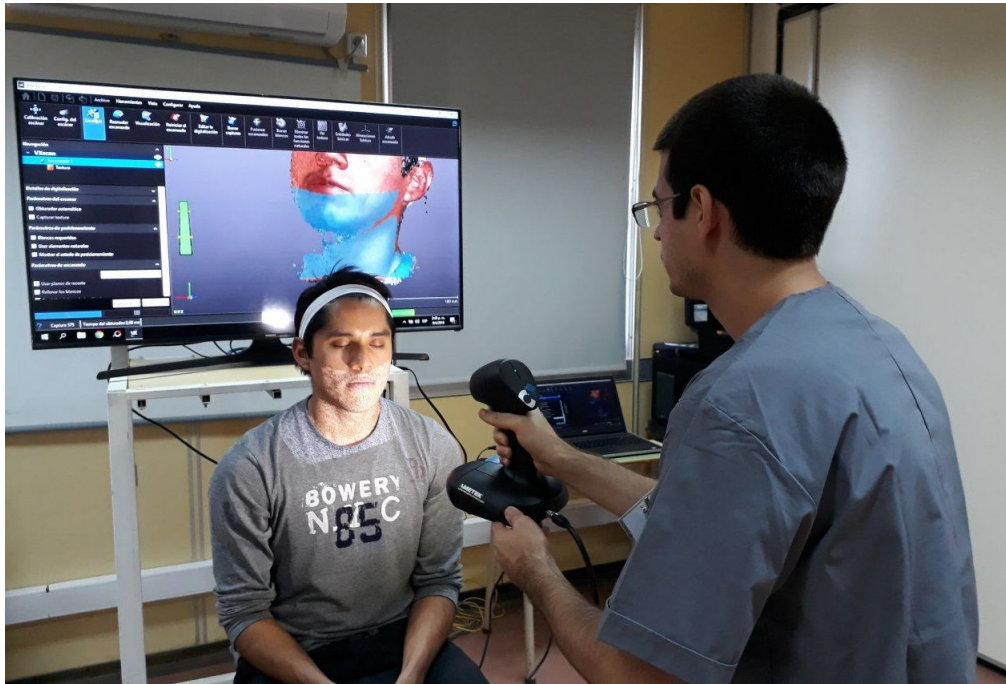


Figura 2.6: Escaneo de rostro con *GoScan20*. Escáner basado en luz estructurada. Foto capturada durante la colecta de datos Raíces en Puerto Madryn.

Basado en contacto

También llamado "Digitalizador", este tipo de dispositivo cuenta con un brazo o lápiz conectado al escáner. Si el dispositivo tiene un brazo móvil, entonces el objeto se mantiene estacionario, de lo contrario es el objeto el que está constante movimiento. Cuando el brazo entra en contacto con la superficie a escanear, se obtienen datos del objeto e información de la ubicación del punto. A medida que se relaciona con diferentes puntos en la superficie, el software de escaneo crea una nube de puntos del objeto en un espacio tridimensional. Este se utiliza principalmente para el control de calidad después de la fabricación o durante las operaciones de mantenimiento. La presente metodología tiene como ventaja una alta precisión y la capacidad de escanear superficies que reflejen la luz o sean transparentes. Como desventaja, relevar una gran cantidad de puntos puede ser un proceso muy tedioso y lento si se realiza manualmente en comparación con otras técnicas y no permite trabajar con objetos en movimiento o cambio de forma.

Basado en pulso LÁSER

Esta tecnología se basa en el tiempo de viaje de un rayo láser. El escáner contiene un dispositivo que ilumina un objeto con un rayo de luz. El reflejo de dicho láser sobre el objeto es capturado por un sensor que mide el tiempo que tarda el láser en viajar desde el punto de emisión hasta el punto de recepción en el sensor. Esta información proporciona la geometría del objeto. Como se conoce la velocidad del láser al viajar en el ambiente, el intervalo de tiempo puede proporcionar la distancia exacta entre el objeto y el escáner 3D. El láser debe emitirse 360 grados alrededor de un punto en particular, ya que cada medida solo recopila datos de un punto. La dirección o trayectoria del láser se cambia con la ayuda de un espejo o superficie reflectante que se fija en el escáner 3D. Algunos modelos cuentan con un tipo de láser que es capaz de modular la potencia del rayo, por lo tanto, ofrecen un mejor rendimiento. La gran ventaja de esta técnica es la posibilidad de escanear objetos de gran tamaño y de poder funcionar en grandes entornos. Como desventaja, esta metodología requiere mayor tiempo que las demás para lograr procesar toda la información de manera exitosa.

En conclusión, el uso de todos estos dispositivos tiene una gran aplicación en la industria automotriz, aeronáutica, dental, videojuegos y efectos especiales. Se puede observar entonces que la elección de la metodología de escaneo adecuada depende de características propias del tipo de objeto o escena a escanear, impactando directamente en el costo del dispositivo necesario y, dependiendo el caso, tornando el proceso en algo costoso y poco accesible.

2.2.4. Fotogrametría

La fotogrametría es una técnica que permite interpolar de manera precisa objetos tridimensionales a partir de pares o grupos de imágenes 2D. Funciona haciendo corresponder una serie de puntos sobre dos o más imágenes. De esta manera se puede inferir la posición de la cámara en cada imagen, con lo cual se puede obtener las coordenadas de cada punto en un espacio 3D. En la fotogrametría tradicional, los puntos son ubicados manualmente sobre las imágenes. El principio fundamental utilizado en esta técnica es el de triangulación.

Tomando al menos dos imágenes desde puntos diferentes sobre un mismo objeto podemos calcular las llamadas “líneas de visión” desde cada posición de la cámara hasta el objeto (Ver figura 2.7). Estas líneas se cruzan matemáticamente para producir las coordenadas tridimensionales de los puntos de interés, de la misma manera que trabaja el ojo humano para medir la distancia (llamada percepción de profundidad). *Structure from Motion (SFM)* es

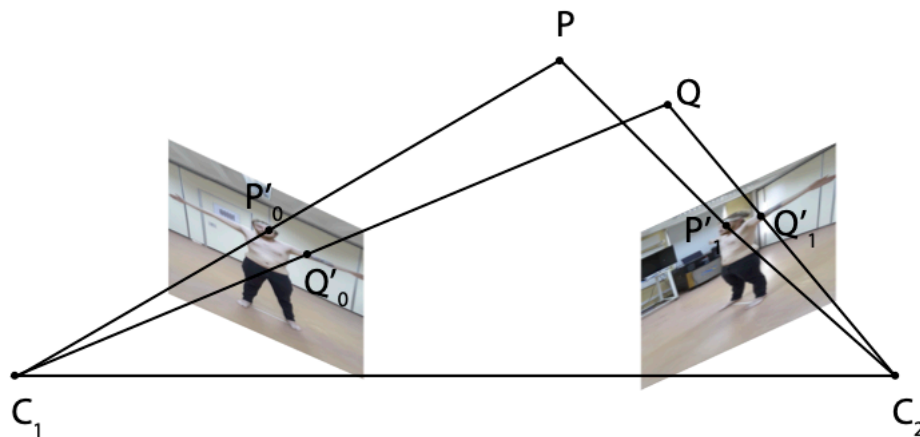


Figura 2.7: P y Q visto de dos cámaras C_1, C_2 . Los puntos 2D P'_1, P'_2, Q'_1 y Q'_2 son los puntos de interés visto en dos imágenes. P y Q es proyección 3D calculada.

una técnica de alta resolución y bajo costo basada en fotogrametría. Se utiliza en combinación con potentes algoritmos de geometría computacional y VC para analizar fotografías de un objeto estático capturado desde diferentes puntos de vista. Ejemplos de estos algoritmos pueden ser las técnicas de *Image Matching* como SURF (Bay et al., 2006), SIFT (Lowe, 2004), BRIEF (Calonder et al., 2010) u ORB (Rublee et al., 2011), entre otras. Todos ellos, en mayor o menor medida, son capaces de resolver automáticamente las características coincidentes entre varias imágenes. Estas características distintivas son a menudo esquinas o segmentos de línea, las cuales se rastrean de una imagen a otra y se utilizan para producir estimaciones de las posiciones y orientaciones de la cámara, finalizando este proceso con una nube de puntos de coordenadas X, Y, Z . En la actualidad, estas técnicas son utilizadas como alternativa para el escaneo 3D sin hacer uso de dispositivos específicos. El mayor desafío de esta tecnología es analizar con precisión múltiples fotos. Por lo tanto, se requiere de capacidad de cómputo muy potente para ejecutar estos algoritmos.

2.3. Pre-procesamiento

Para acelerar y mejorar las metodologías de VC, se aplica una etapa de pre-procesamiento que incluye *la estandarización y normalización*, la limpieza general de los datos y la *reducción de dimensionalidad*. Básicamente, esta etapa consiste en pre-procesar los datos de entrada y obtener características (o en inglés *features*) que representen de mejor manera el dato. Por ejemplo, para el problema de reconocimiento de dígitos, las imágenes generalmente se trasladan y escalan para que cada dígito este contenido dentro de una caja de un tamaño fijo. Esto sirve como ayuda para que los algoritmos obtengan una abstracción que se centre en diferenciar por la forma del carácter y no por el tamaño ni posición dentro de una imagen. Este tipo de procesamiento se lo conoce como *Estandarización*, donde cada dato es procesado con un mismo método para que cumpla con una serie de características al final de esta etapa.

Además de esta técnica, existen otras que son comunes en el pre-procesamiento de datos utilizadas para los métodos de VC. Dado que los modelos de VC son sensibles a las muestras defectuosas en el conjunto de datos, la limpieza de datos es esencial para un mejor rendimiento. Las tareas pueden incluir eliminar o corregir datos faltantes o valores atípicos.

Por otro lado, se aplica una etapa de normalización del conjunto de datos, la cual evita el problema de trabajar con diferente rangos de atributos y ayuda a que el algoritmo funcione con mayor precisión. Si bien la mayoría de los conjuntos de datos consisten en diferentes escalas de valores (por ejemplo, el peso y altura de una persona o información sobre la alimentación) la normalización de estos valores, antes de ser la entrada del algoritmo, proporciona un aprendizaje más fácil y un menor costo de cálculo. La normalización se enfoca en la significancia estadística de los atributos, en caso de que estos puedan parametrizarse, utilizando el ajuste a una distribución. En capítulos posteriores de esta tesis se utiliza el método de normalización implementado en [Pedregosa et al. \(2011a\)](#), definido de la siguiente manera:

$$X_s = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (2.1)$$

$$X_s = X(f_{r_{max}} - f_{r_{min}}) + f_{r_{min}}, \quad (2.2)$$

donde X es el vector de entrada, $\max(X)$ y $\min(X)$ son los valores máximos y mínimos del vector X . X_s son los valores normalizados para X , y $f_{r_{min}}$ y $f_{r_{max}}$ son los valores normalizados al rango $[0, 1]$.

Otro tipo de proceso aplicable es la reducción de dimensionalidad. Si bien esta técnica involucra el uso de las demás, es importante destacarla debido a que este procesamiento es utilizado en muchas aplicaciones de VC. Básicamente, consisten en reducir el número de variables utilizadas como entrada, esto se puede realizar seleccionando características existentes o extrayendo nuevas a partir de un nuevo espacio. La selección de características puede darse de forma aleatoria, eliminando las variables dependientes o seleccionando a partir de un criterio de resultados. La más básica y, por lo tanto no tan efectiva, es seleccionar un conjunto de características de manera aleatoria utilizando algún tipo de distribución estadística. Si bien esto puede disminuir el tiempo de procesamiento, generalmente no se obtienen buenos resultados ya que pueden dejar afuera características importantes.

Otra alternativa es eliminar variables dependientes, que además de disminuir el tiempo de procesamiento permite eliminar la redundancia en los datos sin perder calidad en los resultados. Esta tarea no puede realizarse siempre, porque depende de los datos disponibles (caso en que todas los atributos sean independientes). También se puede seleccionar según el rendimiento del modelo, es decir, seleccionar por un criterio elegido con anterioridad y poner a prueba el modelo de forma tal que podamos evaluar la selección. Este procedimiento es más tedioso, debido a que la cantidad de selecciones que hagamos son la cantidad de veces que debemos probar el modelo, además no nos asegura que hayamos elegido el mejor conjunto de características posibles. Para asegurarnos esto, podemos realizar una selección a *fuerza bruta*, esto nos asegura que todos los casos son puestos a prueba, aunque el costo de cómputo aumenta considerablemente.

Para finalizar con la etapa de preprocesamiento, existe otra manera de reducir la di-

mensionalidad, la cual consiste en extraer características a partir de un nuevo espacio de menores dimensiones. La transformación de datos de un espacio a otro puede ser lineal, como Análisis de Componentes Principales (PCA), o no lineal como *T-distributed Stochastic Neighbor Embedding* (t-SNE) (Maaten and Hinton, 2008) o *Uniform manifold approximation and projection* (UMAP) (McInnes et al., 2018). En capítulos posteriores de esta tesis se utiliza PCA para extraer nuevas características. PCA se basa en buscar conjuntos de combinaciones lineales de las variables originales que contienen la mayor varianza. Como depende de la cantidad de combinaciones que se utilicen, pueden emplearse en conjunto con un método de selección (como por ejemplo: selección al azar) para elegir la cantidad de combinaciones que se usen.

2.4. Inteligencia Artificial

En la actualidad, ayudado por el aumento de la capacidad de cómputo, emergieron nuevas técnicas que si bien fueron planteadas teóricamente en los años 60, el contexto actual permitió que puedan ponerse en práctica. La Inteligencia Artificial (IA) es la principal área de la computación que atiende los objetivos de la VC. El nombre fue acuñado en 1956 (Russel et al., 2013) y está definida como el campo de estudio de “agentes inteligentes”: cualquier dispositivo que perciba su entorno y reaccione a partir de él. Los autores (Russel et al., 2013) lo definen coloquialmente como una máquina que imita las funciones cognitivas que los humanos asocian con otros humanos, como podrían ser las acciones de percibir, razonar, aprender y resolver problemas. Dentro de la IA se puede encontrar dos técnicas muy utilizadas en la actualidad como el Aprendizaje de Máquina o en inglés Machine Learning (ML) y el Aprendizaje Profundo o Deep Learning (DL). La figura 2.8 esquematiza cómo cada técnica está contenida, ya que éstas comparten, en gran medida, principios y conceptos.

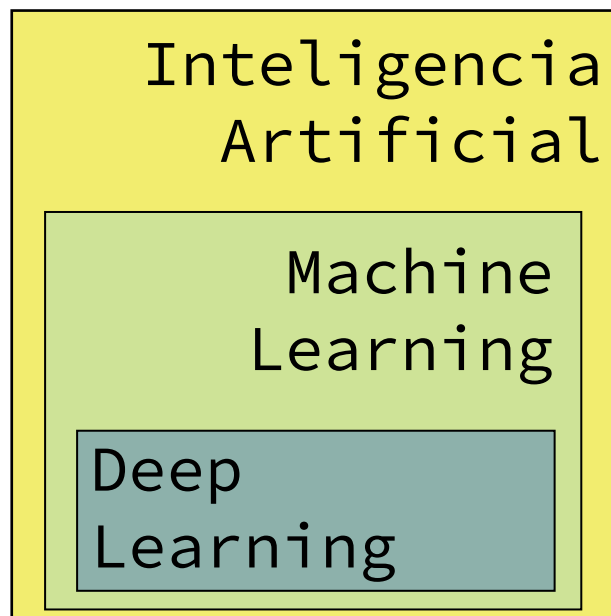


Figura 2.8: Esquema de cómo se compone la IA.

2.4.1. Machine Learning

ML es una colección de algoritmos y técnicas utilizadas para diseñar programas que se construyen a sí mismos a partir de datos de entrada. ML se basa en conceptos matemáticos, estadísticos y de computación. Además de estas ciencias, cada problema aporta conceptos únicos del dominio a resolver, por ejemplo si se quiere reconocer especies de araña a partir de imágenes, necesitamos aplicar algunos conocimientos básicos sobre la taxonomía de las arañas. Los algoritmos se pueden catalogar en **entrenamiento supervisado** y **no supervisado** dependiendo del conjunto de datos de entrenamiento.

Entrenamiento supervisado

Para el entrenamiento supervisado se utiliza un conjunto de datos etiquetado, es decir que contiene un grupo de registros que fueron categorizados o marcados empleando un campo en especial que proporciona información relevante a los datos. Por ejemplo, en un conjunto de datos referido a una población se puede tener un campo que determine si una persona tuvo o no un problema cardíaco, o si hace o no regularmente algún tipo de actividad física. Usando estos datos se puede entrenar modelos que intenten predecir etiquetas para registros que no contengan esa información.

Entrenamiento no supervisado

En el aprendizaje no supervisado, el conjunto de datos utilizado no está etiquetado. Un ejemplo podría ser un conjunto de datos que contienen medidas del cuerpo (altura, circunferencia de cadera y cintura) de 300 personas. Utilizando el aprendizaje no supervisado, el objetivo principal es descubrir y estudiar patrones en los datos. Una forma posible sería utilizando algoritmos de clusterización para encontrar los patrones en el conjunto de datos. El resultado final podría ser el descubrimiento de grupos que expliquen algún tipo de comportamiento.

Además de la clasificación anterior, los problemas que se pueden resolver con estas técnicas suelen ser categorizados a partir de tres preguntas: ¿Pertenece a la clase 0 o 1

? *Clasificación*; ¿Cuánto o cuántas?. *Regresión*; ¿Cómo se agrupan?. *Clusterización* (Ver figura 2.9).

Clasificación

Determinar la clase, origen o cualquier categoría de una muestra, es un problema común en muchas ramas de la ciencia. ML permite identificar a qué conjunto pertenece una nueva observación en función del conjunto de datos de entrenamiento que contiene las categorías observadas. Algunos ejemplos de problemas de clasificación son:

- En la práctica clínica, es necesario conocer el estado nutricional de una persona con el fin de diagnosticar o tratar enfermedades. A partir de imágenes podemos determinar si una persona es saludable, tiene sobrepeso u obesidad.
- En arqueología, clasificar materiales encontrados en nuevos yacimientos ayuda a conocer la cronología, la funcionalidad y el origen aportando evidencias de la producción local, relaciones comerciales, comportamiento de consumidores en poblaciones locales, entre otras.
- En finanzas, puede ser necesario determinar si un conjunto de operaciones bancarias son o no parte de un fraude.

Entre los ejemplos descritos se pueden identificar clasificaciones binarias donde un problema cuenta solo con dos clases posibles (fraude, no fraude) o aquellos con más de dos clases denominados de multi-clase (saludable, sobrepeso, obeso). El resultado de un problema de clasificación es un valor discreto que indica la clase predicha en la que se encuentra una observación aunque también podría ser un valor continuo, que indica la probabilidad de una observación que pertenece a una clase en particular.

Regresión

Si bien la clasificación puede ser resuelta mediante regresión, en esta categoría se ubican todo los problemas que requieren la predicción de valores continuos, por ejemplo:

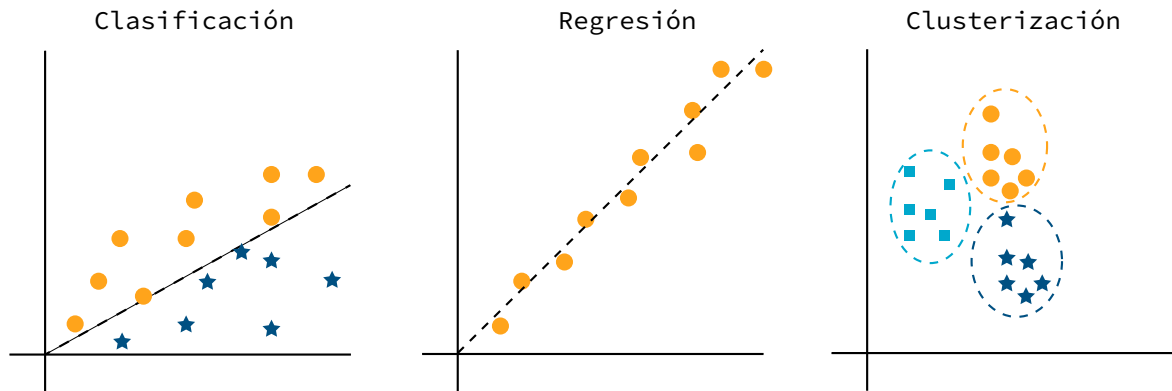


Figura 2.9: Clasificación, Regresión y Clusterización.

- Climatología, predecir la temperatura máxima y mínima de la próxima semana.
- Determinar la vida útil de un producto a partir de sus características.
- Finanzas, predecir la ganancia de una empresa o conjunto de empresas según las tendencias del mercado.

La regresión es el algoritmo más básico que nos permite pronosticar valores al estimar la relación entre variables o conjunto de variables.

Clusterización

La clusterización es la acción de agrupar en conjuntos datos con similares características o comportamiento. Esta agrupación generalmente se da en grupos intuitivos, lo que ayuda a descubrir cómo se organizan naturalmente. La agrupación en clúster es muy útil cuando se desea descubrir un patrón específico en datos. Ejemplos de problemas de clusterización:

- Para publicidad, agrupar usuarios en base a sus gustos es un problema común para la utilización de propaganda dirigida.
- Para arqueología, estudiar agrupaciones de restos arqueológicos, por ejemplo vasijas completas o fragmentos para determinar conexiones entre yacimientos encontrados, rutas comerciales o técnicas de elaboración

- En la fabricación de insumos, la agrupación de productos por características puede ser utilizado como objetivo para encontrar posibles fallas dada ciertas condiciones o propiedades.

Dependiendo del conjunto de datos y el problema a resolver se pueden utilizar diferentes algoritmos. En esta sección se hará una introducción a los utilizados en capítulos posteriores.

Clasificación usando Máquina de Vectores

Máquinas de Vector de Soporte o en inglés *Support Vector Machine (SVM)* es un algoritmo de clasificación supervisada. Este algoritmo es muy utilizado hoy en día, gracias a su exitosa aplicación en grandes problemas, superando a los clasificadores no lineales, como las redes neuronales, en muchos contextos (Khorshid et al., 2015). La idea principal de SVM es, a partir de un espacio de características compuesto por el conjunto de datos, encontrar una función que permita dividir en dos o más clases dentro del espacio. Una vez ubicada la función de separación se pueden incorporar datos nuevos al espacio y así predecir su clase correspondiente.

En la figura 2.10 A. se plantea un espacio de características en dos dimensiones con seis individuos (tres clase estrella y tres círculos). Se puede ver que es posible trazar múltiples funciones que dividan el espacio separando las dos clases. SVM busca la función que genere la línea que cuente con los márgenes más amplios. Estos márgenes se calculan a partir de funciones que estén en contacto con al menos un punto en cada clase. En la figura 2.10 B. se denota una función que separa las dos clases con f_0 y f_1 como la distancia de la función hacia los márgenes. El objetivo es encontrar la línea que permita separar los dos grupos y contenga los f_0 y f_1 más grandes posibles.

Desde aquí se utilizará el término *hiperplano* para referirse a la línea que separa las dos clases, ya que en SVM generalmente se trata con espacios de más de dos dimensiones, este término transmite con mayor precisión la idea de un plano en un espacio multidimensional. Se denomina vectores de soporte a los puntos que forman cada uno de los márgenes. Para realizar una formulación matemática de este algoritmo, se debe entender la necesidad de encontrar una fórmula que nos permita calcular, a partir del hiperplano, a qué clase

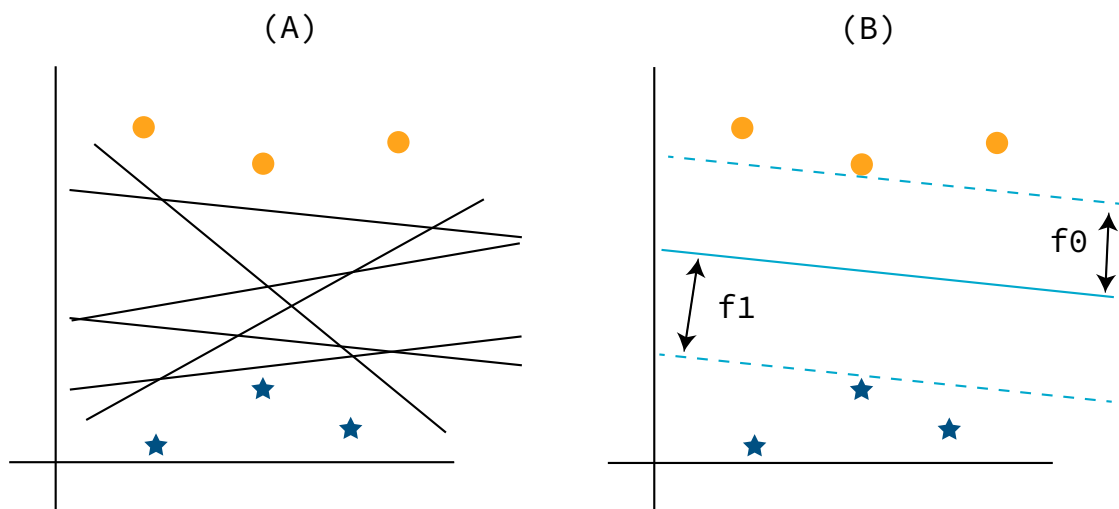


Figura 2.10: Funcionamiento del algoritmo de SVM.

pertenece cada muestra. Para el ejemplo de la figura 2.10, con \vec{V}_0 , \vec{V}_1 como los pesos de los vectores de soporte, y x_0 , x_1 como las coordenadas de la nueva muestra, se puede formular la ecuación de la siguiente manera:

$$f(x) = \vec{V}_0 x_0 + \vec{V}_1 x_1 + b, \quad (2.3)$$

b es el sesgo de la ecuación. El resultado de esta ecuación determina la clase del nuevo individuo. Si $f(x) \geq 1$ pertenece a la primera clase y si $f(x) \leq -1$ a la segunda. Los SVM también se pueden usar cuando los patrones no son linealmente separables. En este caso, los patrones se asignan a un espacio de características diferente en el que se vuelven separables, utilizando una función de *kernel* específica.

Clasificación usando árboles de decisión

Los árboles de decisión son un método de aprendizaje no paramétrico para realizar tareas de clasificación y regresión. Los árboles son construidos a partir de un algoritmo que identifica distintos tipos de divisiones en los datos basados en diferentes condiciones, de forma que generan una estructura abstracta que da una representación de la distribución de los datos. Se denominan árboles porque estos modelos pueden visualizarse como un grá-

fico similar a un árbol con nodos que representan una pregunta (basadas en un atributo); las ramas, las respuestas a la pregunta; y las hojas simbolizan la salida real o la etiqueta de clase. Los árboles de decisión se pueden extender fácilmente para administrar valores continuos, tanto en características como en el resultado. Si el resultado esperado es un valor continuo (generalmente un número real), se llama árbol de regresión (Rokach and Maimon, 2008). Si un atributo es continuo, el nodo correspondiente realiza una partición del espacio del atributo en dos o más subespacios de acuerdo con una función discreta llamada función de corte. En el caso más simple, los nodos internos consideran un solo atributo, lo que implica que la función de corte se limita a subdividir el espacio de acuerdo con el valor del atributo (si el atributo es numérico, generalmente se usa un rango como un límite). Alternativamente, cada hoja puede contener un vector de probabilidad (o afinidad) que indica la probabilidad de que el atributo de salida tenga un cierto valor. La figura 2.11 muestra un ejemplo de la aplicación de árboles para clasificar personas con obesidad abomino-visceral (Elsayed et al., 2008). Uno de los principales problemas de este

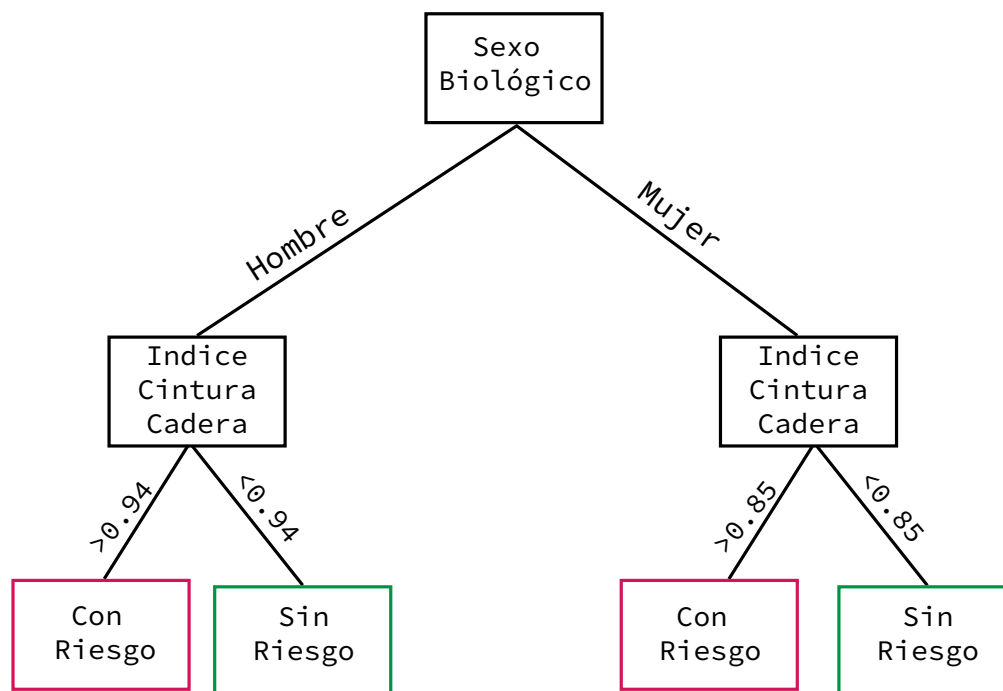


Figura 2.11: Visualización en forma de Árbol. Diferentes valores del índice cintura-cadera junto con el sexo biológico determinar la probabilidad de tener o no riesgo cardiovascular (Elsayed et al., 2008).

modelo, es la selección de los atributos para dividir los datos. Para ello existe una métrica que determina el nivel de incertidumbre que evalúan el comportamiento de los árboles.

Por ejemplo, si se cuenta con un conjunto de datos de 20 muestras, la mitad son de la clase cuadrado y la otra mitad clase círculo. Analizando la figura 2.12, se puede visualizar que si se realiza una primera selección de los datos a partir de un atributo (por ejemplo, color) se genera un conjunto completamente homogéneo (es decir, ese atributo determina completamente la clase) y tiene incertidumbre mínima: mirando una muestra por el atributo, no tenemos dudas de cuál es la clasificación de cualquiera de sus elementos. En este caso, la incertidumbre (o entropía) es 0. Si se analiza una muestra donde el atributo genera conjuntos igualmente distribuidos, como se expone en el árbol de la derecha en la figura 2.12, la incertidumbre es la más alta que se puede obtener. Seleccionar la clase de una muestra equivale a elegir una clase al azar, por lo cual en este caso la incertidumbre es 1.

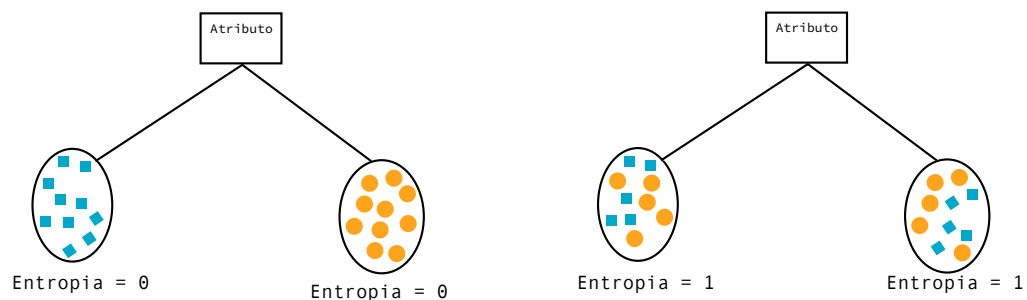


Figura 2.12: Comportamiento de la métrica entropía. El árbol de la izquierda muestra un caso ideal, donde las muestras generan conjuntos homogéneos. El árbol de la derecha es un caso con *incertidumbre* = 1, la selección de la clase es igual a seleccionar una clase al azar.

Se puede definir esta métrica formalmente de la siguiente manera:

$$E(S) = \sum_{n=1}^C -p_i \log_2(p_i), \quad (2.4)$$

donde S es el conjunto de muestras, C es el número de diferentes clasificaciones que usamos, y cada p_i es la proporción de ejemplos que hay de la clasificación i en la muestra.

Existe otro tipo de árboles, los *Extremely Randomized Trees (ERT)* donde tanto sus atributos como la selección de puntos de corte se realizan parcial o totalmente de forma aleatoria. En un caso extremo, un ERT es capaz de construir árboles de clasificación completamente aleatorios cuyas estructuras son independientes de los valores de salida de la

muestra de entrenamiento. La cantidad de aleatoriedad se puede ajustar para obtener comportamientos específicos mediante una elección de parámetro adecuada. Estos parámetros incluyen la cantidad de atributos seleccionados aleatoriamente en cada nodo y el tamaño mínimo de elemento para un nodo. Idealmente, estos parámetros deberían adaptarse al problema específico, ya sea manual o automáticamente (por ejemplo, mediante el uso de un mecanismo de validación cruzada) (Geurts et al., 2006).

2.4.2. Deep Learning

Aprendizaje Profundo o *Deep Learning (DL)* es un tipo de *Machine Learning* donde los modelos utilizados están compuestos por algoritmos de procesamiento multi-capas con el objetivo de aprender representaciones abstractas sobre grandes volúmenes de datos. En estos modelos de DL, cada una de las capas tiene como función extraer nueva información sobre los datos de entrada y generar una nueva representación para la capa siguiente. Este proceso funciona progresivamente hasta que una capa final genere los resultados como salida del algoritmo.

DL actualmente es una tecnología recurrente para resolver distintos problemas en VC. Estas redes obtienen gran popularidad en problemas de reconocimiento de voz, reconocimiento de objetos visuales, detección de objetos y muchos otros dominios. Gracias a los recientes avances sobre clasificación y detección de objetos en imágenes, ha beneficiado a otras áreas de la ciencia como la arqueología, medicina y biología aportando soluciones antes impensadas. En el campo de las imágenes médicas muchos estudios han demostrado resultados prometedores en diagnósticos complejos que abarcan dermatología (Esteva et al., 2017, Haenssle et al., 2018), radiología (Cheng et al., 2016, Cicero et al., 2017), oftalmología (Gulshan et al., 2016, Poplin et al., 2018b) y patología (Cireşan et al., 2013, Charoentong et al., 2017). Los sistemas de aprendizaje profundo podrían ayudar a los médicos ofreciendo segundas opiniones y señalando áreas relacionadas con las imágenes. Un caso particular de algoritmo multi-capas en DL son las redes convolucionales o Convolutional Neural Networks (CNN) (LeCun et al., 2015, Russakovsky et al., 2015).

Proceso de entrenamiento CNN

El éxito para entrenar estos métodos se basa en conceptos como: pesos, función de optimización y función de pérdida. A modo de ejemplo, pensemos que se necesita un clasificador de patologías a partir de imágenes médicas. Aplicando aprendizaje supervisado, y con un gran conjunto de imágenes (tomografías por ejemplo), cada una etiquetada con una patología, se diseña un entrenamiento para una CNN. Durante el entrenamiento, se da como entrada una imagen a la red, la cual produce un vector de salida en donde cada uno de sus componentes corresponde a una categoría que se quiere clasificar. Se quiere lograr que la posición en el vector que tenga el valor más alto sea la que corresponde a la categoría etiquetada (la determinada por el conjunto de datos de entrenamiento). Esto es poco probable que suceda antes del entrenamiento por lo cual debemos calcular mediante una función, el error (o distancia) entre los vectores resultantes y los esperados. A esta función se la denomina función de pérdida. A partir del error calculado se modifican los parámetros internos ajustables para reducir este error en la próxima iteración que la red vea la misma imagen. Estos parámetros ajustables, llamados pesos, son números reales que definen la función de entrada-salida de cada capa de la CNN. En una red, puede haber cientos de millones de estos pesos ajustables, por lo cual es necesario contar con una estrategia para modificarlos. Para ajustar adecuadamente, los algoritmos de aprendizaje calculan un vector de gradiente que, para cada peso, indica en qué cantidad el error aumentaría o disminuiría si el peso aumentará. El vector de peso se ajusta en la dirección opuesta al vector de gradiente.

En la práctica, se utiliza un procedimiento llamado descenso de gradiente estocástico o en inglés *Stochastic Gradient Descent (SGD)* que consiste en ingresar a la red un subconjunto de ejemplos, calcular las salidas y los errores, calcular el gradiente promedio para esos ejemplos y ajustar los pesos en consecuencia. Este proceso se repite para distintos conjuntos de ejemplos hasta que el promedio de la función objetivo deje de disminuir. Este procedimiento simple generalmente encuentra un buen conjunto de pesos sorprendentemente rápido. Además de SGD, existen otros tipos de algoritmos de optimización como por ejemplo ADAM, NESTEROVS y NONE, para la cual cada uno tiene sus particularidades y mejoran su funcionamiento según el problema que se quiere resolver. Después

del entrenamiento, el rendimiento de la red se mide en un conjunto diferente de ejemplos llamado conjunto de prueba. Esto sirve para comprobar la capacidad de generalización del sistema junto con su capacidad de producir salidas acordes para nuevas entradas que nunca ha visto durante el entrenamiento.

A la hora de entrenar las redes un problema común es el sobre-entrenamiento, esto significa que el algoritmo de aprendizaje queda ajustado a unas características muy específicas de los datos de entrenamiento, lo cual determina que el sistema no puede generalizar con datos desconocidos en el proceso de entrenamiento. Es por ello que se utilizan métodos de regularización, los cuales introducen información adicional para impedir el sobre-entrenamiento. En las redes existen varias técnicas de regularización como por ejemplo *regularización por L1 y L2* que básicamente consiste en agregarle un término a la función de pérdida.

A diferencia de los algoritmos de *Machine Learning* vistos anteriormente estos modelos resuelven problemas que requieren que la función de entrada-salida sea invariante a cambios irrelevantes de la entrada, como variaciones en la posición, orientación o iluminación de un objeto pero muy sensible a variaciones minuciosas particulares. Además que los modelos de ML tradicionales requieren de un buen extractor de características que produzcan representaciones que sean selectivas a los aspectos de la imagen. Los cuales son importantes para la discriminación, pero invariantes a aspectos irrelevantes. Todo el preprocesamiento puede evitarse si las buenas características se pueden aprender automáticamente mediante un procedimiento de aprendizaje de propósito general. Esta es la ventaja clave del aprendizaje profundo.

Arquitecturas de red

La arquitectura de las redes neuronales se construye a partir de 3 tipos de capas; Capa de entrada: datos iniciales para la red neuronal; Capas ocultas: capa intermedia, el lugar donde se realiza todo el cálculo; y por último la Capa de salida: produce el resultado (Ver figura 2.13). Cada capa contiene una cantidad de nodos los cuales combinan las entrada de los datos con los pesos, que amplifican o amortiguan esa entrada, asignando así importancia

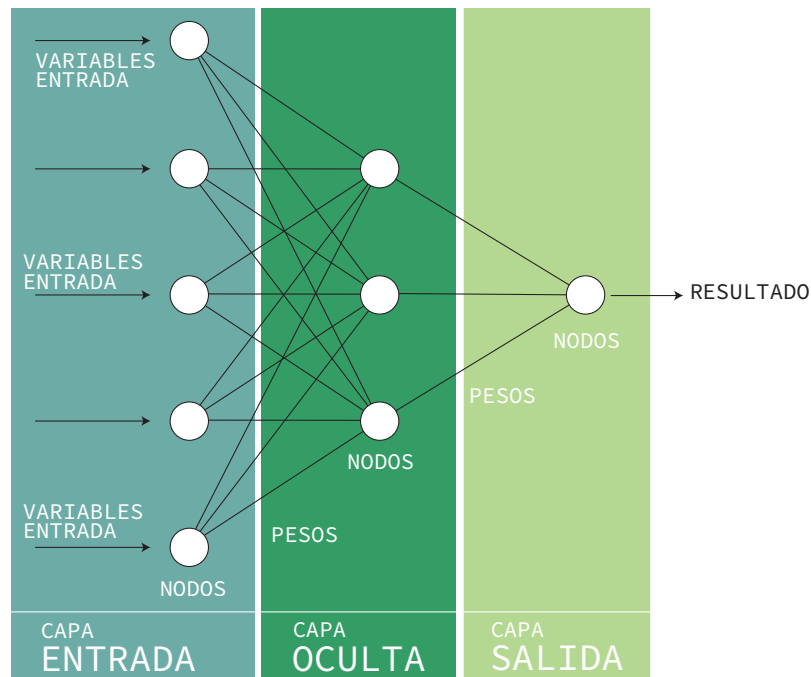


Figura 2.13: Arquitectura de capas en redes neuronales.

a las entradas con respecto a la tarea que el algoritmo está tratando de aprender. Estos pesos de entrada se promedian y utilizando una función de activación se determina si esa entrada debe progresar o no a través de la red y en qué medida deben afectar el resultado final. Existen diferentes funciones de activación que pueden utilizarse en cada nodo, si bien hay un abanico inmenso de funciones podemos citar algunas de las más utilizadas *ReLU*, *leakyReLU*, *sigmoid*, *softmax* y *TANH*. Como este tipo de capas contiene nodos totalmente conectados entre sí, se las denomina capas *fully connected*. Particularmente las redes neuronales convolucionales son redes multi-capas que trabajan con imágenes aplicando operaciones de convolución. La necesidad de su existencia está dada a que la utilización de capas totalmente conectadas (como las explicadas anteriormente) sería inviable, ya que solo con una imagen 100x100 se obtiene 10,000 nodos para cada capa, lo cual es un número considerable incluso en una red de poca profundidad. La operación de convolución brinda una solución a este problema, ya que reduce el número de parámetros, permitiendo que la red contenga un conjunto mayor de capas (más profundidad) con menos pesos para ajustar. Este tipo de red consta, de al menos tres tipos de capas, las *fully connected*, las convolucionales y las capas de *pooling*.

Capa convolucional. En estas capas se realizan operaciones matriciales entre los datos de entrada (puede ser la salida de una capa anterior) y una matriz (*kernel* o filtro) definidos en la capa actual. Se selecciona un conjunto de pixeles vecinos (definido por los parámetros de la capa) en la imagen inicial y se opera matemáticamente con el *kernel* a medida que el mismo se desplaza sobre toda la imagen. En este proceso se obtiene una nueva matriz (imagen) filtrada por el *kernel*. Generalmente existen casos donde no se aplica solo un *kernel*, sino que se aplican un conjunto (denominado filtros), por lo cual se obtiene una salida de más de una matriz. A este conjunto de matrices resultantes se lo denomina *feature mapping*. La entrada a este tipo de capa se puede ejemplificar como una matriz de $n \times h \times w \times p$ donde n es el número de imágenes, h y w la altura y el ancho de la imagen y p la profundidad. Luego de ingresar a una capa convolucional, la imagen se abstrae en un *feature mapping* de $n \times k \times s \times r$, donde n es el número de imágenes, k , s y r son la altura, ancho y profundidad del mapa de *features*. Una capa convolucional dentro de una red neuronal debe contar con los siguientes atributos: *kernels* convolucionales definidos por un ancho y alto (hiper parámetros), el número de canales de entrada y canales de salida (hiper parámetro). También debe contar con la profundidad del filtro de convolución (los canales de entrada) los cuales deben ser igual a la cantidad de canales (profundidad) del mapa de características (o imagen) de entrada.

Capa de pooling. En algunos caso es necesario contar con capas que permita re-dimensionar los datos, ya sea para utilizar como capas intermedia dentro de la red o para generar la salida con los tamaños deseados. La re-dimensión debe hacerse teniendo en cuenta las características más importantes que se detectaron para cada filtro, es decir no puede realizarse al azar, se necesita una estrategias para ello. Existen varias alternativas, aquí se detallará las más conocidas que consiste en combinar las entradas de una capa mediante una operación básica. Dentro de la capa se combina la entrada en pequeños grupos, típicamente 2×2 . Para la capa denominada *MaxPooling* (Yamaguchi et al., 1990), la selección del valor de retorno se realiza quedándose con el valor máximo de ese grupo, para *AveragePooling* (LeCun et al., 1998) se retorna un valor promedio. De esta manera, si se utiliza una vecindad 2×2 , la capa reduce las dimensiones de los datos a la mitad.

Aprendizaje por transferencia

El diseño de la arquitectura es un aspecto fundamental para determinar el éxito de la red neuronal. Estas mismas redes han demostrado un buen rendimiento en el aprendizaje por transferencia (Yosinski et al., 2014). Este concepto permite que una CNN probada y entrenada en un conjunto de datos masivo que no está relacionado con la tarea de interés (por ejemplo, ImageNet2, un conjunto de datos de millones de objetos de uso cotidiano) se ajuste a un conjunto de datos más pequeño (por ejemplo, un conjunto de imágenes médicas). El algoritmo aprovecha grandes cantidades de datos para aprender las estadísticas naturales en imágenes (líneas rectas, curvas, coloraciones, etc.) y en el segundo paso, las capas de nivel superior del algoritmo se vuelven a entrenar.

Finalizado el marco teórico, los próximos capítulos se centraran en solucionar, utilizando las tecnologías ya descritas, diferentes problemas en el campo de la caracterización del fenotipo humano. En el capítulo tres se detallan los protocolos de captura de junto con la descripción de los datos utilizados a lo largo de toda esta investigación.

Capítulo 3

Colecta de datos

Como parte del desarrollo inicial de esta investigación se llevó adelante la creación y ejecución de dos grandes proyectos que tuvieron como objetivo principal, construir un banco de datos en torno a la población de Puerto Madryn. En este capítulo, se describe en detalle este banco de datos: su composición y características como así también los protocolos utilizados y las campañas de divulgación.

3.1. Proyecto Patagonia 3D Lab

El proyecto *Patagonia 3D Lab* comenzó como una idea conjunta del GIBEH (Grupo de Investigación en Biología Evolutiva Humana) y el LCI (Laboratorio de Ciencias de las Imágenes, Universidad Nacional del Sur) con el objetivo principal de desarrollar una aplicación móvil que fuera capaz de crear modelos 3D del cuerpo humano aprovechando características de hardware de los dispositivos móviles (celulares y tablets) y su presencia masiva en la vida cotidiana. Por esta razón, la aplicación sería capaz de realizar análisis volumétricos sobre los modelos 3D, obteniendo diversos índices que ayuden a profesionales de la salud en el diagnóstico y tratamiento de múltiples patologías haciendo hincapié en el sobrepeso como problemática epidémica central.

A principios de 2014 cuando se originó este proyecto, las estadísticas señalaban que más de 1900 millones de adultos en todo el mundo tenían sobrepeso. El aumento del peso

corporal promedio, tanto en países desarrollados como emergentes, está relacionado con un cambio de paradigma en la alimentación. Realizar un diagnóstico temprano y seguimiento de la evolución del estado nutricional del paciente resulta fundamental, pero actualmente las evaluaciones se basan en la toma de un conjunto de medidas antropométricas que demandan tiempo y contienen un margen de error importante. La automatización de estas tareas, a través del registro de imágenes tridimensionales, permitiría obtener mejores resultados. En la actualidad, la obtención automática de medidas antropométricas puede ser llevada a cabo por medio de la utilización de dispositivos de escaneo. Por desgracia, el mercado ofrece pocas o nulas opciones asequibles para las instituciones de salud que las necesitan, imposibilitando su obtención y utilización en la práctica médica habitual. Como grandes desventajas, estos dispositivos suponen un gran desembolso económico para su adquisición y un tiempo de entrenamiento considerable para el manejo del software asociado. Es por eso que la alternativa propuesta por los miembros del proyecto fue desarrollar una aplicación vinculada a un servicio en la nube, capaz de crear modelos 3D del cuerpo del paciente mediante el uso de un dispositivo móvil con cámara (teléfono o tablet) en combinación con algoritmos avanzados de visión computacional. Como resultado, esta aplicación permitiría obtener variables antropométricas para determinar y cuantificar la forma corporal de manera automática mediante un procedimiento simple, rápido, no invasivo y sobre todo, accesible. La solución presentada fue seleccionada dentro de una convocatoria impulsada por el Centro de Innovación Tecnológica Empresarial y Social (CITES). Esta organización es una aceleradora e incubadora de empresas del Grupo SanCor Seguros, que brinda apoyo financiero y recursos a emprendedores tecnológicos. En la edición 2016 de su Programa CITES Ideas¹ se presentaron 122 proyectos provenientes de diversos puntos del país, Latinoamérica, Canadá y Estados Unidos. Quedaron diez finalistas y dos resultaron ganadores, siendo uno de ellos *Patagonia 3D Lab*.

¹CITES Ideas premia a startups innovadoras que contribuyen al desarrollo de soluciones para mercados regionales y globales



Figura 3.1: Campaña de convocatoria en redes sociales.

3.2. Campaña de Voluntarios

La financiación obtenida como premio de aquel programa permitió llevar a cabo una campaña de recolección de datos en la ciudad de Puerto Madryn. Los voluntarios fueron convocados mediante divulgación a través de redes sociales y medios de comunicación locales (Ver Figura 3.1). A cada persona se le contaba acerca de los objetivos del proyecto en una Cartilla de Información, véase en Apéndice A. Si la persona estaba de acuerdo, se procedía a firmar un Término de Consentimiento Libre e Informado, en el que expresaba su voluntad de participar. Toda la documentación y el procedimiento de registro, análisis y almacenamiento de la información contó con el aval del Comité de Bioética del Área Programática Norte, Chubut. El documento de aval de este proyecto puede observarse en el Apéndice B .

3.3. Protocolo de recolección de datos

La campaña de recolección de datos consistió en la toma de medidas antropométricas clásicas e índices asociados al estado nutricional de los voluntarios, acompañado por la toma de imágenes 3D por medio de un dispositivo de escaneo. El grupo de voluntarios contó con 154 individuos en total con un promedio de edad de 39 ± 13 años. El escáner utilizado fue *Structure Sensor* (desarrollado por *Occipital Inc., San Francisco, California*), el cual fue fabricado como un accesorio de uso general para dispositivos iPad (*4ta generación, Apple Inc., Cupertino, California*). Su funcionamiento se basa en la proyección de luz infrarroja (más información en Capítulo 2) y alcanza una precisión de entre $0,5mm$ y $40cm$. El escáner contiene una cámara VGA con una resolución $3-8MP$ y $10-50fps$. Su precisión y exactitud para los estudios del cuerpo humano fue puesta a prueba en comparación con otros escáneres 3D y tomografías computadas, confirmando que se trata de un dispositivo preciso y efectivo para describir detalles anatómicos en áreas redondeadas, incluso en los senos femeninos, la forma de la cabeza, las mejillas y el contorno de la barbilla ([Knoops et al., 2017](#), [Oranges et al., 2019](#)). El procedimiento de adquisición de imágenes 3D fue realizado por el mismo operador en todos los casos siguiendo un protocolo establecido. Para conseguir una representación fiel de la forma corporal de los voluntarios, cada uno de ellos debió vestirse con ropa deportiva o lo más ajustada posible. Se le solicitó a cada voluntario que adoptara una posición específica con sus brazos en perpendicular a su cuerpo (90° del eje sagital) y con sus pies separados, en paralelo, a $40cm$ uno del otro. Luego, se le pidió que respirara lo más lento posible, para evitar o minimizar la introducción de ruido al modelo reconstruido provocado por los movimientos abdominales y del tórax. Mientras tanto, el operador del escáner debía sostener el dispositivo e ir moviéndose alrededor de los voluntarios a $1m$ de distancia, capturando las imágenes necesarias para lograr un registro completo de 360° (Ver Figura 3.2). Este mismo procedimiento se repitió para registrar un vídeo alrededor del voluntario utilizando un teléfono móvil.

Para la obtención de los datos antropométricos, se siguió un protocolo estandarizado ([Cabañas and Esparza, 2009](#)) que incluye (a) altura total, utilizando la cinta métrica Seca 206 (Seca GmbH & Co Kg, Hamburgo, Alemania), con una longitud de graduación de $1mm$ y un rango de 0 a $220cm$; (b) peso total y composición corporal (masa mus-

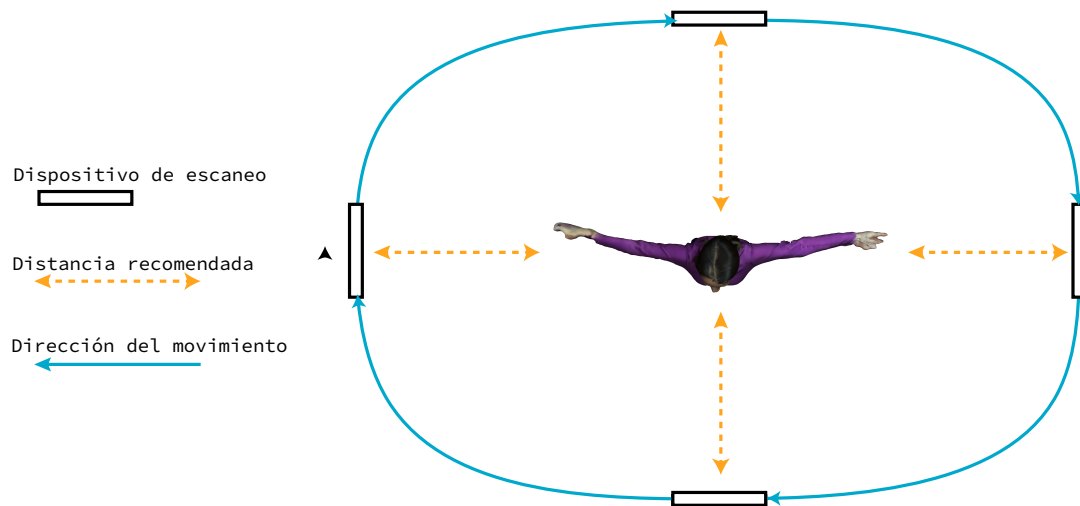


Figura 3.2: Protocolo de captura de datos en 360. Vista superior.

cular, masa libre de grasa y grasa corporal, y porcentajes) estimados con una escala de bioimpedancia (Tanita BC 1100F); (c) circunferencias de muslo, bíceps, cadera y cintura, utilizando cinta métrica ergonómica Seca 201 (Seca GmbH & Co Kg, Hamburgo, Alemania) con una división de $1mm$ y un rango de 0 a $205cm$ y (d) pliegues subescapular y bicipital, con un calibre de precisión Vernier con un rango hasta $0,02mm$ a $200mm$. Estas mediciones se registraron en el lado derecho del cuerpo de los voluntarios. Por último, para controlar el posible error intra-observador, todas las mediciones se obtuvieron dos veces, tolerando una diferencia entre ambas mediciones de no más de $0,5cm$.



Figura 3.3: Protocolo de medición.

3.4. Datos Artificiales

Además de los escaneos y medidas a los voluntarios, se generaron 450 modelos artificiales que respetan la misma posición corporal utilizada por los voluntarios en la muestra real. Estos modelos fueron creados por un artista utilizando el software *Poser Studio* proporcionando parámetros de los diferentes tipos de aspectos morfológicos presentes en una muestra de una población real (por ejemplo diferentes alturas, estados nutricionales, características físicas, edades, etc.). Como resultado, la muestra artificial quedó conformada por 118 modelos (36 hombres, 82 mujeres) que representan personas con sobrepeso u obesas y 332 (190 hombres y 142 mujeres) que representan otras condiciones físicas: promedio, atlético, bajo peso.

3.5. Configuración de landmarks

Para enriquecer aún más los datos recolectados, se colocó una serie de *landmarks* en cada uno de los modelos 3D, incluidos modelos reales y artificiales (Ver Figura 3.4). Este conjunto de *landmarks* se corresponden con puntos ubicados anatómicamente (Ver

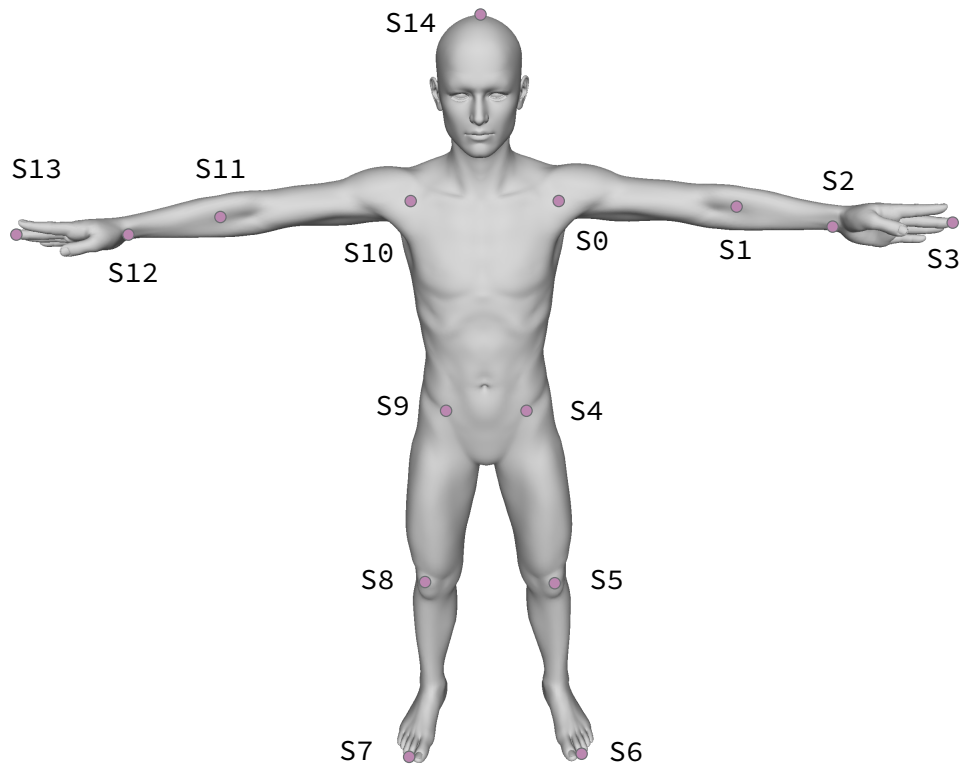


Figura 3.4: Configuración de *landmarks* corporales 3D sobre un modelo artificial.

Tabla 3.1) que pueden ser identificados repetitivamente y con precisión a lo largo de todos los individuos. Este pequeño conjunto de puntos abre la posibilidad a nuevas investigaciones, así como a nuevos procedimientos que permitan contar con un sustento anatómico.

Número	Definición anatómica
S0, S10	Articulación del húmero.
S1, S11	Extremo proximal del radio.
S2, S12	Punto extremo distal del radio.
S3, S13	Punto extremo del dedo mayor.
S4, S9	Punto de articulación de la cabeza del fémur en la cadera
S5, S8	Punto medio de la rótula
S6, S7	Extremo distal segundo dedo
S14	Punto extremo de la cabeza

Tabla 3.1: Definición anatómica de los 15 *landmarks*.

3.6. Proyecto Raíces

Habiendo finalizado la primera etapa del proyecto *Patagonia 3D Lab* y con el éxito obtenido durante la convocatoria de voluntarios, se pensó una campaña para continuar con el objetivo de generar un banco de datos provenientes de diversas fuentes (imágenes corporales en 2D/3D, datos genéticos y de estilo de vida) conformando una muestra representativa de la población patagónica. Es por esto que el Instituto Patagónico de Ciencias Sociales y Humanas (IPCSH-CONICET) puso en marcha el proyecto Raíces. En su etapa inicial, Raíces convocó en Puerto Madryn a voluntarios mayores de 18 años de edad con la condición de ser hija o hijo de personas nacidas en Argentina o en alguno de sus países limítrofes.

A la fase de recolección de imágenes y modelos 3D del cuerpo completo de los voluntarios se le sumó un escaneo facial, una encuesta sociocultural y una extracción de sangre por punción venosa. La incorporación de estos nuevos tipos de datos permitió conocer la variabilidad genotípica, fenotípica y sociocultural de la muestra, las posibles asociaciones, así como también la realización de una estimación de la ancestría genética de cada voluntario. Los nuevos procedimientos y protocolos para estos nuevos datos fueron incluidos en un nuevo Aval del Comité de Bioética del Área Programática Norte, Chubut. Como resultado, se recolectaron datos de un total de 144 voluntarios, 70,2% mujeres y 29,8% hombres. Si bien estos datos no fueron incorporados totalmente al objetivo de esta tesis, como ya se dijo anteriormente, la existencia y disponibilidad de este biobanco es un caso único en Argentina y será considerado una parte fundamental para el trabajo futuro de la presente investigación.

Los datos recolectado en el Proyecto Patagonia 3D Lab fueron utilizados para el diseño y entrenamiento de los modelos en los siguientes capítulos. Además para el capítulo cuatro, se utilizaron los modelos artificiales para obtener una mejor generalización en los algoritmos.

Capítulo 4

Reconocimiento automático de sexo biológico utilizando capturas 3D

En este capítulo se presenta un clasificador de sexo biológico basado en el análisis de mallas 3D del cuerpo humano. Se describen las técnicas de pre-procesamiento utilizadas, junto con los algoritmos de aprendizaje supervisado que permiten resolver la tarea de clasificación. Finalmente, se realiza una comparación entre los resultados obtenidos con dos modelos de aprendizaje diferentes: máquinas de vectores de soporte y árboles de decisión.

4.1. Introducción

En múltiples escenarios se requiere contar con mediciones antropométricas precisas de una manera rápida, confiable y económica, incluyendo aspectos como: forma, tamaño, perímetro y volumen de diferentes partes del cuerpo humano. En la práctica médica, un apropiado estudio de la forma corporal en relación con el sexo biológico del paciente es clave para el diagnóstico, tratamiento y seguimiento clínico apropiado de diversas condiciones como la detección de malformaciones, cánceres específicos, entre otros ([Treleaven and Wells, 2007](#)). Para ser más precisos, el sexo es un factor principal al analizar la ubicación del tejido adiposo. De hecho, la obesidad central tiende a ser un fenotipo masculino. Las mujeres son, en general, más eficientes en la división de la grasa en la periferia del cuerpo

en comparación con los hombres (Geer and Shen, 2009, Power and Schulkin, 2008). Este dimorfismo sexual aparece en parte debido al efecto de las hormonas sexuales en los cuales, se observan varios cambios en el patrón de obesidad y los consiguientes factores de riesgo asociados, después de la menopausia. Además, los factores genéticos y ambientales también contribuyen a las diferencias étnicas en la distribución de grasas (Guglielmi and Sbraccia, 2018). Al mismo tiempo, en otros contextos (por ejemplo, comerciales y publicitarios) puede utilizarse la identificación del sexo biológico para generar mejores indicadores del público objetivo. Ejemplos de ellos son: el campo del diseño y venta de indumentaria en línea (Azouz et al., 2006, Daneshmand et al., 2016), deportología y estudios ergonómicos (Lovato et al., 2014).

Actualmente, los avances tecnológicos en el campo de la visión computacional en combinación con el aprendizaje de máquina, permiten facilitar y acelerar el procedimiento de detección o reconocimiento del sexo biológico. La estrategia habitual para el reconocimiento de esta característica es el análisis de imágenes y vídeos utilizando los rasgos faciales sobresalientes que permiten distinguir entre mujeres y hombres adultos (por ejemplo, Moghaddam and Yang (2002)). En Li et al. (2008), los autores analizaron en imágenes, las siluetas de cuerpos enteros. La ventaja de estos enfoques es que el único dato que se requiere es una proyección 2D, la cual puede ser obtenida de vídeos o imágenes utilizando técnicas de segmentación. (Shanthi.G.C and Saravanan.E, Kim et al., 2005, Cao et al., 2008).

El análisis 3D se ha comenzado a utilizar recientemente, en relación con la madurez de los métodos basados en nubes de puntos. En Fouts et al. (2012), Shan (2012), Moghaddam and Ming-Hsuan Yang (2000), el reconocimiento de sexo se logra utilizando histogramas obtenidos del análisis de densidad de diferentes segmentos de nubes de puntos 3D. Aunque la extracción y clasificación de características en nubes de puntos sigue siendo problemática, es una buena aproximación. Otra estrategia es la de realizar un reconocimiento de sexo basado en la triangulación de la nube de puntos, o con la utilización directa de la malla (Blaha et al., 2017).

En Tang et al. (2011), se realizan diferentes análisis sobre modelos 3D utilizando múltiples mediciones del área del pecho y el torso del individuo como datos de entrenamiento

para clasificadores. Este análisis excluye personas con sobrepeso o completamente vestidas, con lo cual excluyen algunas características del cuerpo completo. En el presente capítulo se detallaran técnicas de pre-procesamiento junto con algoritmos de aprendizaje supervisado utilizados en la construcción de un nuevo algoritmo que permite resolver la tarea de clasificación de sexo biológico. Por último, se elabora una comparación entre los resultados obtenidos con dos modelos de aprendizaje diferentes: Support Vector Machine (SVM) y Extremely Randomized Trees (ERT).

4.2. Materiales y Métodos

4.2.1. Descripción de la muestra

Como dato de entrada se utilizó el conjunto de modelos artificiales del cuerpo humano previamente mencionados en el Capítulo 3. Dicho conjunto cuenta con 223 cuerpos femeninos y 255 cuerpos masculinos que comparten la misma pose, piernas separadas a una distancia fija y brazos extendidos al nivel de los hombros como puede verse en la figura 4.1. También se realizaron pruebas utilizando modelos 3D de personas reales que participaron como voluntarios del mismo proyecto. Para la adquisición de estos últimos modelos, se realizaron escaneos de un total de 151 adultos (111 mujeres y 40 hombres) con el dispositivo *3D Structure Sensor* (Ver figura 4.2).

4.2.2. Extracción de características y entrenamiento

Si bien cualquier algoritmo de aprendizaje de máquina podría ser entrenado con un conjunto de datos sin procesar, la evidencia empírica demuestra que el hecho de utilizar técnicas de preprocesamiento conocidas mejora significativamente tanto el rendimiento como la calidad de los resultados obtenidos. Con este objetivo en mente, a continuación se detalla el flujo de pre-procesamiento utilizado para entrenar el algoritmo de clasificación propuesto. El primer paso consiste en obtener las coordenadas de los vértices x , y , z de los cuerpos 3D (tanto el conjunto de entrenamiento como los de prueba). Para que la muestra

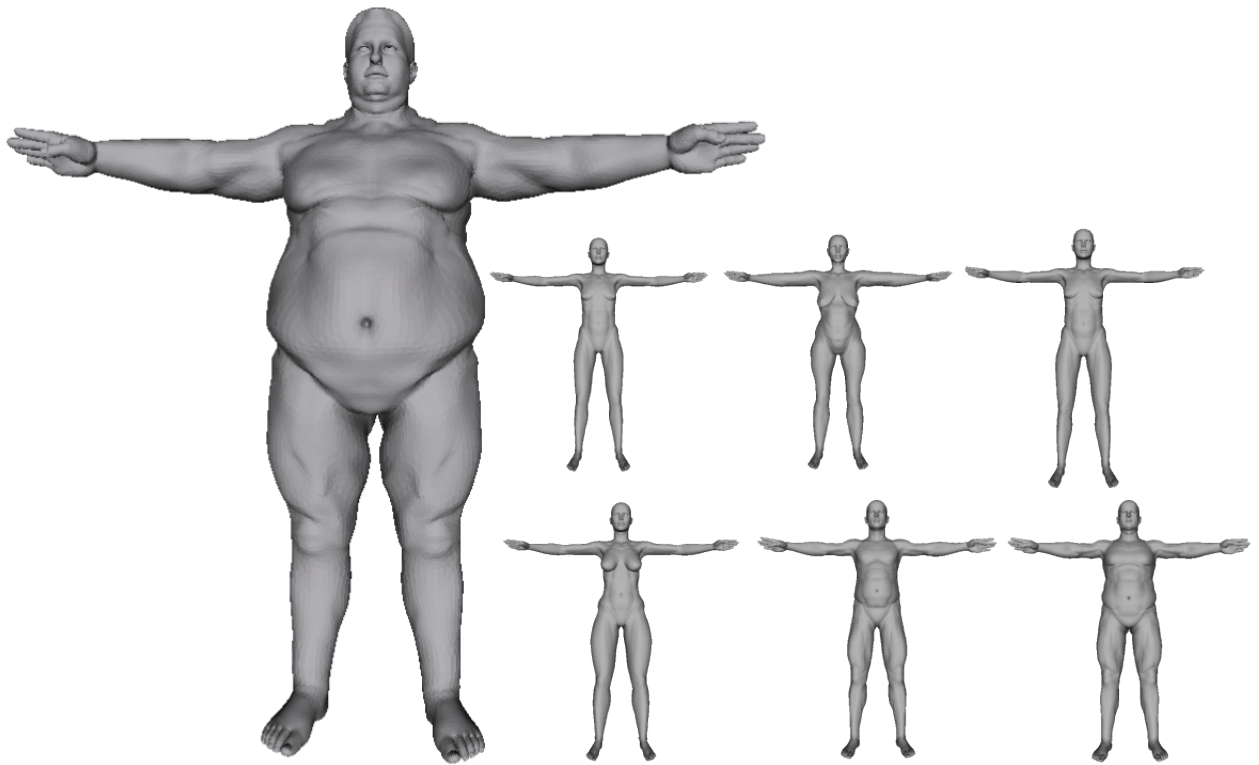


Figura 4.1: Modelos artificiales del dataset (Navarro et al., 2018).

tenga una variabilidad representativa, los modelos utilizados en este trabajo tienen diferentes características antropométricas, lo que altera la cantidad de vértices requeridos para conformar las mallas 3D de los diferentes cuerpos. En particular, las mallas del conjunto de datos cuentan con un número de vértices que oscila entre 67290 y 68300 según las diferencias en forma de los cuerpos que representan. Esto provocará varios problemas durante los pasos posteriores, porque tanto el tamaño del espacio de características como la importancia geométrica de cada uno de los vértices utilizados durante el proceso de aprendizaje del clasificador será diferente. Para solventar estas dificultades, se propone una estrategia que consiste en utilizar un número estándar de vértices para todos los modelos 3D a través de un mecanismo conocido como *resampling*. Luego de la adquisición inicial, cada malla (real y artificial) vuelve a calcular sus vértices usando una distribución de probabilidad uniforme (Vitter, 1984), la cual garantiza que todos los puntos tengan la misma probabilidad de ser elegidos. Finalmente, cada modelo contará con la cantidad más baja de vértices registrada en las mallas de conjunto de datos ($|V| = 67290$).

La naturaleza del problema conduce a tener espacios de *features* de grandes dimensiones, lo cual el rendimiento de los clasificadores puede verse perjudicado. De este modo, se

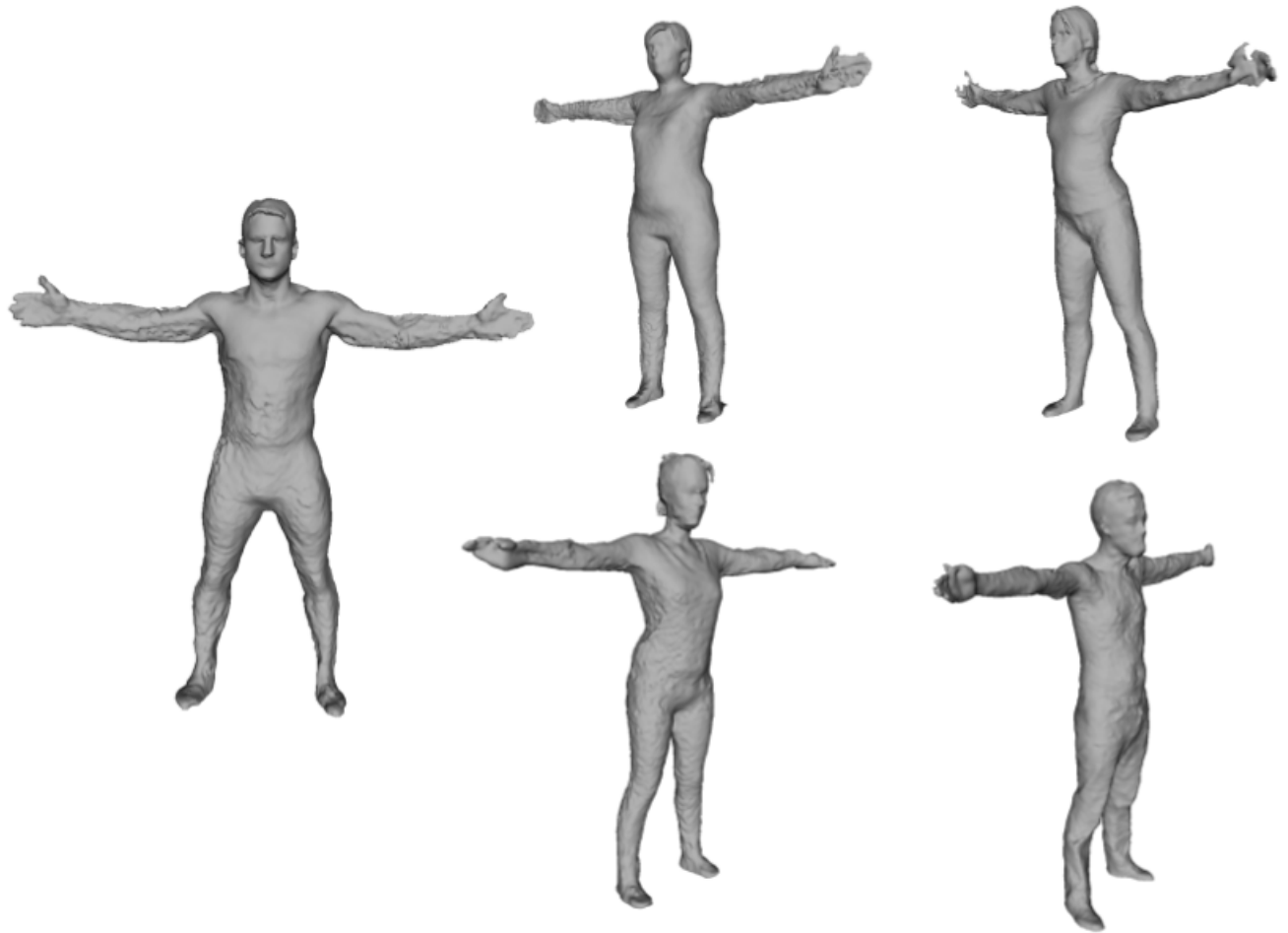


Figura 4.2: Modelos reales del dataset (Navarro et al., 2018).

requiere algún medio para reducir la complejidad intrínseca de los datos (Fodor, 2002) (y por lo tanto de las formas) eliminando las redundancias y reduciendo el tamaño del espacio de características a dimensiones convenientes. Además es necesario aplicar un proceso de normalización para que el clasificador no sea sensible a valores extremos, sesgando sus decisiones con respecto a las medidas unitarias de los *features* (LeCun et al., 2012). Existen varios enfoques para reducir la dimensión de los espacios, en este capítulo se utiliza el Análisis de Componentes Principales (PCA) ya que es una de las técnicas más utilizadas a la hora de reducir los tamaños de espacios de características. Para más información sobre PCA, ver Capítulo 2. En particular, se realizó un análisis PCA modificando el parámetro de cantidad de componentes principales hasta lograr un tamaño aceptable que explique un buen porcentaje de variación de los datos de entrenamiento y comparando los resultados del clasificador con la métrica *F1-score* (Ver Figura 4.3. Finalmente se optó con un con-

junto de 350 componentes principales, capaces de explicar aproximadamente el 80 % de la varianza en la muestra y obteniendo un valor de $F1$ -score prometedor.

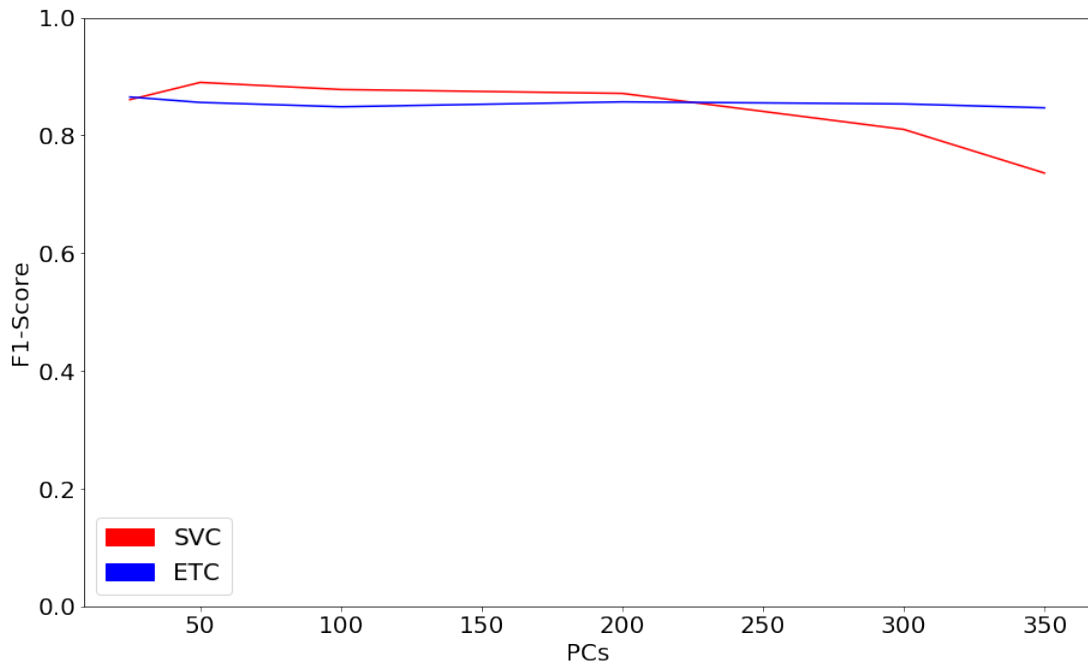


Figura 4.3: $F1$ -score usando diferentes cantidades de componentes para dos clasificadores (Navarro et al., 2018).

4.3. Experimentos

Se comparó el desempeño de dos clasificadores, *Support Vector Machine (SVM)* y *Extremely Randomized Trees (ERT)*. Para este capítulo se utilizaron dos implementaciones de estas técnicas *Support Vector Classifier (SVC)* y *Extra Trees Classifier (ETC)*. Ambos pertenecientes a la biblioteca de *Machine Learning: Scikit-learn* (Pedregosa et al., 2011b) desarrollada en *Python*.

4.3.1. Selección y optimización de hiper parámetros

La gran mayoría de los algoritmos de aprendizaje automático permiten al usuario ajustar la configuración del modelo por medio de parámetros que luego impactarán en los resultados. Una selección adecuada de estos parámetros puede convertirse en una difícil tarea si el propio espacio de parámetros es complejo. Por esta razón, es importante

encontrar una solución automatizada y replicable (Snoek et al., 2012). Dentro de las más aplicadas en la actualidad encontramos Búsqueda Aleatoria (*Random Search*) (Bergstra and Bengio, 2012) y Búsqueda de Cuadrícula (*Grid Search*). A continuación, se describirá la última de ellas debido a que una solución aleatoria no es adecuada para una selección correcta de los parámetros.

Grid Search se define como un método que especifica un espacio de hiper parámetros que determina las variables que se quieren optimizar. Usando un algoritmo de búsqueda exhaustiva (o fuerza bruta), esta técnica selecciona los valores indicados para cada hiperparámetro y compara el desempeño del algoritmo de aprendizaje utilizándolos con una métrica en particular. Naturalmente, esta técnica es computacionalmente costosa por lo que se definen combinaciones de parámetros acotados determinando un espacio reducido de búsqueda. Dado que la exploración de los posibles resultados se puede realizar de forma independiente, es fácil de paralelizar (Hsu et al., 2003).

En particular, la búsqueda del espacio de parámetros para el algoritmo ETC se definió utilizando un número de estimadores en un rango de valores entre 60 y 100, las características máximas en un rango de valores entre 2 y 350 y por último, la profundidad máxima en valores entre 1 y 10. Finalizado el proceso de búsqueda, el mejor rendimiento obtenido para ETC arrojó un *F1-score* del 89 %, utilizando 70 estimadores, junto con un tope máximo de 97 características y una profundidad máxima de 4. Para el algoritmo SVC, el espacio de búsqueda apuntó a determinar con qué funciones de kernel el clasificador obtenía mejores resultados, determinando que el kernel basado en la función radial obtuvo el mejor rendimiento, con un *F1-score* del 71 %.

4.3.2. Selección de datos de Entrenamiento.

Se llevó a cabo una división inicial del conjunto de datos original compuesto por modelos artificiales, reservando un 75 % para la tarea de entrenamiento y el 25 % restante para la etapa de prueba. A su vez, este último grupo fue dividido en estratos de tal forma que se asegure que cada división del dataset sea representativa para cada clase, es decir que cada estrato esté balanceado en términos de cantidad de ejemplos disponibles para cada clase.

Fold	Precision	Recall	F1-score
1	0.79	0.75	0.74
2	0.70	0.69	0.68
3	0.77	0.75	0.74
4	0.67	0.66	0.65
5	0.72	0.72	0.72
Avg/Total	0.73	0.71	0.71

Tabla 4.1: SVC, promedio por corrida (Navarro et al., 2018).

Fold	Precision	Recall	F1-score
1	0.90	0.90	0.90
2	0.89	0.88	0.88
3	0.91	0.91	0.91
4	0.90	0.88	0.88
5	0.86	0.85	0.85
Avg/Total	0.89	0.88	0.88

Tabla 4.2: ETC, promedio por corrida (Navarro et al., 2018).

Se utilizaron 5 estratos que fueron obtenidos utilizando el algoritmo *Stratified-K-Fold*, lo que garantiza un tamaño de datos relativamente bajo, beneficiando el rendimiento durante la tarea de selección. Para definir los conjuntos de datos que se utilizarían para entrenar el algoritmo se utilizó el mecanismo de Validación Cruzada (CV), procesando cada uno de los estratos por separado y evaluando el rendimiento del algoritmo utilizando el conjunto de prueba. El proceso de CV, toma cada una de los estratos como parámetros de entrada y ejecuta una cierta cantidad de corridas (en este caso 10), para garantizar que el resultado del algoritmo de clasificación seleccionado sea independiente de la partición de datos realizada (entrenamiento y prueba). Pudo observarse que la combinación de *Stratified-K-Fold* y CV, permite lograr un buen equilibrio sesgo/varianza del clasificador y mejorar sus resultados con respecto al CV simple como método para selección de muestra Kohavi et al. (1995). En las Tablas 4.1 y 4.2 se ilustran los resultados de este proceso cuando se aplica a los dos algoritmos seleccionados, ETC y SVC.

Time	SVM	ETC
Pipeline	1min 10s ± 340 ms	4min 12s ± 342 ms
Train Batch (313 indiv.)	233 ms ± 3.09 ms	4.24 s ± 132 ms
Prediction Batch (135 indiv.)	14.4 ms ± 226 us	16.9 ms ± 199 us
Individual Prediction	169 us ± 813 ns	12.9 ms ± 84.3 us

Tabla 4.3: Tiempos de entrenamiento y predicción (Promedio de 1000 corridas) (Navarro et al., 2018).

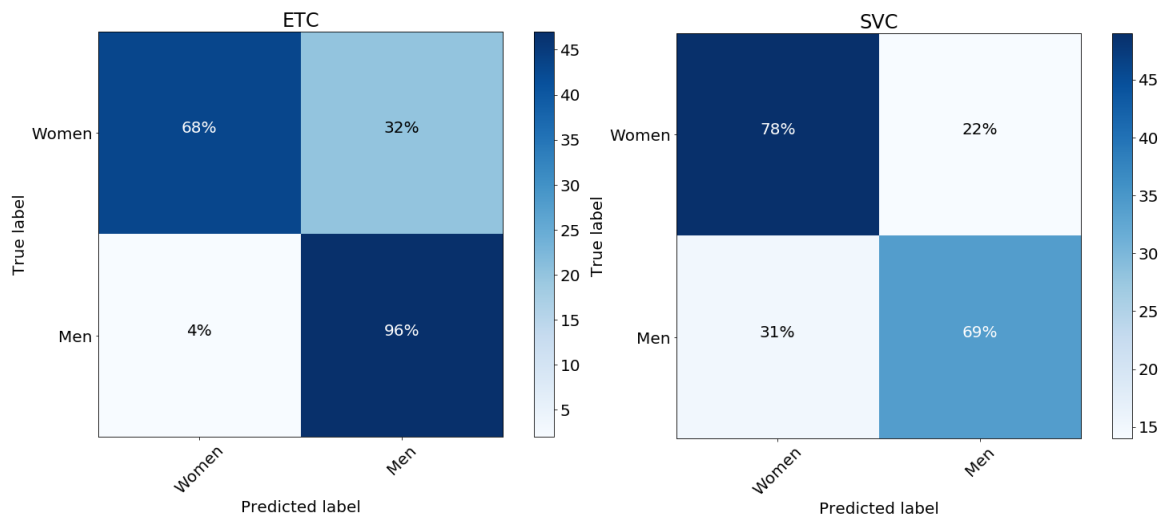


Figura 4.4: Matriz de confusión obtenida con el conjunto de prueba (Navarro et al., 2018).

4.3.3. Tiempo de entrenamiento y predicción

Los tiempos promedio de entrenamiento y predicción se calcularon utilizando ambos algoritmos y siguiendo el *pipeline* antes mencionado. Estos resultados pueden observarse en la tabla 4.3. En términos generales, SVM resultó más veloz que ETC.

4.4. Resultados

El conjunto de datos destinados a pruebas (25 % del conjunto de datos original) fue utilizado para realizar una serie de experimentos, debido a que contiene datos nunca antes vistos por los clasificadores entrenados reflejando así, el comportamiento de estos algoritmos en escenarios de uso más cercanos a la realidad. El algoritmo ETC obtuvo los mejores resultados en términos de clasificación con un 85 % de precisión y un 80 % de exactitud, mientras que el SVC logró un 71 % de precisión y un 76 % de exactitud. Adicionalmente,

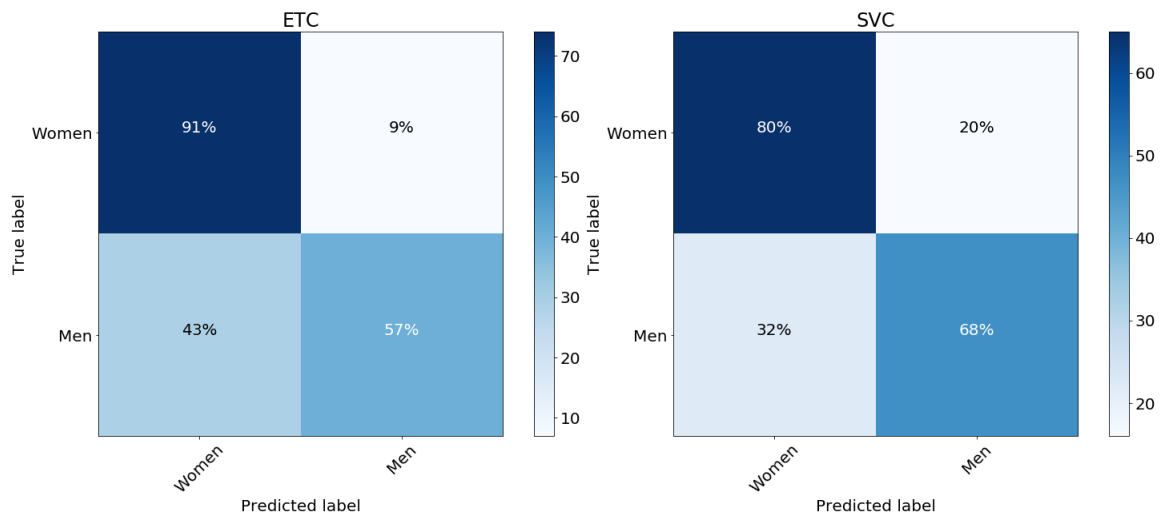


Figura 4.5: Matriz de confusión obtenida agregando el conjunto de datos reales (Navarro et al., 2018).

puede observarse en las matrices de confusión (Figura 4.4) que ETC reconoció correctamente al 68 % de las mujeres y al 96 % de los hombres.

Si bien para el algoritmo SVC, los valores porcentuales de éxito fueron mejores para el grupo de mujeres con un 78 % de aciertos, la tasa de éxito al clasificar a los hombres cayó notablemente al 69 % en comparación con el algoritmo ETC. Éste mismo *pipeline* volvió a ejecutarse pero esta vez, combinando los modelos artificiales con mallas de personas reales, conformando así un conjunto de datos de un total de 450 modelos. Para este nuevo conjunto de datos, ETC logró una precisión del 74 % y una exactitud del 86 %, mientras que SVC se desempeñó con una precisión del 75 % y una exactitud del 79 % (ver Figura 4.5).

4.5. Conclusiones

En este capítulo se introdujo el método propuesto para reconocimiento de sexo biológico basado en análisis de modelos 3D utilizando algoritmos de aprendizaje supervisado y para lo cual, se exploraron múltiples enfoques que utilizan diferentes metodologías de procesamiento. Se llegó a la conclusión de que el modelo con mejor rendimiento se basa en el uso directo de vértices de malla para la extracción de atributos que se toman como datos de entrada para dos algoritmos de clasificación diferentes, uno basado en SVM y el otro

en ETC. Se construyó un *pipeline* de procesamiento incluyendo técnicas de normalización y reducción de dimensionalidad de los datos que aminoran, significativamente el esfuerzo computacional requerido sin comprometer la calidad de los resultados. Dicho pipeline contiene un procedimiento de selección de datos de entrenamiento y de selección/optimización de hiper-parámetros para los cuales se mostraron los resultados obtenidos. Finalmente, las validaciones correspondientes se realizaron utilizando una parte del conjunto de datos dedicada exclusivamente a este fin, obteniendo los mejores resultados con el algoritmo SVC.

Capítulo 5

Extracción automática de características de la forma corporal para el estudio de obesidad

En este capítulo se presentará una metodología que permita obtener un descriptor para las estimaciones de sobrepeso y obesidad, utilizando evidencia de que preserva fielmente la naturaleza de la distribución de adiposidad abdominal como un aspecto de la forma. Este método está basado en imágenes 3D utilizando los modelos capturados en el proyecto *Patagonia 3D Lab* para lo cual, se detallará la metodología utilizando técnicas de morfometría geométrica y segmentación 3D.

5.1. Introducción

En la actualidad, el diagnóstico y tratamiento de obesidad están basados en indicadores clásicos como la relación entre peso y altura (Índice de Masa Corporal o IMC) ([Keys et al., 1972](#)), la circunferencia de cintura y cadera (Índice Cintura Cadera o ICC) o el Índice de Cintura Talla (ICT) ([Ashwell et al., 2012](#), [Larsson et al., 1984](#), [Lapidus et al., 1984](#)). En algunos casos, se utilizan evaluaciones antropométricas más complejas y costosas para obtener un marco general de proporciones corporales y distribución de grasa. Establecer un

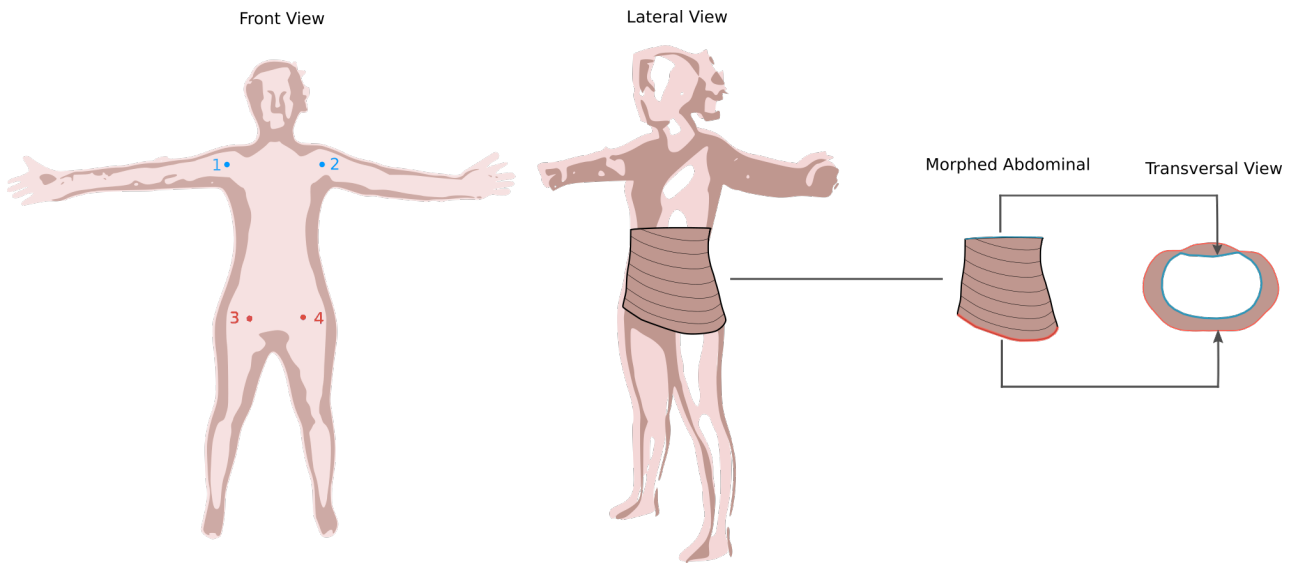


Figura 5.1: Ejemplo de imagen 3D del conjunto de datos, junto con la configuración de *landmarks* usada para la extracción de las características de la forma corporal. Los *landmarks* 1 – 2 corresponde a la articulación del hombro (glenohumeral) y los *landmarks* 3 – 4 a los punto donde articula la cabeza del fémur en la cadera (Navarro et al., 2019a)

umbral adecuado o puntos de corte en los índices para predecir estados nutricionales (como obesidad y sobrepeso), está directamente relacionado con las características de la población de referencia. Estudios anteriores indican las diferencias observadas entre las poblaciones de las islas de Asia, Europa o el Pacífico sobre los estándares de IMC para clasificar el sobrepeso y la obesidad en adultos (Carroll et al., 2008, Gu et al., 2018, Kagawa et al., 2007, Madan et al., 2012, Rønn et al., 2017). Estas diferencias también son reportadas para los índices ICC e ICT. Este problema sigue sin estar claro en poblaciones mixtas o mestizas, como en grandes partes de América Latina. Debido a su extensa mezcla entre las tres principales poblaciones continentales (africanos, europeos y nativos americanos), los latinoamericanos tienen un fondo biológico complejo (Koehl and Long, 2018, Norris et al., 2018, Ruiz-Linares et al., 2014).

En un estudio reciente se evaluó mediante regresión lineal múltiple, la relación entre obesidad definida como IMC, ICT y ICC, ascendencia genómica y estado socioeconómico en una muestra de latinoamericanos mezclados (6776 voluntarios adultos de Brasil, Chile, Colombia, México y Perú; Ruderman et al. (2019)). La proporción de obesidad depende en gran medida del índice y de la población considerada, y los índices no se superponen constantemente al clasificar la obesidad (Ruderman et al., 2019). Es importante establecer

un corte específico para el gran porcentaje de la población mestiza de América Latina y estudiar cuáles serían los parámetros para un diagnóstico correcto de sobrepeso y obesidad. Una mejor respuesta a este problema podría incluir el registro y el análisis de imágenes en 3D como se propone en este capítulo en el cual se combinan técnicas de computación aplicadas a la antropología.

5.2. Método

5.2.1. Descripción de la muestra

Para este procedimiento se utilizaron los datos colectados bajo el protocolo de *Proyecto 3D Lab* (Ver Capítulo 3), de los cuales únicamente se utilizaron los modelos de voluntarios reales debido a que fue necesario contar con información asociada para validar el método. Con el objetivo de no sesgar negativamente el experimento, se descartaron los individuos de la muestra que contenían deformaciones anómalas ubicadas en el torso de sus representaciones 3D. Estas anomalías se atribuyen a errores de escaneo, producidos por el dispositivo al capturar algún tipo superficie o simplemente por error del operador. El conjunto de individuos ($n = 112$) utilizado se describe en detalle en la Tabla 5.1, indicando los diferentes promedios y desvío estándar para las medidas tomadas junto con los diferentes estados nutricionales acorde a los puntos de corte correspondientes a los tres índices utilizados. Considerando la muestra, 15,2 % de la muestra fueron considerados obesos de acuerdo a su IMC, 31,2 % con sobrepeso, y 53,6 % saludables. El porcentaje de voluntarios con obesidad según el criterio de corte para ICT es de 81,2 %. La misma tendencia se muestra de acuerdo al ICC con 60 % de hombres y 69 % de mujeres obesas.

En las imágenes 3D fueron colocados cuatro puntos de anclaje anatómico también denominados *landmarks*. Estos *landmarks* fueron ubicados por un operario capacitado ya que se requiere tener experiencia y conocimientos de anatomía. La utilización de estos *landmarks* permite obtener puntos de correspondencia entre todos los modelos de la muestra. Es decir, cómo estos puntos están posicionados anatómicamente es posible ubicarlos en cualquier representación del cuerpo humano. En la figura 5.1 se puede observar la posición

		Hombre	Mujer
Medidas Lineales	Altura(cm)	173,6 ± 8,5	160,4 ± 5,8
	Peso(kg)	85,0 ± 17,11	64,6 ± 13,1
	Circunferencia de Cintura (cm)	96,6 ± 14,2	89,6 ± 11,0
	Circunferencia de Cadera (cm)	104,27 ± 9,38	101,16 ± 9,30
IMC	Saludable	6	54
	Sobrepeso	13	22
	Obeso	6	11
ICT	Saludable	6	15
	Obeso	19	72
ICC	Saludable	10	27
	Obeso	15	60

Tabla 5.1: Fila Superior: promedio y desviación estándar de las medidas lineales. Filas Inferiores: Tamaños de la muestra por sexo y estado nutricional. Índice de Masa Corporal (IMC) puntos de corte de acuerdo a la Organización Mundial de la Salud: *Obeso* ≥ 30 , *Sobrepeso* ≥ 25 *Saludable* $\geq 18,50 - 24,99$ WHO (2018). Índice de Cintura Talla (ICT) con punto de corte en 0.5 para ambos géneros. Browning et al. (2010) Índice Cintura Cadera (ICC), valores sobre 0,85 y 0,9 son considerados no saludables para mujeres y hombres, respectivamente. Navarro et al. (2019a).

de estos cuatro puntos de anclaje situados sobre un modelo ejemplo (ver Capítulo 3 para más detalle). Los *landmarks* 1 – 2 corresponde a la articulación del hombro (glenohumeral) y los *landmarks* 3 – 4 a los punto donde articula la cabeza del fémur en la cadera.

5.2.2. Extracción automática de características de la forma corporal

Como ya se mencionó, el escáner 3D es una herramienta potente a la hora de capturar muestras a gran escala y necesaria, para realizar estudios de población (Löfler-Wirth et al., 2016, Jaeschke et al., 2015). El problema con estas representaciones es como extraer características (*features*) discriminantes que permitan estudiar los objetos capturados de una manera eficiente. Para resolver este problema, el presente capítulo propone utilizar los *landmarks*, anteriormente nombrados, como referencia a la hora de extraer *features*. Es por esto que se calcularon dos puntos s_0 y s_1 entre los pares de *landmarks* 1,2 y 3,4. Como estos puntos fueron obtenidos a partir de puntos anatómicamente anclados

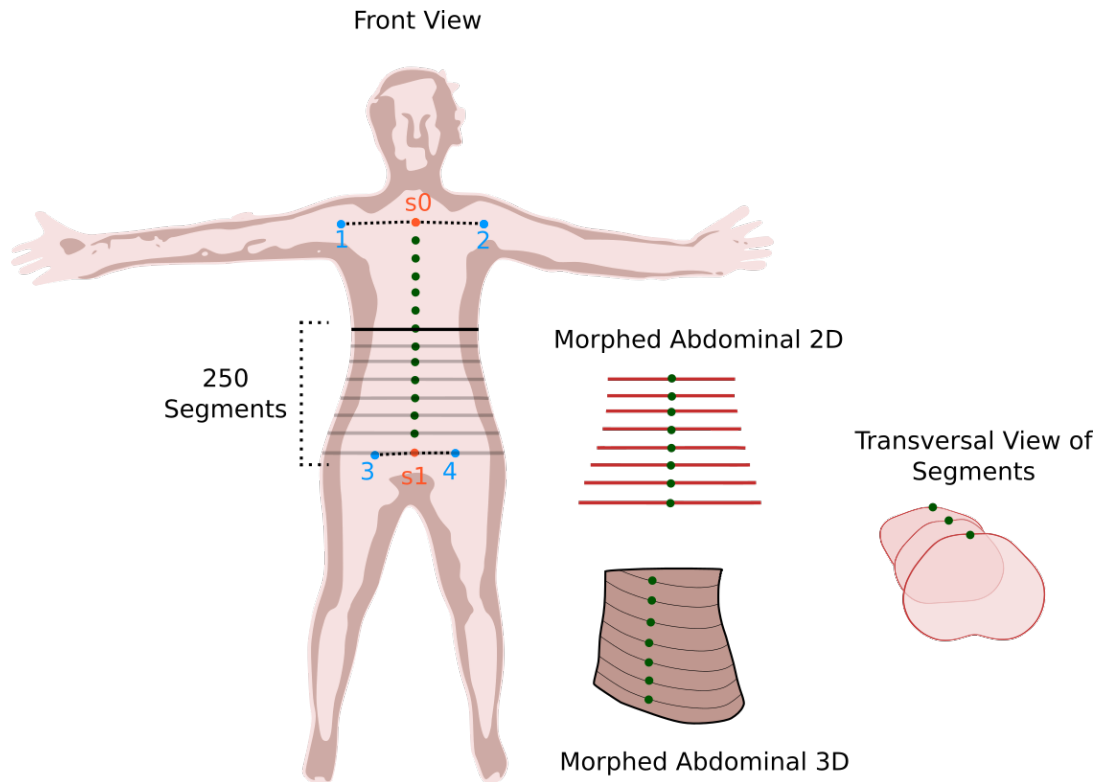


Figura 5.2: Localización de los *semi-landmarks* s_0 y s_1 en Naranja. Los 500 puntos ajustados por el criterio de mínimo *BendingEnergy* en verde. El área abdominal descrita en distintas vistas (2D, 3D y vista transversal.) generados con 250 segmentos. (Navarro et al., 2019a)

se los puede denominar *semi-landmarks* (Ver Figura 5.2). A partir de aquí, se calculó la distancia Euclidiana entre los *semi-landmarks*, con el fin de obtener información sobre la posición relativa al área abdominal. Utilizando la distancia que se calculó previamente, se posicionaron 498 puntos a lo largo del área correspondida entre s_0 y s_1 y se ajustaron con el criterio de menor *Bending energy* Gunz and Mitteroecker (2013). Finalizado este proceso se obtuvieron 500 *semi-landmarks* (incluyendo s_0 y s_1) posicionados sobre el modelo 3D (Ver Figura 5.2). La zona inferior del área abdominal constituye la región en donde se centra la mayor concentración de grasa corporal, por lo cual este es un área de interés para el estudio de obesidad y sobrepeso. Con esto en mente, se utilizaron los 250 *semi-landmarks* inferiores que, al estar fijados a *landmarks* anatómicos, constituye exactamente la mitad inferior.

A partir de aquí, y utilizando operaciones básicas en 3D, se generaron 250 láminas de corte, es decir 250 prismas rectangulares en tres dimensiones (con 0,01 espesor). Estos rectángulos están alineados con el modelo (uno con cada *semi-landmark*) y, mediante

operaciones de intersección, se obtuvieron 250 segmentos del área inferior abdominal. Estos segmentos son utilizados como dato principal para construir un espacio de *feature* capaz de detectar y analizar las diferencias de forma entre sexos y estado nutricional.

5.3. Resultados

Los 112 modelos 3D fueron procesados de manera automática como se detalló previamente. Para representar la variación de la forma abdominal se utilizaron las 250 circunferencias de cada voluntario, las cuales fueron proyectadas dentro de un espacio computado a través de PCA. El primer componente principal (PC1) y el segundo componente principal (PC2) explican 12,5% y 7,5% de la variación respectivamente. En Figura 5.3 se observa la distribución de los individuos dentro del espacio definido por estos dos primeros componentes. Además se visualiza la variación de la forma abdominal a lo largo de los ejes. Recorriendo el primer eje (PC1) desde los valores negativos hacia los positivos, los valores discriminan casi por completo los cambios anteriores-posteriores en los torso basado en reconstrucciones 3D correspondientes a los voluntarios con menor adiposidad central a aquellos con sobrepeso marcado. En el segundo eje, todos los valores negativos corresponden con torsos clasificados como saludables (de acuerdo a los tres índices). Estas variaciones de forma están afectadas por el efecto combinado del estado nutricional y el dimorfismo sexual de los individuos. El peso, la altura y las circunferencias de cintura y cadera de cada voluntario, medidas lineales que no se refieren completamente a la tridimensionalidad del cuerpo humano por sí mismas, se analizaron en relación con los parámetros de forma expresados por el espacio generado. Dado que los índices antropométricos están mutuamente y altamente correlacionados, se redujo todos estos valores en un nuevo componente a través de un PCA específico calculado solo en datos antropométricos. El primer eje representa el 73.08% de la varianza total y está dominado principalmente por el peso, seguido de la circunferencia de la cintura. Los valores residuales fueron introducidos como covariables dentro de un análisis de covarianza. Este estudio estadístico fue realizado con el objetivo de evaluar si las medias de los valores de PC1 y PC2 obtenidos de los cortes

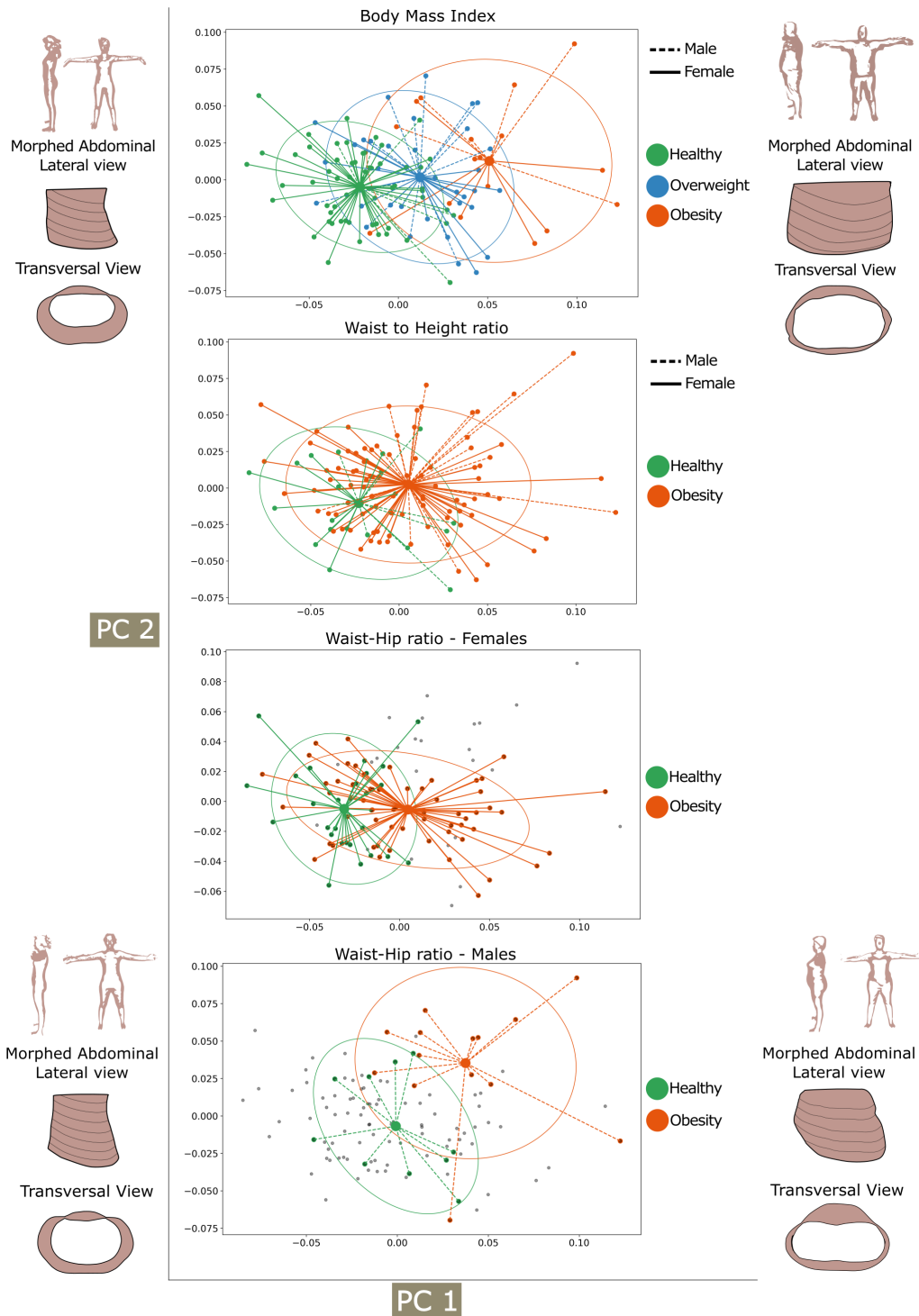


Figura 5.3: Análisis de componentes principales. Gráfico de dispersión y variación de forma y distribución de la grasa abdominal con el descriptor propuesto. Elipses coloreadas basadas en IMC, ICT, ICC estados nutricionales. Los cambios de forma representan el cambio de forma promedio observado en los extremos de los ejes. (Navarro et al., 2019a)

fueron diferentes para hombres y mujeres después de ajustar el efecto de las covariables (es decir, formas corporales relacionadas con diferentes estados de salud de los individuos).

La homogeneidad de la varianza entre los grupos se estableció con la prueba de Levene. El rasgo principal que explica la variación del PC fue la forma del cuerpo, seguida en menor medida ($P < ,05$), por dimorfismo sexual.

La variación en la forma que describe la obesidad abdominal se expuso a lo largo de PC1 y PC2 (Figura 5.3) donde el estado nutricional de los individuos están indicados con diferentes colores. Las elipses agrupan a los voluntarios de acuerdo con los valores de corte referidos para cada índice como “Saludable / Sobrepeso / Obesidad” para el IMC o “Saludable / Obesidad” para ICT e ICC. Las líneas coloreadas indican la distancia de cada voluntario al centroide de la distribución. Puede observarse que cada persona puede clasificarse en más de una categoría, de acuerdo con diferentes índices. Esta discrepancia es especialmente notable entre IMC y ICT. La clasificación es congruente solo para aquellos fenotipos considerados como obesos en relación con su peso y estatura, pero varía notablemente para el resto. En el espacio definido por los dos ejes las diferentes elipses se superponen, y muchos de estos voluntarios están más cerca del centroide de una categoría cercana que el que corresponde según su designación siguiendo cada uno de los tres índices.

La distribución del tejido adiposo abdominal debe considerarse como un aspecto de la forma y no como una relación lineal entre las medidas antropométricas. Se plantea la cuestión de si el análisis de varianza de un segmento de reconstrucción 3D puede verse como un enfoque más preciso para el diagnóstico de sobrepeso y obesidad. Para evaluar esta hipótesis, se realizaron análisis de regresión lineal entre el primer y el segundo eje (PCA calculado para hombres y mujeres por separado) y los tres valores de índice clásicos: IMC, ICT y ICC. Las regresiones fueron estadísticamente significativas para PC1 (Ver Tabla 5.2).

5.3.1. Medición de error y adquisición de características

Para medir qué tan robusto es el método aquí propuesto se evaluó cómo la ubicación de los puntos de referencia afectan el proceso y los resultados finales de la extracción automática de características para el área abdominal. Se realizó un análisis estadístico de error de la siguiente manera: para cada individuo en la muestra ($N = 112$) se replicaron

		PC - 1		PC - 2	
Genero	Index	r^2	P -value	r^2	P -value
Mujer	IMC	0,45	$1,25e^{-12}$	$8,3e^{-4}$	0,79
	ICT	0,62	$1,15e^{-19}$	0,002	0,67
	ICC	0,44	$1,62e^{-12}$	$1,41e^{-5}$	0,97
Hombre	IMC	0,40	$0,06e^{-2}$	0,07	0,18
	ICT	0,54	$2,3e^{-5}$	0,20	0,20
	ICC	0,55	$2,14e^{-5}$	0,14	0,14

Tabla 5.2: Regresión Lineal PC1 y PC2 contra índices clásicos (IMC, ICT, ICC) (Navarro et al., 2019a)

24 escenarios de error. Las coordenadas cartesianas x, y, z de los pares de puntos de referencia 1 a 2 y 3 a 4, fueron transformados en coordenadas polares, luego, se generaron 24 posiciones aleatorias alternativas en un radio de 1,5 cm para cada punto de referencia utilizando una distribución uniforme. Al final de este procedimiento, se obtuvo un total de 2688 individuos replicados, en los cuales se calcularon los *semi-landmarks* s_0, s_1 y los 250 segmentos (ver Figura 5.2). Los cortes obtenidos (2688 individuos emulados y 112 originales) fueron procesados a partir del método previamente detallado.

Para cada grupo de originales y réplicas ($N = 25$), se calculó el centroide de la varianza de esa distribución y la distancia entre cada uno de ellos al centroide para estimar el error intra-individual. Se realizó el mismo procedimiento para toda la muestra (error inter-individual). En comparación, la distancia inter-individual promedio fue de 0,05, mientras que la distancia intra-individual fue de 0,002. Por lo tanto, la varianza interindividual, incluso en las pruebas simuladas, es más del doble que la varianza intra-individual. Lo que concluye que en todos los escenarios simulados, la posible colocación incorrecta de puntos de referencia introduce un error en la segmentación 3D no significativo a los efectos de este análisis.

5.4. Conclusiones

El enfoque clásico utilizado en todo el mundo para definir el sobrepeso y la obesidad es el IMC, el cual es una herramienta económica y rápida de usar durante el diagnóstico

clínico. Sin embargo, como se informó anteriormente, el IMC y otros índices relacionados tienden a no describir la obesidad con precisión, debido a las variaciones dentro y entre poblaciones en la distribución de grasa abdominal ([Chandalia et al., 2007](#), [Gallagher et al., 2000](#), [Rush et al., 2007, 2009](#), [Wulan et al., 2010](#)).

El método presentado en este capítulo permite a los investigadores obtener de manera masiva información sobre la forma y variabilidad del cuerpo humano. Además este tipo de enfoque provee metodologías apropiadas para detectar la obesidad central ([Medina-Inojosa et al., 2016](#)) y determinar el volumen y la superficie del cuerpo junto con medidas antropométricas con gran precisión ([Kuehnafel et al., 2016, 2017](#), [Liu et al., 2017](#)).

Las imágenes 3D y los dispositivos para obtenerlas se han vuelto cada vez más accesibles y se deben desarrollar nuevos métodos de análisis que faciliten el procesamiento de esta información. En nuestra vida diaria, ya se utilizan tecnologías con gran capacidad de cómputo, como teléfonos inteligentes, tabletas y computadoras personales, con un potencial para la práctica profesional. Los avances en los métodos de análisis 3D permitirán a los médicos aprovechar las nuevas tecnologías (por ejemplo, aplicaciones para teléfonos inteligentes) para obtener definiciones más realistas de la forma del cuerpo en diferentes condiciones de salud y sus causas relacionadas. Después de la validación, estos enfoques futuros harán posibles mejores descripciones de la distribución de grasa corporal. Esto es necesario para capturar la variación “faltante” debido a factores intra e inter-poblacionales.

El método presentado en este capítulo va más allá de los índices tradicionales y problemáticos y alcanza un alto grado de precisión multivariante, el cual puede usarse para analizar grandes conjuntos de datos epidemiológicos. Al facilitar la captura de la forma del cuerpo, nuestro enfoque proporciona un conjunto de medidas, perímetros y proporciones que, colectivamente, representan un método rápido, no invasivo, económico y fácil de usar que puede incorporarse a la investigación y la práctica clínica centrada en obesidad y trastornos relacionados. Esta herramienta, enfocada en medir la adiposidad central, se puede adaptar a conjuntos de datos de imágenes en 3D de diferentes poblaciones, cada una de ellas con un objetivo biológico particular.

Capítulo 6

Descriptores Multi-vista basado en aprendizaje profundo

En este capítulo se presenta el desarrollo de un *framework*, denominado *Sketchzoom*, que permite calcular descriptores que soporten cambios significativos en la representación de forma. Si bien la presente tesis se encuentra dirigida a estudios relacionados al fenotipo humano, en este capítulo se utiliza un dataset de objetos inanimados no relacionados a esa área debido a su facilidad de obtención, mayor volumen y variedad, permitiendo además la comparación con otros métodos ya desarrollados para este fin. El funcionamiento de este *framework* se basa en el diseño de un espacio de características capaz de agrupar puntos de anclaje semánticamente similares, y disgregar a los diferentes.

6.1. Introducción

En el campo de la visión computacional, un problema recurrente es la necesidad de encontrar formas de representar características visuales presentes en imágenes, videos y/o modelos 3D, entre otros. Estas "formas", reciben el nombre de *descriptores visuales* o *descriptores*. El conjunto de datos utilizado consta de representaciones gráficas de objetos a través de contornos y con detalles no definidos e insinuados de forma esquemática (bocetos) los cuales darán al algoritmo la información necesaria para corresponder distintos tipos de

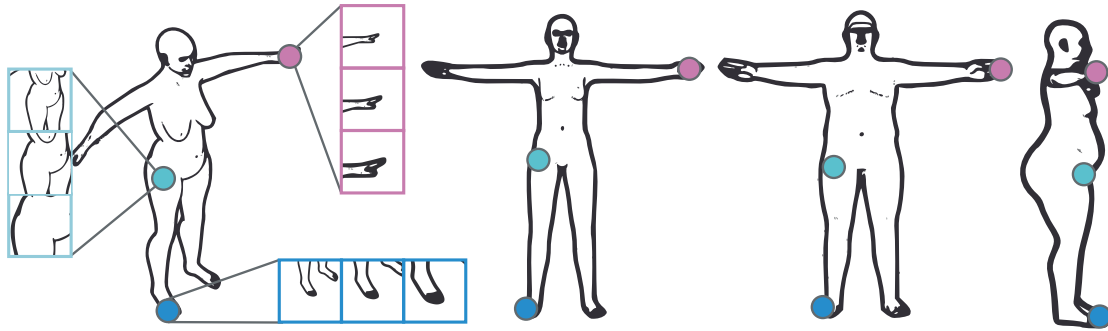


Figura 6.1: Dibujo de humanos con diferentes formas corporales.

representaciones. En comparación con fotografías, en estos dibujos las proyecciones son imprecisas y las formas a menudo están compuestas por varias líneas incompletas. (ver Figura 6.1) Las ventajas y posibilidades que brinda este *framework* son extrapolables a otros conjuntos de datos como los utilizados para el tema de esta tesis. Hacia el final del capítulo se describe el funcionamiento del mismo en el espacio de características o *features* obtenido a partir de modelos 3D del cuerpo humano.

Para aprender a extraer estos descriptores, se entrenó una red neuronal convolucional a partir de una sucesión de vistas ampliadas de un objeto. Dicha red consta de una adaptación de arquitecturas de redes neuronales de múltiples vistas (Redes Neuronales Convolucionales de Múltiples Vistas (MVCNN)) con un entrenamiento basado en *triplets* (Su et al., 2015, He et al., 2018) con el fin de obtener *features* de imágenes que transformen el contexto local y global de un dibujo a una espacio de múltiples vistas. El punto de consulta está representado por un conjunto de vistas 2D capturadas desde la vecindad inmediata del punto desde diferentes distancias (zooms). La red produce vectores de características para cada vista y aprende a fusionarlos mediante una capa de agrupación para construir un único vector descriptivo para el punto.

6.2. Materiales y Métodos

La Figura 6.2 proporciona una descripción general de los componentes principales del *pipeline*. Dada una colección de formas 3D, se renderizan automáticamente imágenes desde

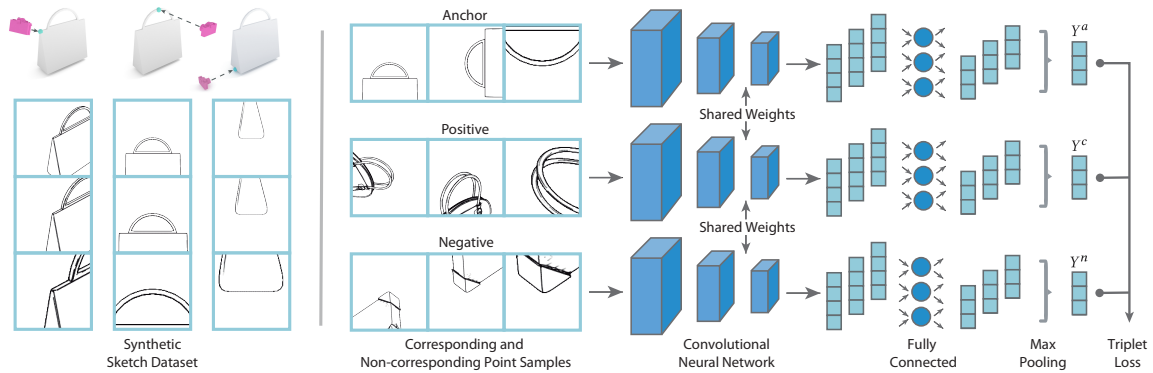


Figura 6.2: Vista general del *pipeline* propuesto. Izquierda: Dado un conjunto de datos precompilados de modelos 3D con datos de correspondencias, se generan automáticamente dibujos a diferentes escalas y posiciones utilizando un motor de renderizado no fotorrealista de última generación. Derecha: se utilizan estos dibujos como entradas para la red neuronal convolucional a fin de aprender los descriptores de múltiples vistas. Además se observa cómo los puntos coincidentes se asignan juntos independientemente del ángulo de proyección. Se utiliza una arquitectura de múltiples vistas conjuntamente con capas lineales para reducir el tamaño del descriptor y las capas de *MaxPooling* que agregan información importante a las vistas ampliadas. (Navarro et al., 2019b)

ángulos predefinidos. Los datos generados se combinan con la información de registro semántico disponible en el conjunto de datos para entrenar una red neuronal convolucional (representada a la derecha).

Conjunto de datos.

Para aprender características discriminantes y consistentes es necesario contar con un gran volumen de datos. En este caso, se tiene una vasta colección de más de 5 millones de bocetos, distribuidos en cuatro categorías semánticas: *Carteras*, *Sillas*, *Tazas* y *Auriculares*. Estos bocetos fueron obtenidos a partir de la renderización de 300 modelos 3D utilizando técnicas no fotorrealistas que imitan este tipo de diseño. Para construir este conjunto de datos, se utilizó el método de registro semántico y no rígido de Huang et al. (2017). Los datos fueron capturados a partir de tres vistas ortográficas principales comúnmente adoptadas por ingenieros y diseñadores: *de lado*, *frontal* y *vista 3/4* (isométrico) centradas en un punto p (ver figura 6.1).

Para evaluar el funcionamiento de la red se utilizaron 40 bocetos adicionales con dife-

rentes estilos y vistas. Se registraron manualmente las correspondencias de puntos entre cada uno de estos bocetos con el fin de obtener métricas cuantitativas junto con resultados cualitativos.

Red convolucional de múltiples vistas

Como fue mencionado previamente, el modelo MVCNN utilizado en esta tesis, toma un conjunto de imágenes como entrada (Figura 6.2). Esta secuencia consiste en una serie de vistas ampliadas centradas en un punto particular desde una ventana gráfica fija. Los distintos enfoques (*zooms*) sobre las imágenes de entrada, proporcionan a la red información de contexto local y global sobre el punto objetivo. Esta representación basada en vistas es cercana a la forma en que los humanos perciben los objetos en nuestro mundo 3D, y se ha demostrado que es efectiva en tareas de reconstrucción que involucran formas 3D y datos de dibujo (Lun et al., 2017).

El objetivo de nuestra configuración de entrenamiento basado en *triplets* es asegurar que: (i) los puntos coincidentes tengan sus vectores de característica próximos en el espacio de características del descriptor, y (ii) los puntos no correspondientes tengan sus vectores alejados entre sí. La arquitectura de múltiples vistas aprende a combinar datos importantes de cada una a través de la capa de *Max Pooling*, ignorando la información superflua. Dado que este tipo de entrenamiento a menudo hace que la inserción colapse en grupos excesivamente pequeños, se refinó el procedimiento mediante una estrategia de selección personalizada que evita este problema.

6.2.1. Entrenamiento de la red

Datos de entrada: data augmentation

Como se mencionó anteriormente, la red calcula descriptores locales a partir de un conjunto de imágenes que representan vistas del mismo objeto con 3 tipos de enfoques o *zooms*. Antes de alimentar la red, las imágenes se normalizan y se escalan a una resolución de 224x224 píxeles. Dado que en el mundo real es común toparse con imágenes que

no están perfectamente orientadas con respecto a la posición de la cámara, el proceso de entrenamiento de la red permite producir descriptores invariantes a este problema mediante rotaciones aleatorias de las imágenes de entrada. Cada una de las tres vistas tiene la misma probabilidad de permanecer sin transformarse o de rotarse individualmente 90, 180 o 270 grados.

Para generar un descriptor robusto a diferentes resoluciones, se redujo el tamaño de la imagen de entrada entre un 30 % y 60 % con una probabilidad de 0,2, respectivamente. Luego se procedió a restaurar las entradas al tamaño original usando interpolación lineal para simular diferentes cantidades de degradación de la calidad. En experimentos preliminares se notó que las características de SketchZooms eran altamente sensibles a la distancia del punto objetivo de la cámara. Por lo tanto, se agregó ruido al parámetro que computa el *zoom* al muestrear los desplazamientos de la cámara desde una distribución normal (con $\mu = 0$ y $\sigma^2 = 0,3$, donde 0,3 significa un incremento del tamaño del 30 % con el tamaño de la imagen original) Este ruido del *zoom* se aplica directamente sobre las imágenes renderizadas como un proceso posterior durante el entrenamiento. Los procesos de aumento de datos fueron iterativos y guiados empíricamente por los resultados obtenidos durante la etapa de experimentación utilizando un conjunto de validación.

Arquitectura de red

La red diseñada se basa parcialmente en la arquitectura *AlexNet* (Krizhevsky et al., 2012). Está compuesto por cinco capas convolucionales, seguidas por capas *ReLU* y *Max-Pooling*. De la arquitectura original, se procedió a excluir las dos últimas capas relacionadas a tareas de clasificación en *ImageNet* (Krizhevsky et al., 2012). En su lugar, se utilizó una capa de agrupación de vistas que fusiona los descriptores generados para cada una de las tres vistas de entrada formando así un único vector descriptor. La fusión se realiza mediante el máximo de elementos en las vistas de entrada.

Para acelerar el cálculo de los descriptores y hacer que las consultas sean más eficientes, se antepone una capa lineal que aprende a mantener información relevante y transforma la salida de dimensión 4096 (fc6) a un vector de tamaño 128. Esta reducción de dimensión

se realiza antes de la capa de agrupación (*MaxPooling*) la cual beneficia enormemente el cómputo máximo de agrupación sin notar ninguna degradación en la calidad del descriptor (Ver figura 6.2).

Función de error basada en Triplets

Mientras que en trabajos recientes como Wang et al. (2015) se utiliza una arquitectura siamesa y la función de pérdida por contraste (*Contrast Loss*), en el enfoque aquí propuesto se adopta una función de pérdida basada en Triplets o (*Triplet Loss*), introducido por primera vez en *FaceNet* (Schroff et al., 2015). El motivo principal para esta decisión es que las distancias entre features obtienen mayor información semántica cuando se ponen en contexto y además, el punto de anclaje introducido por la *Triplet Loss* aporta información de forma al espacio de *features* (Wu et al., 2017). Idealmente, dos puntos que se corresponden deberían encontrarse cerca (corta distancia) en el espacio de *features* del descriptor mientras que dos muestras que no se corresponden se encontrarán alejadas en tal espacio. *Triplet Loss* tiene en cuenta ambos objetivos al minimizar la distancia entre un descriptor (también llamado positivo) Y^c y "ancla" Y^a correspondiente con tal descriptor (ver Figura 6.2). Simultáneamente, maximiza la distancia entre el ancla y un descriptor de punto que no corresponde (negativo) Y^n .

Matemáticamente, esta función de pérdida se representa de la siguiente forma:

$$D^2(Y^a, Y^c) + \alpha < D^2(Y^a, Y^n), \quad (6.1)$$

donde D representa la distancia euclidiana entre descriptores, y α es un margen impuesto entre pares positivos y negativos ($\alpha = 1$ en la implementación utilizada). Si se formula la ecuación 6.1 como un problema de optimización sobre los parámetros de la red (\mathbf{w}), se obtiene:

$$\mathcal{L}(\mathbf{w}) = \sum_i^N \max(D^2(Y_i^a, Y_i^c) - D^2(Y_i^a, Y_i^n) + \alpha, 0), \quad (6.2)$$

donde N es la cardinalidad del conjunto de entrenamiento. Si se intenta minimizar la ecuación 6.2, las distancias entre pares positivos $D^2(Y_i^a, Y_i^c)$ se acercan a cero y la distancias de pares negativos se hacen mayores a $D^2(Y_i^a, Y_i^c)$ más el margen α . Entrenar con todos

los tripletes posibles en un dataset de 5 millones de imágenes es extremadamente costoso. Además, el uso de todos los triplets es altamente ineficiente ya que cuanto más progresa el entrenamiento, más triplets satisfarán la ecuación 6.1, haciendo que el entrenamiento sea más lento a medida que avanza el tiempo (Schroff et al., 2015).

Para solucionar este problema se seleccionan ternas de forma adaptativa por lo que en cada paso de entrenamiento se utilizan sólo aquellos casos que satisfagan:

$$\begin{cases} D^2(Y^a, Y^c) < D^2(Y^a, Y^n), \\ D^2(Y^a, Y^n) < D^2(Y^a, Y^c) + \alpha, \end{cases} \quad (6.3)$$

donde $\{Y^a, Y^c, Y^n\}$ se refiere a la terna formada por el ancla, el positivo y el negativo respectivamente. En esta ecuación se busca obtener muestras de entrenamiento que se encuentren dentro del área del margen delimitado por α .

En la práctica, se computa $\{Y^a, Y^c, Y^n\}$ a partir de imágenes de entrada $\{X_v^a, X_v^c, X_v^n\}$ utilizando la red en el último estado. Las ternas son construidas al mismo tiempo que se produce el entrenamiento, poniendo a prueba a los descriptores para determinar si infringen o no la ecuación 6.3. En la configuración utilizada, se agrupan muestras individuales en grupos G para usar de forma secuencial durante cada iteración de entrenamiento. Para construir un bloque de entrenamiento, se toma una muestra al azar de pares positivos de G de la forma $[Y_i^a, Y_j^c]$, $i, j \in G$. Luego se comprueba la ecuación 6.3 sobre un número aleatorio s de muestras negativas $[Y_i^a, Y_k^n]$, $i, k \in G$. Se experimentó con distintos valores para s y se encontró que un $s = 5$, minimiza el tiempo dedicado a la búsqueda aleatoria y al mismo tiempo proporcionar ternas adecuadas para el entrenamiento.

Herramientas utilizadas

La arquitectura de la red fue implementada en el lenguaje *Python*¹ haciendo uso de la librería *PyTorch 1.0*² y fue entrenada utilizando una placa gráfica *NVIDIA Titan Xp*. El entrenamiento de la red demoró aproximadamente 3 semanas por categoría. Además

¹<https://python.org/>

²<https://pytorch.org/>

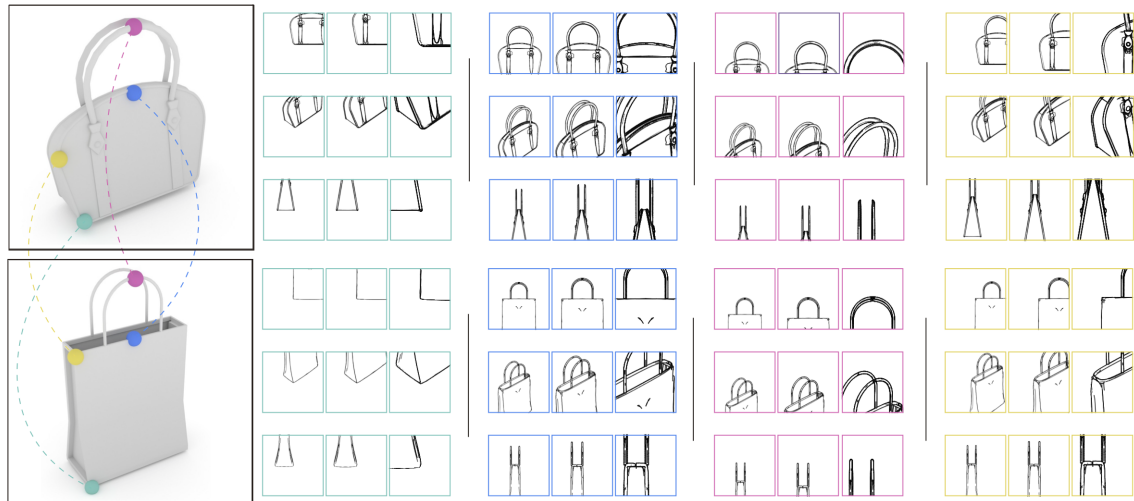


Figura 6.3: Visualización de imágenes del conjunto de datos. Dado un par de modelos 3D densamente registrados (izquierda), se realizaron renderizaciones desde tres ángulos (columnas) predefinidos y tres *zooms* (filas) diferentes. Cada imagen se centra en el punto de interés correspondiente al punto coincidente (indicado con un contorno de color).

se diseñó una configuración utilizando cuatro categorías en simultáneo, en este caso el entrenamiento demoró 5 semanas. Se inicializan las capas de la red utilizando los pesos de *AlexNet* entrenada con el conjunto de datos de *ImageNet* provisto por *PyTorch*. La tasa de aprendizaje (o *learning rate*) se estableció en $l = 1 \times 10^{-5}$ y ambas redes se entrenaron durante 30 épocas. Se utilizó la función de pérdida 6.2 junto con el algoritmo de optimización *Adam* (Kingma and Ba, 2014) ($\beta_1 = 0,9, \beta_2 = 0,999$) y un tamaño de *batch* de 128 ternas. No se utilizaron capas de normalización o capas *Dropout* a parte de las que ya se encuentran en *AlexNet*.

6.2.2. Construcción del conjunto de datos

En esta sección se describe en mayor detalle el conjunto de datos 3D y la técnica de renderización empleada para crear las imágenes sintéticas utilizadas en la etapa de entrenamiento y evaluación del método propuesto.

Modelos 3D

En general, existe una gran cantidad de *datasets* como [Eitz et al. \(2012\)](#) o [Ha and Eck \(2017\)](#), en los cuales se compila una gran cantidad de dibujos. Las imágenes de estos conjuntos de datos a menudo provienen de autores que no necesariamente tienen habilidades de dibujo o conocimiento específico sobre diseño además de no proporcionar correspondencias anotadas entre bocetos. Por lo tanto y de manera similar a lo propuesto por trabajos recientes dirigidos a bocetos y aprendizaje automático ([Delanoy et al., 2017](#), [Huang et al., 2016](#), [Su et al., 2018](#)), se procedió a generar dibujos sintéticos directamente capturados de las formas 3D, las cuales si cuentan con una correspondencia semántica ([Huang et al., 2017](#)). De las categorías disponibles en ese conjunto de datos, se seleccionaron aquellas que ofrecen la mayor cantidad de modelos para garantizar diversidad en la muestra. Es entonces que de los 4932 modelos originales distribuidos en 24 categorías, se eligieron 79 de la categoría cartera, 62 de sillas, 73 de tazas y 68 de auriculares. Los modelos de la muestra están centrados, escalados y orientados de la misma manera, lo que facilita el posicionamiento de la cámara durante el renderizado. Además, estos modelos 3D cuentan con archivos de correspondencias que proporcionan una lista de puntos que coinciden en cada posible par de formas dentro de cada categoría. Por cada forma 3D se obtiene una muestra al azar con alrededor de 10,000 puntos sobre la superficie del modelo. Las correspondencias se calcularon con un algoritmo de registro basado en segmentación que realiza una alineación no rígida de todos los pares con la misma etiqueta sobre dos formas de destino.

Dibujos sintéticos. Los métodos utilizados para representar bocetos emplean técnicas de procesamiento de imágenes como *Canny Edge*[1987] o contornos como [Saito and Takahashi \(1990\)](#). Para este algoritmo se adoptó *Aparent Ridges* de [Judd et al. \(2007\)](#), la cual es una buena aproximación a las líneas de los artistas como se muestra en [Cole et al. \(2012\)](#). Además del estilo, la selección de las vista es crucial para transmitir información de forma. La literatura sobre diseño recomienda adoptar vistas específicas para reducir la ambigüedad del boceto y mostrar simultáneamente la mayor parte de la forma del objetivo ([Eissen and Steur, 2011](#)). La mayoría de los algoritmos de reconstrucción 3D a partir de bocetos se basan en suposiciones como paralelismo, ortogonalidad y simetría ([Cordier et al., 2016](#)).

Con base en los lineamientos de diseño, se seleccionó un conjunto de vistas ortográficas para representar desde los modelos 3D. Se utilizó un total de tres vistas: frontal, lado derecho y un ángulo isométrico único, también llamado vista informativa: frontal-derecha o 3/4 (ver Figura 6.3). Para capturar las imágenes, se centró la cámara en cada punto de interés de la muestra y se capturó desde tres distancias diferentes y tres vistas distintas mientras que los puntos ocluidos fueron descartados. En total, nuestro conjunto de datos consta de 5,053,038 imágenes en una resolución de 512x512 píxeles.

Parámetros para la renderización

Se utilizó el algoritmo *Apparent Ridges* incorporado en el software RTSC de [DeCarlo et al. \(2003\)](#). Además, se utilizó información del contorno para el renderizado. Se adoptó un modelo de cámara ortográfico, capturando las imágenes de las vistas laterales y frontales a distintas distancias del objetivo. Se capturaron los *zooms* en 1,0x, 1,5x y 2x. Para suavizar los modelos, se ejecutaron dos iteraciones de la subdivisión de malla y los algoritmos de suavizado de curvatura implementados en la biblioteca de manipulación 3D *Trimesh*³.

Colocación manual de puntos de referencia

La transferencia de conocimiento en redes aprendidas con datos sintéticos a entradas reales ha demostrado ser efectiva, antes de lidiar con las dificultades de compilar y procesar masivamente datos reales ([Wang et al., 2015](#), [Delanoy et al., 2017](#)). Para validar el rendimiento al trabajar con imágenes reales, se recopiló un conjunto de 40 bocetos dibujados a mano, disponibles gratuitamente en línea o en repositorios como *Shutterstock*⁴. Se seleccionaron 10 bocetos por categoría ya que es una cantidad de imágenes razonable para que usuarios puedan hacer coincidir manualmente. El conjunto compilado incluye dibujos lineales con estilos, sombras, líneas de construcción, decoraciones y ruido de proyección/cámara significativamente diferentes. Posteriormente, se procesaron para estandarizar la escala y resolución de las imágenes. Para todas las imágenes recopiladas, se seleccionaron cuatro puntos aleatorios y se solicitó a los usuarios mediante una aplicación experimental que

³<https://trimsh.org/>

⁴<https://www.shutterstock.com>

hicieran coincidir manualmente cada par de bocetos dentro de la misma categoría. Para recopilar correspondencias se consideraron todas las combinaciones posibles de pares de imágenes de origen-destino (combinatorias de 2 elementos de 10 = 90) para cada clase de objeto. Se dividió la tarea completa en cuatro usuarios sin superposición y se obtuvo un total de 1440 correspondencias.

6.3. Diseño experimental

Se realizaron experimentos teniendo en cuenta múltiples aspectos del enfoque aquí presentado. (i) Se probó el funcionamiento en una serie de ejemplos dibujados a mano para evaluar la capacidad de generalización de la red (ver Figura 6.4). (ii) Se examinó la capacidad del espacio aprendido para distribuir adecuadamente los descriptores. (iii) Se calcularon métricas de precisión en las correspondencias realizadas por la red para evaluar el rendimiento de coincidencia sobre el espacio de la imagen. (iv) Se realizó un estudio perceptual para evaluar los aspectos semánticos de los *features* generados. Dado que el *pipeline* es lo suficientemente general como para permitir el aprendizaje en un entorno de múltiples categorías, se evaluó adicionalmente el modelo entrenado en las cuatro categorías de objetos simultáneamente. Además para comprobar el funcionamiento del descriptor utilizando figuras humanas se aplicó la misma metodología detallada anteriormente con dibujos sintéticos directamente capturados a partir los modelos artificiales del *Proyecto 3DLab* (Ver Figura 6.1), las cuales cuentan con una correspondencia semántica a partir de la utilización de los *landmarks* como puntos de coincidencia entre modelos (Ver Capítulo 3).

6.3.1. Configuración de la red de categoría específica

Calidad del descriptor

Se evaluó la expresividad del descriptor utilizando *Cumulative Match Characteristic (CMC)*, una métrica de calidad utilizada para la correspondencia de imágenes ([Karanam](#)

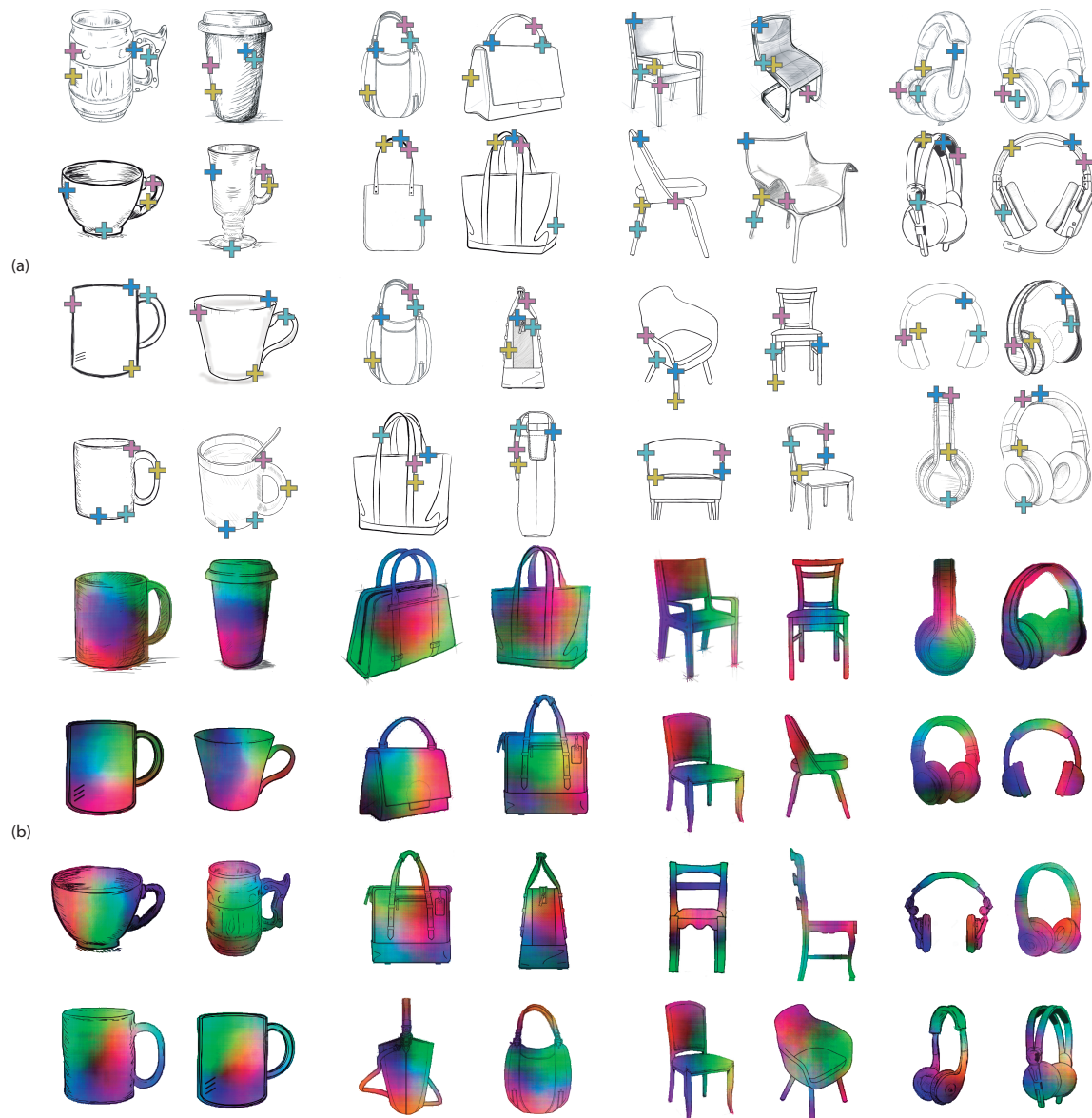


Figura 6.4: (a) Correspondencias por pares en cuatro categorías de objetos. Los puntos que se corresponden tienen el mismo color. (b) Visualización de correspondencias densas utilizando el descriptor aquí presentado en diferentes categorías. Los píxeles correspondientes se indican en colores similares. Se utilizó una máscara binaria para excluir píxeles fuera de la región del boceto. A pesar de las diferencias extremas en la geometría y las posiciones de la cámara, el descriptor logra registrar correctamente los bocetos. (Navarro et al., 2019b)

et al., 2015, Wagg and Nixon, 2004). Esta métrica captura la proximidad entre dos puntos dentro del espacio del descriptor, calculando distancias sobre pares de descriptores en dos bocetos destino. Dado un punto en una de las imágenes, se recupera una lista de correspondencias candidatos en la otra imagen. Luego, los candidatos se clasifican utilizando una

medida de proximidad como por ejemplo, la distancia euclidiana en el espacio descriptor. Al repetir este proceso y acumular la lista de candidatos en todos los pares de imágenes de prueba, se creó un gráfico en el que el eje Y representa el número de puntos coincidentes (*ground-truth matching*) cuya lista esté por debajo del número representado en el eje X. Se realizó la comparación del método con los descriptores comúnmente utilizados para las tareas de correspondencia en bocetos, incluidos los histogramas radiales de *Shape Context* (SC) (Belongie et al., 2002) y el descriptor *GALIF*, basado en filtros *Gabor* de Eitz et al. (2012). Además, se consideró un descriptor artesanal que consiste en un análisis de componentes principales (PCA) sobre una pequeña vecindad de píxeles que rodean el punto objetivo. Esto proporciona una línea base de comparación simple para todos los métodos probados.

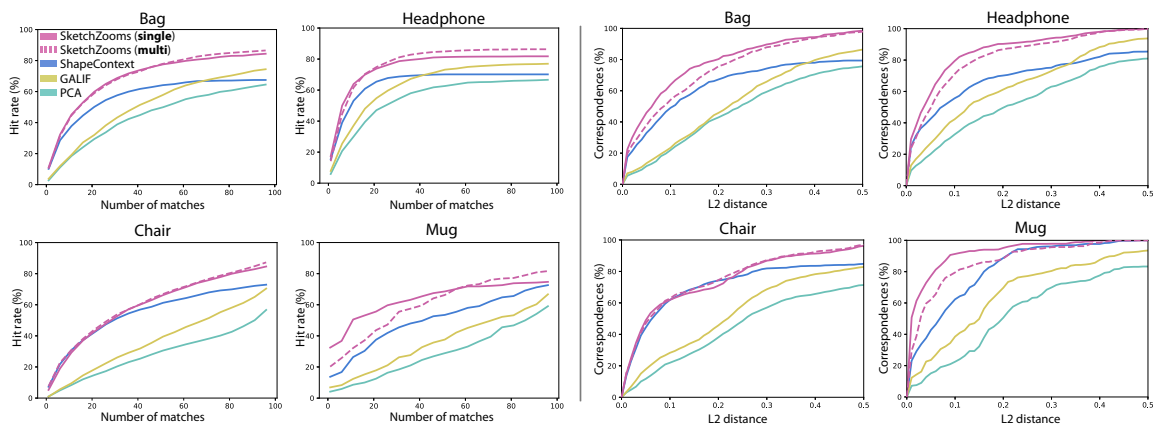


Figura 6.5: Izquierda: Gráficos de CMC para los descriptores evaluados en los conjuntos de datos de prueba para cada categoría. El eje Y representa el porcentaje de coincidencias recuperadas debajo de la posición indicada en X. Derecha: curvas de precisión de correspondencia donde el eje X muestra un error de distancia euclidiana normalizado. El eje Y representa el porcentaje de coincidencias recuperadas por debajo del margen de error indicado en X. (Navarro et al., 2019b)

Se calculó la métrica CMC utilizando el conjunto de datos de prueba que consistió en bocetos sintéticos de múltiples vistas tomados de modelos no utilizados durante la fase de entrenamiento. Dadas dos formas, se consideraron todas las combinaciones entre el conjunto de vistas (frente-frente, frente-lado, frente-3/4, lado-frente, lado-lado, lado-3/4, 3/4-frente, 3/4-lado, 3/4-3/4). Dado que los métodos como *ShapeContext* no están bien definidos sobre regiones vacías del boceto (es decir sobre áreas en blanco), solo se consideraron los puntos que se encuentran en píxeles negros para garantizar una comparación

equitativa entre todos los métodos. Sin embargo, cabe aclarar que el método aquí presente puede ser calculado incluso en áreas alejadas de píxeles negros, como se muestra en la Figura 6.4. En total, las muestras de prueba consisten en 2,000 puntos correspondientes para cada categoría seleccionada. La figura 6.5 muestra el rendimiento de los descriptores evaluados sobre el conjunto de pruebas sintéticas. Se reportan los valores de la métrica en la Tabla 6.1. En todos los casos, el descriptor aprendido supera a los otros métodos capturando mejor la semántica de los puntos objetivo a través de los cambios de vista.

6.3.2. Configuración de la red multicategoría

En un intento de construir un descriptor más genérico, se realizó un experimento adicional para estudiar la viabilidad de aprendizaje de la red neuronal propuesta en un entorno de múltiples categorías. Con este fin, se entrenó el método utilizando un conjunto de datos unificado que comprende cuatro categorías de objetos en total. Estos modelos presentan una variabilidad significativa en cuanto a características de forma, con más de 350 objetos diferentes considerados para esta configuración. La figura 6.5 incluye el CMC y los gráficos de precisión de correspondencia para 8,000 muestras de prueba, en comparación con otros métodos de referencia. Ambas configuraciones de entrenamiento mostraron un comportamiento similar, superando el rendimiento de los descriptores informados. Más importante aún, el entrenamiento en varias categorías superó al entrenamiento individual al calcular las métricas de calidad del espacio del descriptor (ver Tabla 6.1). A pesar de la mayor complejidad de este escenario, la red multicategoría puede producir descriptores de forma generales que funcionan favorablemente en comparación con otras alternativas. Sin embargo, se observa que los modelos entrenados con categorías específicas aseguran emparejamientos más confiables entre componentes de objetos con la misma funcionalidad, incluso si su forma o apariencia difiere significativamente.

Modelos	SC	GALIF	PCA	Red (Por categoría)	Red (Multicategoría)
Carteras	67.47 %	74.40 %	64.51 %	84.39 %	86.51 %
Sillas	72.91 %	70.4 %	56.67 %	84.60 %	87.25 %
Tazas	72.61 %	66.71 %	59.07 %	74.71 %	81.65 %
Auriculares	70.11 %	76.96 %	66.61 %	81.75 %	86.30 %

Tabla 6.1: Resultados de la métrica *Cumulative Match Characteristic (CMC)* medidos en 4 métodos. CMC es reportado sobre 100 correspondencias. Aquí se presentan los resultados para la red entrenada con cada categoría y para la red multicategoría. (Navarro et al., 2019b)

6.4. Resultados

6.4.1. Imágenes dibujadas a mano

La figura 6.4 ilustra los resultados del descriptor en un conjunto de datos dibujados a mano en dos escenarios de correspondencia, dispersa y densa. Observamos que el método es capaz de utilizar las características aprendidas a partir de un conjunto de entrenamiento con datos sintéticos cuando se trabaja con imágenes dibujadas a mano. En particular, el rendimiento se mantiene estable para estas imágenes, incluso si la posición de la cámara no está perfectamente alineada con los ejes de los objetos. Más importante aún, las decoraciones y las líneas de sombreado no introducen discontinuidades significativas sobre los mapeos densos. Las correspondencias predichas son suaves y visualmente plausibles. Además, se puede observar que la mayoría de los pares corresponden a dibujos lineales representados desde diferentes posiciones de cámara. Los descriptores pueden capturar aproximadamente la semántica y la información 3D proporcionada por los bocetos. Para acelerar el cálculo de mapas densos, se utilizaron máscaras binarias que eliminan los píxeles fuera de la región de interés. Se visualizaron los mapas de colores densos proyectando descriptores en R^3 (Espacio de colores RGB).

Métricas

Se proporcionaron métricas cuantitativas calculadas sobre el conjunto de imágenes dibujadas a mano. Dado que el conjunto de datos contiene correspondencias escasas, se

consideró apropiado informar el Error cuadrático medio (RMSE) normalizado. Se calcularon las correspondencias utilizando los *features* de las redes de entrenamiento de una sola categoría y se obtuvieron las distancias L2. Para esta operación, se utilizó un conjunto de datos de 1440 correspondencias no densas. El descriptor aquí presentado obtuvo un mejor rendimiento comparado con los métodos considerados: *SketchZooms*: 0.069, *Shape Context*: 0.074, *GALIF*: 0.195 y *PCA*: 0.206. Sin embargo, es importante resaltar que métodos como *Shape Context* son independientes del modelo y en consecuencia, son más generales. El objetivo no es reemplazar los descriptores tradicionales sino introducir una nueva solución a los problemas relacionados con la vista 3D y la información de las piezas directamente en 2D.

6.4.2. Estudio de usuario

Los humanos poseen la extraordinaria capacidad de resolver correspondencias semánticas en escenarios de múltiples vistas gracias a su conocimiento previamente adquirido sobre el mundo real en 3D. Para comprender mejor el potencial del descriptor aquí desarrollado, se realizó un estudio de usuario. Para lo cual se convocó a 10 voluntarios a utilizar una aplicación experimental desarrollada para este fin. A cada uno de los 10 voluntarios del estudio se les presentó un par de imágenes con $m = 4$ puntos en la primera y se le solicitó que encontrará los m puntos correspondientes en la segunda imagen. Se utilizó un total de 40 pares de imágenes distribuidas en cuatro categorías de evaluación: cartera, silla, taza y auriculares. Los puntos se seleccionaron previamente de manera aleatoria de una lista *features* calculados sobre todas las imágenes de estudio utilizando el detector de esquina de [Shi and Tomasi \(1993\)](#). Similar al estudio presentado en *BestBuddies* ([Aberman et al., 2018](#)), se ajustó a una distribución Gaussiana 2D sobre las coordenadas de puntos anotadas por los usuarios. Se definió una medida de similitud evaluando la función de densidad de probabilidad ajustada en los puntos de consulta preseleccionados. Luego se asignaron puntuaciones de similitud más altas a las regiones donde el consenso entre los usuarios fue más fuerte y viceversa. Se promediaron todos los puntos basados en la puntuación obtenida. Los resultados se resumen en la Tabla 6.2. En todos los métodos que fueron probados, se filtraron los valores atípicos fuera del umbral de error $mean \pm std$ medido en píxeles.

	SC	GALIF	PCA	Red (Por categoría)	Red (Multicategoría)
Score	0.043	0.01	0.009	0.047	0.048
NRMSE	0.068	0.164	0.17	0.062	0.067

Tabla 6.2: Estudio de métricas cuantitativas. Los valores en la primera fila representan la probabilidad de que cada descriptor probado forme parte de las estadísticas del usuario, modeladas con una distribución gaussiana en 2D sobre los datos del estudio. En la segunda fila se detalla los valores de Error cuadrático medio normalizados (RMSE) resultado del rendimiento de los descriptores con respecto a los datos del usuario.

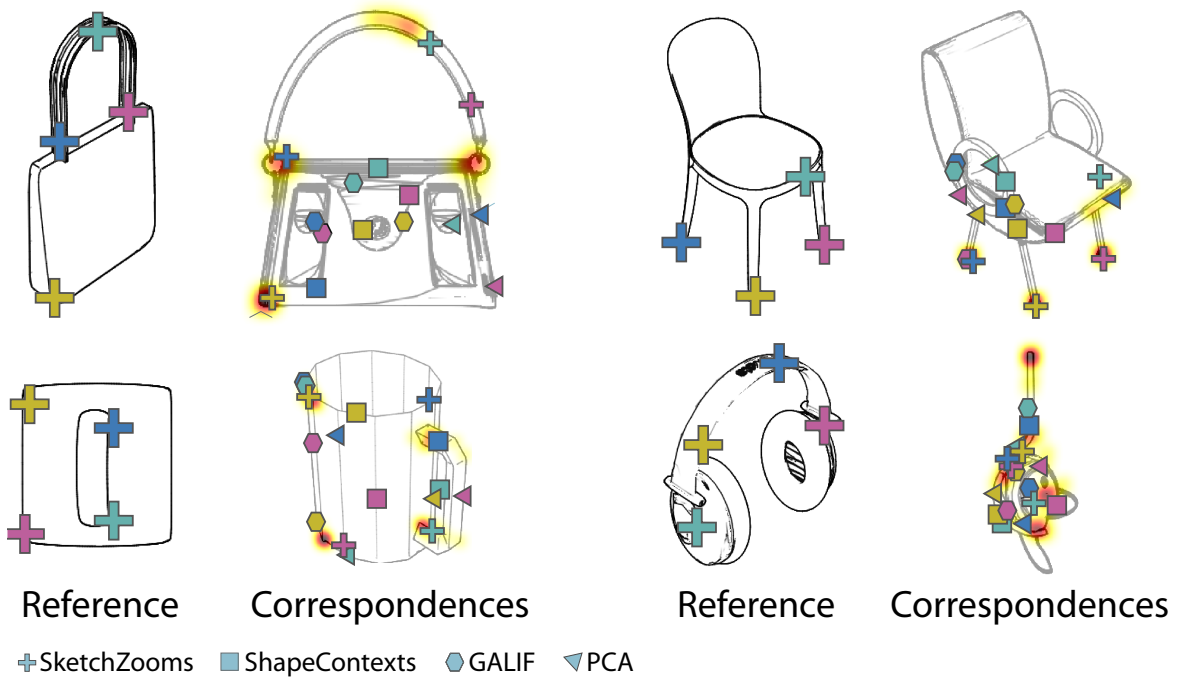


Figura 6.6: Mapa de calor con las correspondencias anotadas por los usuarios. En cada par, la imagen de la izquierda muestra los puntos de referencia que debían ser correspondidos por el usuario, la imagen de la derecha representa un mapa de calor con las coincidencias de los usuarios más fuertes. Los símbolos indican las correspondencias para cada método evaluado.

Los mapas de calor en la Figura 6.6 representan áreas donde el consenso de coincidencia de los usuarios fue más fuerte. Las correspondencias obtenidas con el descriptor están más cerca de las áreas “calidad” que las producidas con otros métodos. Se utilizaron redes entrenadas por categorías para calcular las correspondencias.

6.4.3. Pruebas con figuras humanas

Con el fin de observar el funcionamiento del descriptor sobre figuras humanas se realizó una evaluación cualitativa. Se entrenó la red con imágenes de 450 modelos artificiales, de los cuales se le registraron 15 puntos por individuo. Si bien los modelos cuentan con pocos puntos de correspondencia, la cantidad de modelos determina un lote aceptable de imágenes para el entrenamiento. Además se comprobó el funcionamiento de la red utilizando imágenes RGB capturadas de voluntarios reales del *Proyecto 3D Lab*. En Figura 6.7 se visualiza los frutos del entrenamiento de la red. Cada una de las fila es el resultado de corresponder tres imágenes al mismo tiempo. La red es capaz de trasladar los puntos entre imágenes incluso con diferentes proyecciones y formas. Cabe aclarar que el *pipeline* fue entrenado con imágenes en blanco y negro similares a un dibujo y aún así se observa en la Figura 6.7, que el funcionamiento del descriptor es correcto utilizando fotografías a color, inclusive conservando el fondo original.

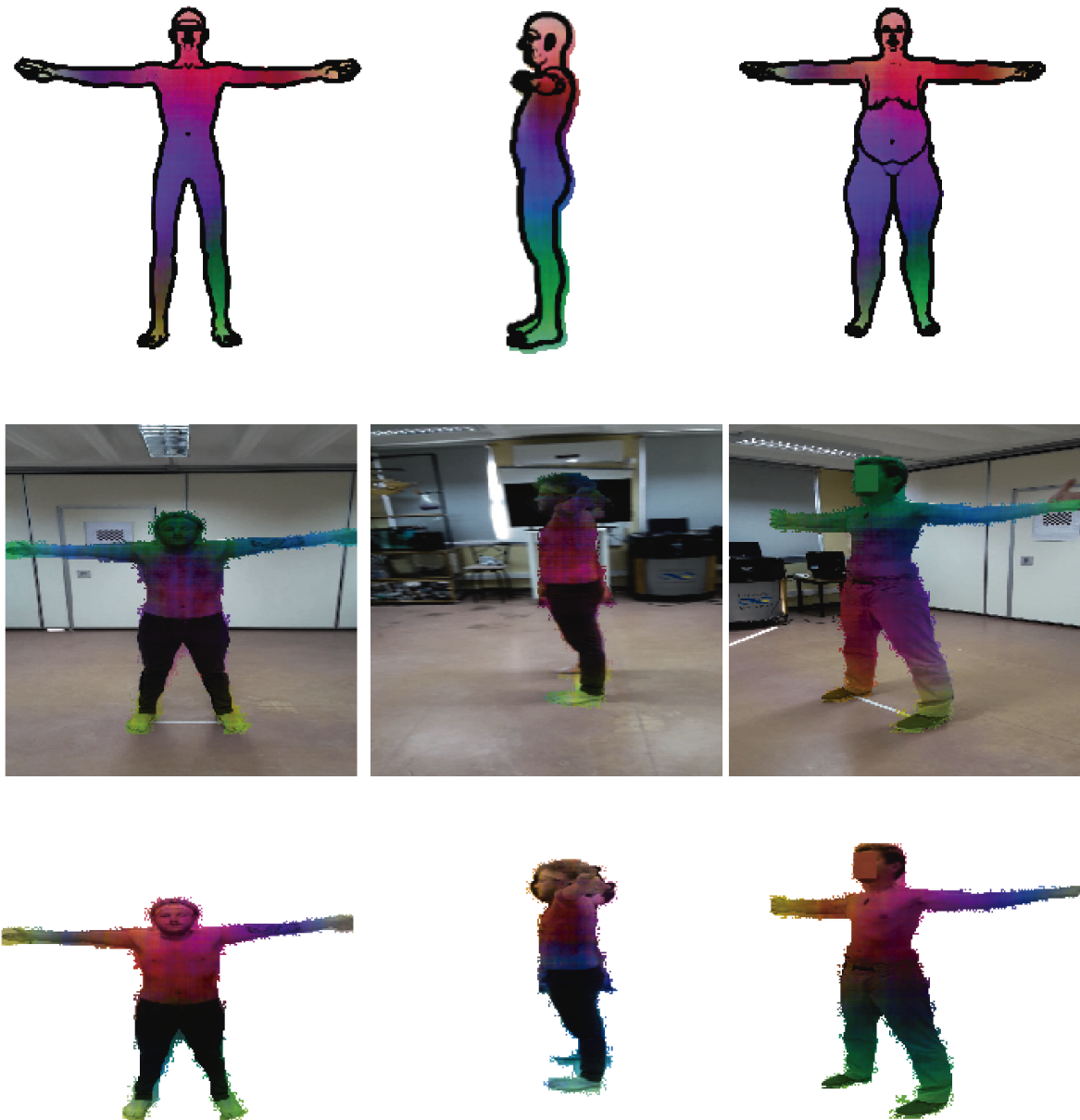


Figura 6.7: Resultados preliminares de la red en cuerpos humanos. Cada fila determina la correspondencia densa entre tres imágenes. Los puntos que se corresponden cuentan con el mismo color.

6.5. Conclusión y trabajo futuro

Se presentó un nuevo descriptor automático de imágenes basado en aprendizaje profundo para obtener correspondencias en bocetos. La técnica aquí propuesta, está basada en una red neuronal convolucional de múltiples vistas, la cual aprende a discriminar punto a punto a través de un espacio de *feature*. *SketchZooms* es el primer enfoque basado en datos que aprende automáticamente descriptores semánticamente coherentes para coincidir en

bocetos con un contexto de múltiples vistas. La realización de esto con redes neuronales era inviable hasta el momento debido a la limitación de datos, ya que recopilar bocetos de artistas y diseñadores es extremadamente costoso. Para abordar este problema, se utilizaron colecciones de formas 3D registradas semánticamente y fueron representadas como bocetos utilizando técnicas de representación no fotorrealistas de vanguardia. Más importante aún, la técnica propuesta fue capaz de generalizar a bocetos reales directamente a partir de los datos sintéticos.

Los resultados ofrecen interesantes interrogantes para futuras investigaciones. Una de estas es si otras configuraciones de vistas son posibles sin introducir ambigüedad en el espacio del descriptor. Sin embargo, esto está limitado por el tiempo dedicado a generar representaciones de bocetos sintéticos. En el aspecto técnico, los enfoques recientes han propuesto utilizar imágenes dibujadas a mano semi-supervisadas para mejorar el rendimiento de la red ([Simo-Serra et al., 2018](#)). Además, al realizar el estudio del usuario en la Sección 6.4.2, quedó en evidencia que los participantes aplican diferentes criterios al seleccionar los puntos. Sería invaluable obtener una visión más profunda de cómo los humanos realizan tareas de correspondencia en el dominio de la imagen.

Capítulo 7

Conclusiones Generales

7.1. Conclusiones

El avance de la tecnología ha permitido el acople masivo de datos asociados a la salud poniendo a disposición de la ciencia y de la medicina, grandes bases de datos con información fenotípica. Estos datos incluyen desde imágenes 2D hasta escaneos tridimensionales del cuerpo. Con dichos bio-bancos creciendo y estableciéndose cada vez con mayor solidez, es necesario investigar e innovar en el desarrollo de nuevas técnicas para el apoyo al diagnóstico y seguimiento de los pacientes. Estas nuevas innovaciones mejoran y aceleran tanto la práctica clínica como el diseño de políticas innovadoras en salud pública.

La heredabilidad de muchos fenotipos, como la distribución del tejido en el cuerpo, el peso, la altura o la masa libre de grasa, ha sido estimada en distintas poblaciones y diferentes grupos de edad, con observaciones consistentes de la contribución de factores genéticos a la variación de estos rasgos ([Loos and Bouchard, 2003](#)). Tanto los factores genéticos como los no genéticos pueden predisponer a ciertas personas a distintas enfermedades, por lo que estudiarlas dentro de una población específica es de gran importancia para la prevención, diagnóstico y tratamiento de estas afecciones. La necesidad de tener datos de población específica es crucial para desarrollar y validar nuevas herramientas tanto en el ámbito de la ciencia como de la medicina clínica. En la investigación biomédica, el avance de la tecnología en la captura de datos complejos debe acompañarse con algoritmos y procedimientos que

permitan extraer automáticamente información, a fin de hacer comparaciones intra e inter poblacionales, establecer correlaciones y posibles causas, así como también promover el desarrollo de nuevas aplicaciones que asistan en el ámbito clínico.

En esta tesis se propuso investigar y desarrollar técnicas avanzadas que permitan aplicar métodos de computación con el objetivo de generar automáticamente descriptores que representen y discriminen características importantes del fenotipo humano. Dentro de este marco, es necesario que estos nuevos descriptores cuenten con una fuerte correlación sobre los índices utilizados actualmente por los expertos en el ámbito clínico. Utilizando información de la población local, se diseñaron, implementaron y evaluaron tres tipos de descriptores utilizando métodos avanzados en procesamiento de imágenes junto con técnicas de ML. El primero emplea información de escaneos 3D del cuerpo, con el objetivo de construir un espacio de descriptores capaz de discriminar el sexo biológico de una persona (Navarro et al., 2018). Determinar esta característica en un individuo es el factor principal a la hora de analizar la ubicación del tejido adiposo para estudiar en detalle la obesidad y el sobrepeso (Geer and Shen, 2009, Power and Schulkin, 2008). En segundo lugar se propuso un método de trabajo que permite a partir de procesamiento en 3D, generar un descriptor para cuantificar las estimaciones de sobrepeso y obesidad (Navarro et al., 2019a). Utilizando análisis estadístico se compararon los índices clásicos utilizados en la práctica clínica (IMC, ICC e ICT.) junto con el espacio del descriptor, la evidencia determinó, que este espacio construido preserva fielmente la naturaleza de la distribución de adiposidad abdominal como un aspecto de la forma. Por último se presentó un descriptor basado en DL junto con un *pipeline* que permite obtener espacios de características capaces de discriminar puntos en imágenes 2D con diferencias en la proyección y vista del objeto de interés (Navarro et al., 2019b). Más precisamente, el pipeline se basa en una red neuronal convolucional de múltiples vistas, el descriptor aquí presentado es el primer enfoque que permite aprender automáticamente descriptores semánticamente coherentes para coincidir en imágenes con un contexto de múltiples vistas.

Los métodos aquí desarrollados son un primer enfoque a la caracterización del fenotipo humano. Los descriptores diseñados, implementados y evaluados permiten a los investigadores obtener de manera masiva información sobre la forma y variabilidad del cuerpo

humano. Esta caracterización dentro de los espacios generados provee el uso de metodologías apropiadas para proporcionar una detección y seguimiento de enfermedades. Además permite una mejor y más completa gestión de la información, obteniendo resultados rápidamente visibles en la recolección, la verificación de hipótesis de diverso tipo la animación de los resultados, la extrapolación de predicciones, etc., generando herramientas para poder manejar de forma rápida, correcta y eficaz la gran cantidad de datos con la que se encuentran los investigadores. Además permite impulsar investigaciones interdisciplinarias novedosas, entre ellas la determinación de patrones de variación intra e inter poblacional, predisposición de determinados grupos étnicos a padecer ciertas enfermedades, el desarrollo de tecnologías aplicadas a la ciencia biológica, etc.

7.2. Trabajo Futuro

Como trabajo futuro, se propone investigar y desarrollar métodos computacionales para analizar las relaciones entre fenotipo y genotipo, buscando determinar aquellas regiones del genoma humano donde existen variantes o marcadores genéticos asociados a determinados rasgos físicos. Además se propone continuar en la investigación de descriptores automáticos de forma asociados a genotipos, para contribuir al diagnóstico y seguimiento de condiciones como obesidad, sobrepeso y condiciones asociadas. El diseño de dichos descriptores permitirá estudios de diversidad específicos de población y el desarrollo de aplicaciones informáticas en medicina y bioantropología. Se profundizará en el uso de metodologías de *Machine Learning* para el diseño y desarrollo de los modelos que permitirán analizar y procesar la información ([Pazos et al., 2018](#), [Cintas et al., 2019](#)).

Con la iniciativa PoblAr ([Dopazo et al., 2019](#)) en mente, el acople masivo de datos permitirá comprender cómo un conjunto de genes interactúa entre sí y con diversos factores ambientales en la propia población argentina, para dar como resultado un determinado fenotipo (por ejemplo: una enfermedad o un aumento en la susceptibilidad de padecerla). Como ya se mencionó con anterioridad estas particularidades genómicas y ambientales, así como sus tipos de interacción, son específicos de cada población, de modo que la construcción de estos mapas genotipo/fenotipo/ambiente también lo son.

La información posee un altísimo valor en la investigación, pero a su vez también en la determinación de políticas públicas preventivas y en la mejora de la calidad de los servicios de salud. Las aplicaciones en relación con el proyecto PoblAr serían innumerables. Entre las que se pueden destacar: la realización de prototipos de prueba que permitan la predicción de fenotipos a partir de marcadores genéticos y sus correspondientes variaciones, la investigación y desarrollo de algoritmos de búsqueda en bases de datos genotipo-fenotipo junto con la correlación de información poblacional a fin de obtener un conjunto de predictores que permitan inferir tendencias de interés sanitario (estimaciones de riesgo de diferentes grupos étnicos frente a ciertas dolencias, influencias de factores ambientales, etc.).

Existen casos de éxito aplicando *Deep Learning* a estudios de biología molecular, por ejemplo, para la secuenciación del genoma. Entrenando modelos mejora su extracción, se pueden capturar múltiples escalas genómicas sin necesidad de cambiar de método y evitando las tareas manuales (Alipanahi et al., 2015, Zhou and Troyanskaya, 2015). Otras investigaciones utilizan estos tipos de métodos para resolver la asociación genotipo-fenotipo, como el caso de Ma et al. (2018). Los autores predicen fenotipos a partir de genotipos utilizando una red neuronal convolucional. Cabe aclarar que la aplicación de estas técnicas en el ámbito de salud se ven afectadas por la falta de datos. Recopilar y mantener suficientes datos médicos con los detalles necesarios para que los métodos de *Deep Learning* funcionen correctamente es costoso y laborioso (Che et al., 2017).

El grupo de investigación cuenta con el material colectado por el consorcio CANDELA (Ruderman et al., 2019, Ruiz-Linares et al., 2014, Quinto-Sánchez et al., 2015, 2017, Adhikari et al., 2019), que es administrado y compilado en el IPCSH (CCT CONICET CENPAT) por el Dr. González-José. Comprende información asociada de 7500 individuos, residentes de ciudades cosmopolitas de cinco países (México, Perú, Colombia, Brasil, Chile). Este banco incluye fotografías faciales 2D, datos antropométricos (peso, talla, perímetro de cintura y cadera), socio-culturales, de autopercepción y genotipos; es decir, para cada persona se han obtenido, mediante secuenciación de microarreglos, 600.000 variantes polimórficas o mutaciones de punto de ADN (Ruiz-Linares et al., 2014). Además, para aumentar la cantidad de datos, se utilizan algoritmos que permiten aumentar artificialmente la información pero conservando la distribución y estructura de los datos reales.

En análisis de imágenes existen trabajos previos donde el aumento de datos se realiza utilizando métodos tradicionales (rotar, re-dimensionar, voltear, cambiar de color., etc.) y, más recientemente, existen modelos que sintetizan datos completamente nuevos ([Wei et al., 2019](#)). Estos modelo se denominan Redes Generativas Antagónicas (conocidas como GANs en inglés) las cuales se basan en entrenar simultáneamente dos redes neuronales, una generativa que captura la distribución de los datos, y otra discriminativa que distingue los datos generados de datos originales. Las redes GANs se utilizan principalmente para comprender y replicar características en datos de imagen, video y texto ([Che et al., 2017](#), [Wei et al., 2019](#)). La relevancia de las GANs para los objetivos de este plan de trabajo no ha sido probada aún, pero consideramos que es una vía metodológicamente plausible para acometer el estudio del cruce geno-feno en particular, y el mapa genotipo fenotipo, incluyendo aspectos del desarrollo en general en contextos acotados ([Pigliucci, 2010](#)) como el planteado aquí relacionado con el sobrepeso.

Anexos

Anexo A

Cartilla de Información

"Nos complace invitarlo/a a participar del proyecto. Buscamos desarrollar una aplicación para dispositivos móviles (teléfonos celulares y tablets) para realizar escaneos corporales en tres dimensiones que permitan cuantificar la forma corporal y obtener medidas precisas. Mediante un procedimiento simple y no invasivo, con una sencilla filmación obtenida con un celular, la aplicación calculará variables y volúmenes, permitiendo diagnosticar obesidad o condiciones similares, accediendo a su vez a la historia clínica del paciente para conocer datos específicos, su evolución particular o valores de referencia de la Organización Mundial de la Salud. Poder generar este conocimiento con recursos y profesionales argentinos es altamente beneficioso para una mejor atención médica así como para el desarrollo de políticas de Salud Pública más eficaces. Para alcanzar este objetivo, necesitamos formar una base de referencia, integrada por filmaciones a voluntarias/os. Si Usted decide participar, se le realizarán dos tomas de vídeo de 360 grados con diferentes dispositivos y se registrarán un conjunto de medidas corporales (circunferencia de la cintura, de la cadera, del muslo, del bíceps, pliegues subescapular y bicipital, la estatura y el peso corporal). Todos estos procedimientos son absolutamente indoloros y se realizaran en un ambiente apropiado bajo normas de estricta confidencialidad y privacidad. El procesamiento posterior de la imagen impide la identificación, garantizando el anonimato y protección de su identidad"

Anexo B

Comité de ética

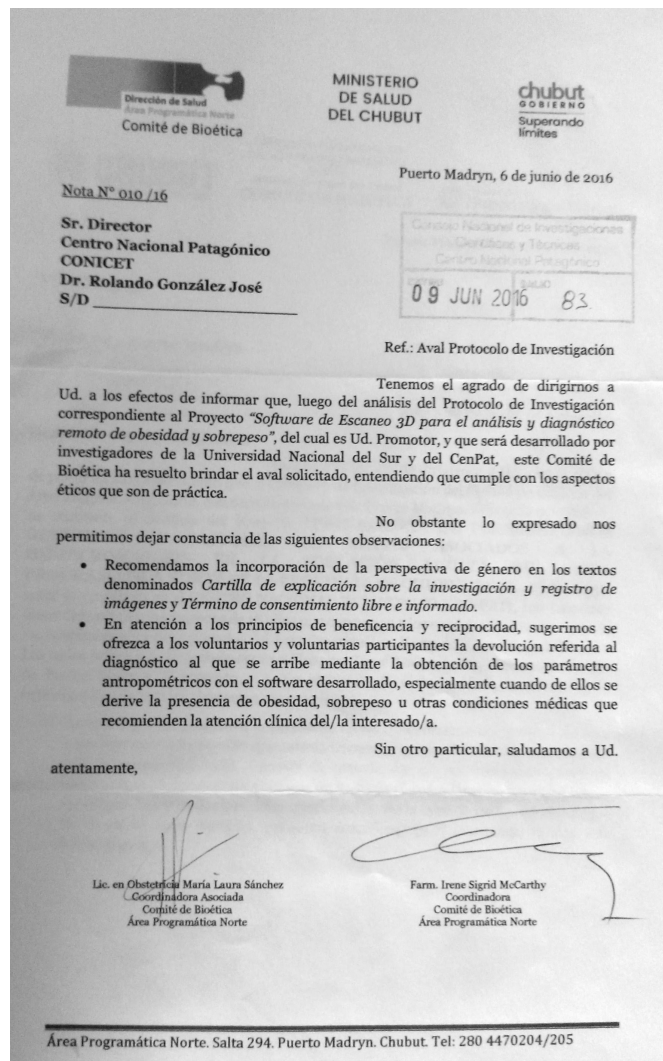


Figura B.1: Aval del comité de ética.

Bibliografía

Tilke Judd, Frédo Durand, and Edward Adelson. Apparent ridges for line drawing. *ACM transactions on graphics (TOG)*, 26(3):19–es, 2007.

Dronica. Nube de puntos 3d es el primer producto resultante del escaner láser. <https://dronica.es/nube-de-puntos-que-es-y-para-que-sirve/>. (Accessed on 11/19/2020).

ItSeez3D. itseez3d turns ipad into ünbelievable”3d scanner with help of a structure sensor - 3dprint.com | the voice of 3d printing / additive manufacturing. <https://3dprint.com/7738/itseez3d-3d-scanning-app/>. (Accessed on 11/19/2020).

Essam F Elsayed, Hocine Tighiouart, Daniel E Weiner, John Griffith, Deeb Salem, Andrew S Levey, and Mark J Sarnak. Waist-to-hip ratio and body mass index as risk factors for cardiovascular events in ckd. *American Journal of Kidney Diseases*, 52(1):49–57, 2008.

Pablo Navarro, Bruno Pazos, Celia Cintas, Virginia Ramallo, Rolando González-José, and Claudio Delrieux. Gender recognition using 3d human body scans. *2018 IEEE Biennial Congress of Argentina*, page 6, 2018.

Pablo Navarro, Virginia Ramallo, Celia Cintas, Anahí Ruderman, Soledad de Azevedo, Carolina Paschetta, Orlando Pérez, Bruno Pazos, Claudio Delrieux, and Rolando González-José. Body shape: Implications in the study of obesity and related traits. *American Journal of Human Biology*, page e23323, 2019a.

Pablo Navarro, José Ignacio Orlando, Claudio Delrieux, and Emmanuel Iarussi. Sketch-zooms: Deep multi-view descriptors for matching line drawings. *arXiv preprint arXiv:1912.05019*, 2019b.

WHO. WHO :: Global Database on Body Mass Index. Technical report, 2018. URL www.who.int/bmi.

Lucy M Browning, Shiun Dong Hsieh, and Margaret Ashwell. A systematic review of waist-to-height ratio as a screening tool for the prediction of cardiovascular disease and diabetes: 0.5 could be a suitable global boundary value. *Nutrition research reviews*, 23(2):247–269, 2010.

Massimo Pigliucci. Genotype–phenotype mapping and the end of the ‘genes as blueprint’ metaphor. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1540):557–566, 2010.

Ruth JF Loos and Claude Bouchard. Obesity—is it a genetic disorder? *Journal of internal medicine*, 254(5):401–425, 2003.

Eitan Segev, Yoram Hemo, Shlomo Wientroub, Dror Ovadia, Michael Fishkin, David M Steinberg, and Shlomo Hayek. Intra-and interobserver reliability analysis of digital radiographic measurements for pediatric orthopedic parameters using a novel pacs integrated computer software program. *Journal of children's orthopaedics*, 4(4):331–341, 2010.

Axel Kamoen, Luc Dermaut, and Ronald Verbeeck. The clinical significance of error measurement in the interpretation of treatment results. *The European Journal of Orthodontics*, 23(5):569–578, 2001.

Jian Zhou, Chandra L Theesfeld, Kevin Yao, Kathleen M Chen, Aaron K Wong, and Olga G Troyanskaya. Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. *Nature genetics*, 50(8):1171–1179, 2018.

Babak Alipanahi, Andrew DeLong, Matthew T Weirauch, and Brendan J Frey. Predicting the sequence specificities of dna-and rna-binding proteins by deep learning. *Nature biotechnology*, 33(8):831–838, 2015.

Qiangguo Jin, Zhaopeng Meng, Changming Sun, Leyi Wei, and Ran Su. Ra-unet: A hybrid deep attention-aware network to extract liver and tumor in ct scans. *arXiv preprint arXiv:1811.01328*, 2018.

- Hao Dong, Guang Yang, Fangde Liu, Yuanhan Mo, and Yike Guo. Automatic brain tumor detection and segmentation using u-net based fully convolutional networks. In *annual conference on medical image understanding and analysis*, pages 506–517. Springer, 2017.
- Riccardo Miotto, Li Li, Brian A Kidd, and Joel T Dudley. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Scientific reports*, 6(1):1–10, 2016.
- C Begué, M Galante, M Gaudio, Goldberg L, A King, V O'donel, and V Sciarretta. *Tercera Encuesta Nacional de Factores de Riesgo para Enfermedades No Transmisibles*. Ministerio de Salud de la Nación, Instituto Nacional de Estadísticas y Censos, 2013.
- Mariana Galante, Victoria O'donel, Magalí Gaudio, Carolina Begué, Ana King, and Lucila Goldberg. Situación epidemiológica de la obesidad en Argentina. *Revista Argentina de Cardiología*, 2016. ISSN 18503748. doi: 10.7775/rac.v84.i2.8028.
- M Ashwell, P Gunn, and S Gibson. Waist-to-height ratio is a better screening tool than waist circumference and BMI for adult cardiometabolic risk factors: systematic review and meta-analysis. *Obesity reviews : an official journal of the International Association for the Study of Obesity*, 13(3):275–86, 2012. ISSN 1467-789X. doi: 10.1111/j.1467-789X.2011.00952.x. URL <http://www.ncbi.nlm.nih.gov/pubmed/22106927>.
- Margaret Ashwell, Les Mayhew, Jon Richardson, and Ben Rickayzen. Waist-to-height ratio is more predictive of years of life lost than body mass index. *PLoS ONE*, 9(9), 2014. ISSN 19326203. doi: 10.1371/journal.pone.0103483.
- Nir Y. Krakauer and Jesse C. Krakauer. A new body shape index predicts mortality hazard independently of body mass index. *PLoS ONE*, 7(7), 2012. ISSN 19326203. doi: 10.1371/journal.pone.0039504.
- Guy Fagherazzi, Alice Vilier, Beverley Balkau, Françoise Clavel-Chapelon, and Dianna J. Magliano. Anthropometrics, body shape over 12 years and risk of cancer events in pre- and post-menopausal women. *International Journal of Cancer*, 133(3):740–748, 2013. ISSN 00207136. doi: 10.1002/ijc.28069.

- Abel Romero-Corral, Victor M. Montori, Virend K. Somers, Josef Korinek, Randal J. Thomas, Thomas G. Allison, Farouk Mookadam, and Francisco Lopez-Jimenez. Association of bodyweight with total mortality and with cardiovascular events in coronary artery disease: a systematic review of cohort studies. *The Lancet*, 368(9536):666–678, 2006. ISSN 0140-6736. doi: 10.1016/S0140-6736(06)69251-9. URL <http://www.sciencedirect.com/science/article/pii/S0140673606692519><http://www.sciencedirect.com/science/article/pii/S0140673606692519/pdf?md5=f87cd76cc92f78185030176641737877&pid=1-s2.0-S0140673606692519-main.pdf>.
- Henry Löffler-Wirth, Edith Willscher, Peter Ahnert, Kerstin Wirkner, Christoph Engel, Markus Loeffler, and Hans Binder. Novel anthropometry based on 3D-bodyscans applied to a large population based cohort. *PLoS ONE*, 11(7), 2016. ISSN 19326203. doi: 10.1371/journal.pone.0159887.
- Fernando Lanas, Pamela Serón, Sergio Muñoz, Paula Margozzini, and Teresa Puig. Latin american clinical epidemiology network series – paper 7: Central obesity measurements better identified risk factors for coronary heart disease risk in the chilean national health survey (2009–2010). *Journal of Clinical Epidemiology*, 86:111 – 116, 2017. ISSN 0895-4356. doi: <https://doi.org/10.1016/j.jclinepi.2016.04.018>. URL <http://www.sciencedirect.com/science/article/pii/S089543561630556X>.
- Ancel Keys, Flaminio Fidanza, Martti J. Karvonen, Noboru Kimura, and Henry L. Taylor. Indices of relative weight and obesity. *Journal of Chronic Diseases*, 25(6-7):329–343, 1972. ISSN 00219681. doi: 10.1016/0021-9681(72)90027-6.
- B Larsson, K Svardsudd, L Welin, L Wilhelmsen, P Bjorntorp, and G Tibblin. Abdominal adipose tissue distribution, obesity, and risk of cardiovascular disease and death: 13 year follow up of participants in the study of men born in 1913. *BMJ*, 288(6428): 1401–1404, 1984. ISSN 0959-8138. doi: 10.1136/bmj.288.6428.1401. URL <http://www.bmj.com/cgi/doi/10.1136/bmj.288.6428.1401>.
- L Lapidus, C Bengtsson, B Larsson, K Pennert, E Rybo, and L Sjöström. Distribution of adipose tissue and risk of cardiovascular disease and death: a 12 year follow up of

- participants in the population study of women in Gothenburg, Sweden. *British medical journal (Clinical research ed.)*, 289(6454):1257–61, 1984. ISSN 0267-0623. doi: 10.1136/bmj.289.6454.1257.
- Jingjing Sun, Bugao Xu, Jane Lee, and Jeanne H Freeland-Graves. Novel body shape descriptors for abdominal adiposity prediction using magnetic resonance images and stereovision body images. *Obesity*, 25(10):1795–1801, 2017.
- Thorsten Bretschneider, Urte Koop, Volker Schreiner, Horst Wenck, and Sören Jaspers. Validation of the body scanner as a measuring tool for a rapid quantification of body shape. *Skin Research and Technology*, 15(3):364–369, 2009. ISSN 0909752X. doi: 10.1111/j.1600-0846.2009.00374.x.
- M. Reese Pepper, Jeanne H. Freeland-Graves, Wurong Yu, Philip R. Stanforth, Jodi M. Cahill, Michael Mahometa, and Bugao Xu. Validation of a 3-dimensional laser body scanner for assessment of waist and hip circumference. *Journal of the American College of Nutrition*, 29(3):179–188, 2010. ISSN 15411087. doi: 10.1080/07315724.2010.10719832.
- Lina Jaeschke, Astrid Steinbrecher, and Tobias Pischon. Measurement of waist and hip circumference with a body surface scanner: Feasibility, validity, reliability, and correlations with markers of the metabolic syndrome. *PLoS ONE*, 10(3), 2015. ISSN 19326203. doi: 10.1371/journal.pone.0119430.
- Jonathan CK Wells, Philip Treleaven, and Tim J Cole. Bmi compared with 3-dimensional body shape: the uk national sizing survey. *The American Journal of Clinical Nutrition*, 85(2):419–425, 2007. doi: 10.1093/ajcn/85.2.419. URL <http://dx.doi.org/10.1093/ajcn/85.2.419>.
- 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>.
- Anahí Ruderman, Luis O Pérez, Kaustubh Adhikari, Pablo Navarro, Virginia Ramallo, Carla Gallo, Giovanni Poletti, Gabriel Bedoya, Maria C Bortolini, Victor Acuña-Alonzo, et al.

- Obesity, genomic ancestry, and socioeconomic variables in latin american mestizos. *American Journal of Human Biology*, 31(5):e23278, 2019.
- Hernán Dopazo, Andrea Llera, Mariana Berenstein, and Rolando González José. Genomas, enfermedades y medicina de precisión: un proyecto nacional. *Ciencia, Tecnología y Política*, 2019.
- Magda Alexandra Trujillo-Jiménez, Pablo Navarro, Bruno Pazos, Leonardo Morales, Virginia Ramallo, Carolina Paschetta, Soledad De Azevedo, Anahí Ruderman, Orlando Pérez, Claudio Delrieux, et al. body2vec: 3d point cloud reconstruction for precise anthropology with handheld devices. *Journal of Imaging*, 6(9):94, 2020.
- Thomas Huang. *Computer vision: Evolution and promise*. 1996.
- Jyoti Islam and Yanqing Zhang. A novel deep learning based multi-class classification method for alzheimer's disease detection using brain mri data. In *International Conference on Brain Informatics*, pages 213–222. Springer, 2017.
- Kanishka Sharma, Christian Rupprecht, Anna Caroli, Maria Carolina Aparicio, Andrea Remuzzi, Maximilian Baust, and Nassir Navab. Automatic Segmentation of Kidneys using Deep Learning for Total Kidney Volume Quantification in Autosomal Dominant Polycystic Kidney Disease. *Scientific reports*, 7(1):2049, may 2017. ISSN 2045-2322. doi: 10.1038/s41598-017-01779-0. URL <http://www.nature.com/articles/s41598-017-01779-0><http://www.ncbi.nlm.nih.gov/pubmed/28515418><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5435691>.
- Yu-Jen Yu-Jen Chen, Kai-Lung Hua, Che-Hao Hsu, Wen-Huang Cheng, and Shintami Chusnul Hidayati. Computer-aided classification of lung nodules on computed tomography images via deep learning technique. *OncoTargets and Therapy*, 8:2015, aug 2015. ISSN 1178-6930. doi: 10.2147/OTT.S80733. URL <http://www.dovepress.com/computer-aided-classification-of-lung-nodules-on-computed-tomography-i-peer-reviewed-article-OTT>.
- Ryan Poplin, Avinash V Varadarajan, Katy Blumer, Yun Liu, Michael V McConnell, Greg S Corrado, Lily Peng, and Dale R Webster. Prediction of cardiovascular risk factors from

retinal fundus photographs via deep learning. *Nature Biomedical Engineering*, 2(3): 158–164, 2018a. ISSN 2157-846X. doi: 10.1038/s41551-018-0195-0. URL <https://doi.org/10.1038/s41551-018-0195-0>.

Huahui Xiong, Xiaoqing Huang, Yong Li, Jianhong Li, Junfang Xian, and Yaqi Huang. A Method for Accurate Reconstructions of the Upper Airway Using Magnetic Resonance Images. *PloS one*, 10(6):e0130186, 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0130186. URL <http://www.ncbi.nlm.nih.gov/pubmed/26066461><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4465749>.

Dominik C. Benz, Georgios Benetos, Georgios Rampidis, Elia von Felten, Adam Bakula, Aleksandra Sustar, Ken Kudura, Michael Messerli, Tobias A. Fuchs, Catherine Gebhard, Aju P. Pazhenkottil, Philipp A. Kaufmann, and Ronny R. Buechel. Validation of deep-learning image reconstruction for coronary computed tomography angiography: Impact on noise, image quality and diagnostic accuracy. *Journal of Cardiovascular Computed Tomography*, 2020. ISSN 1934-5925. doi: <https://doi.org/10.1016/j.jcct.2020.01.002>. URL <http://www.sciencedirect.com/science/article/pii/S1934592519304642>.

Celia Cintas, Manuel Lucena, José Manuel Fuertes, Claudio Delrieux, Pablo Navarro, Rolando González-José, and Manuel Molinos. Automatic feature extraction and classification of iberian ceramics based on deep convolutional networks. *Journal of Cultural Heritage*, 41:106 – 112, 2020. ISSN 1296-2074. doi: <https://doi.org/10.1016/j.culher.2019.06.005>. URL <http://www.sciencedirect.com/science/article/pii/S1296207418307775>.

Priyan Malarvizhi Kumar, Ushadevi Gandhi, R Varatharajan, Gunasekaran Manogaran, R Jidhesh, and Thanjai Vadivel. Intelligent face recognition and navigation system using neural learning for smart security in internet of things. *Cluster Computing*, 22(4):7733–7744, 2019.

Divyarajsinh N Parmar and Brijesh B Mehta. Face recognition methods & applications. *arXiv preprint arXiv:1403.0485*, 2014.

Wenbo Guo, Dongliang Mu, Jun Xu, Purui Su, Gang Wang, and Xinyu Xing. Lemna:

- Explaining deep learning based security applications. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, pages 364–379, 2018.
- 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, et al. Automatic ear detection and feature extraction using geometric morphometrics and convolutional neural networks. *IET Biometrics*, 6(3):211–223, 2016.
- J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.
- Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton. Cifar-10 and cifar-100 datasets. *URI: <https://www.cs.toronto.edu/kriz/cifar.html>*, 6, 2009.
- Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.
- Mark Everingham, Luc Van Gool, Christopher K I Williams, John Winn, and Andrew Zisserman. The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*, 88(2):303–338, 2010. ISSN 1573-1405. doi: 10.1007/s11263-009-0275-4. URL <https://doi.org/10.1007/s11263-009-0275-4>.
- Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *European conference on computer vision*, pages 404–417. Springer, 2006.
- David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.
- Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua. Brief: Binary robust independent elementary features. In *European conference on computer vision*, pages 778–792. Springer, 2010.
- Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. Orb: An efficient alternative to sift or surf. In *2011 International conference on computer vision*, pages 2564–2571. Ieee, 2011.

- 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, 2011a.
- Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.
- Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.
- Stuart Russel, Peter Norvig, et al. *Artificial intelligence: a modern approach*. Pearson Education Limited, 2013.
- M Khorshid, T Abou-El-Enien, and G Soliman. A comparison among support vector machine and other machine learning classification algorithms. *IPASJ International Journal of Computer Science*, 3(5):26–35, 2015.
- Lior Rokach and Oded Z Maimon. *Data mining with decision trees: theory and applications*, volume 69. World scientific, 2008.
- Pierre Geurts, Damien Ernst, and Louis Wehenkel. Extremely randomized trees. *Machine learning*, 63(1):3–42, 2006.
- Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *nature*, 542(7639):115–118, 2017.
- HA Haenssle, C Fink, R Schneiderbauer, F Toberer, T Buhl, A Blum, A Kalloo, A Ben Hadj Hassen, L Thomas, A Enk, et al. Reader study level-i and level-ii groups. man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann Oncol*, 29(8):1836–1842, 2018.
- Jie-Zhi Cheng, Dong Ni, Yi-Hong Chou, Jing Qin, Chui-Mei Tiu, Yeun-Chung Chang, Chiun-Sheng Huang, Dinggang Shen, and Chung-Ming Chen. Computer-aided diagnosis

- with deep learning architecture: applications to breast lesions in us images and pulmonary nodules in ct scans. *Scientific reports*, 6(1):1–13, 2016.
- Mark Cicero, Alexander Bilbily, Errol Colak, Tim Dowdell, Bruce Gray, Kuhan Perampaladas, and Joseph Barfett. Training and validating a deep convolutional neural network for computer-aided detection and classification of abnormalities on frontal chest radiographs. *Investigative radiology*, 52(5):281–287, 2017.
- Varun Gulshan, Lily Peng, Marc Coram, Martin C Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Kasumi Widner, Tom Madams, Jorge Cuadros, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, 316(22):2402–2410, 2016.
- Ryan Poplin, Avinash V Varadarajan, Katy Blumer, Yun Liu, Michael V McConnell, Greg S Corrado, Lily Peng, and Dale R Webster. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nature Biomedical Engineering*, 2(3):158, 2018b.
- Dan C Cireşan, Alessandro Giusti, Luca M Gambardella, and Jürgen Schmidhuber. Mitosis detection in breast cancer histology images with deep neural networks. In *International conference on medical image computing and computer-assisted intervention*, pages 411–418. Springer, 2013.
- Pornpimol Charoentong, Francesca Finotello, Mihaela Angelova, Clemens Mayer, Mirjana Efremova, Dietmar Rieder, Hubert Hackl, and Zlatko Trajanoski. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell reports*, 18(1):248–262, 2017.
- Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436, 2015.
- Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.

- Kouichi Yamaguchi, Kenji Sakamoto, Toshio Akabane, and Yoshiji Fujimoto. A neural network for speaker-independent isolated word recognition. In *First International Conference on Spoken Language Processing*, 1990.
- Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? In *Advances in neural information processing systems*, pages 3320–3328, 2014.
- Paul GM Knoops, Caroline AA Beaumont, Alessandro Borghi, Naiara Rodriguez-Florez, Richard WF Breakey, William Rodgers, Freida Angullia, NU Owase Jeelani, Silvia Schievano, and David J Dunaway. Comparison of three-dimensional scanner systems for craniomaxillofacial imaging. *Journal of Plastic, Reconstructive & Aesthetic Surgery*, 70(4):441–449, 2017.
- Carlo M Oranges, Srinivas Madduri, Philipp Brantner, Bilal Msallem, Salvatore Giordano, Benito Benitez, Daniel F Kalbermatten, Dirk J Schaefer, and Florian M Thieringer. Three-dimensional assessment of the breast: Validation of a novel, simple and inexpensive scanning process. *in vivo*, 33(3):839–842, 2019.
- MD Cabañas and F Esparza. Compendio de cineantropometría. *Madrid: CTO Editorial*, 2, 2009.
- Philip Treleaven and Jonathan Wells. 3d body scanning and healthcare applications. *IEEE Computer*, 40(7), 2007.
- Eliza B Geer and Wei Shen. Gender differences in insulin resistance, body composition, and energy balance. *Gender medicine*, 6:60–75, 2009.
- Michael L Power and Jay Schulkin. Sex differences in fat storage, fat metabolism, and the health risks from obesity: possible evolutionary origins. *British Journal of Nutrition*, 99(5):931–940, 2008.

- Valeria Guglielmi and Paolo Sbraccia. Obesity phenotypes: depot-differences in adipose tissue and their clinical implications. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity*, 23(1):3–14, 2018.
- Zouhour Ben Azouz, Chang Shu, and Anja Mantel. Automatic locating of anthropometric landmarks on 3d human models. In *3D Data Processing, Visualization, and Transmission, Third International Symposium on*, pages 750–757. IEEE, 2006.
- Morteza Daneshmand, Alvo Aabloo, Cagri Ozcinar, and Gholamreza Anbarjafari. Real-time, automatic shape-changing robot adjustment and gender classification. *Signal, Image and Video Processing*, 10(4):753–760, 2016.
- 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.
- Baback Moghaddam and Ming Hsuan Yang. Learning gender with support faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002. ISSN 01628828. doi: 10.1109/34.1000244.
- Xuelong Li, Stephen J. Maybank, Shuicheng Yan, Dacheng Tao, and Dong Xu. Gait components and their application to gender recognition, 2008. ISSN 10946977.
- Shanthi.G.C and Saravanan.E. Background Subtraction Techniques: Systematic Evaluation and Comparative Analysis. *International Journal of Modern Engineering Research (IJMER) www.ijmer.com*, 3(1):514–517.
- Kyunghnam Kim, Thanarat H. Chalidabhongse, David Harwood, and Larry Davis. Real-time foreground–background segmentation using codebook model. *Real-Time Imaging*, 11(3):172–185, jun 2005. ISSN 10772014. doi: 10.1016/j.rti.2004.12.004. URL <http://linkinghub.elsevier.com/retrieve/pii/S1077201405000057>.
- Liangliang Cao, Mert Dikmen, Yun Fu, and Thomas S Huang. Gender recognition from body. In *Proceedings of the 16th ACM international conference on Multimedia*, pages 725–728. ACM, 2008.

- Aaron Fouts, Ryan McCoppin, Mateen Rizki, Louis Tamburino, and Olga Mendoza-Schrock. Exploring point-cloud features from partial body views for gender classification. pages 84020L–84020L–9, may 2012. doi: 10.1117/12.921880. URL <http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1354883>.
- Caifeng Shan. Learning local binary patterns for gender classification on real-world face images. *Pattern Recognition Letters*, 33(4):431–437, 2012. ISSN 01678655. doi: 10.1016/j.patrec.2011.05.016. URL <http://www.sciencedirect.com/science/article/pii/S0167865511001607>.
- B. Moghaddam and Ming-Hsuan Yang. Gender classification with support vector machines. In *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*, pages 306–311. IEEE Comput. Soc, 2000. ISBN 0-7695-0580-5. doi: 10.1109/AFGR.2000.840651. URL <http://ieeexplore.ieee.org/document/840651/>.
- Maros Blaha, Mathias Rothmel, Martin R Oswald, Torsten Sattler, Audrey Richard, Jan D Wegner, Marc Pollefeys, and Konrad Schindler. Semantically informed multiview surface refinement. *arXiv preprint arXiv:1706.08336*, 2017.
- Jinshan Tang, Xiaoming Liu, Huaining Cheng, and Kathleen M Robinette. Gender recognition using 3-d human body shapes. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 41(6):898–908, 2011.
- Jeffrey Scott Vitter. Faster methods for random sampling. *Communications of the ACM*, 27(7):703–718, 1984.
- IK Fodor. A survey of dimension reduction techniques (no. ucl-id-148494). *Lawrence Livermore National Lab., CA (US)*, 2002.
- Yann A. LeCun, Leon Bottou, Genevieve B. Orr, and Klaus Robert Müller. Efficient backprop. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7700 LECTURE NO:9–48, 2012. ISSN 03029743. doi: 10.1007/978-3-642-35289-8-3.

- Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct): 2825–2830, 2011b.
- Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems*, pages 2951–2959, 2012.
- James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *Journal of machine learning research*, 13(Feb):281–305, 2012.
- Chih-Wei Hsu, Chih-Chung Chang, Chih-Jen Lin, et al. A practical guide to support vector classification, 2003.
- Ron Kohavi et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145. Montreal, Canada, 1995.
- Joan F Carroll, Ana L Chiapa, Mayra Rodriguez, David R Phelps, Kathryn M Cardarelli, Jamboor K Vishwanatha, Sejong Bae, and Roberto Cardarelli. Visceral fat, waist circumference, and bmi: impact of race/ethnicity. *Obesity*, 16(3):600–607, 2008.
- Zhan Gu, Ping Zhu, Qiao Wang, Huayu He, Jingjuan Xu, Li Zhang, Dong Li, Jianying Wang, Xiaojuan Hu, Guang Ji, et al. Obesity and lipid-related parameters for predicting metabolic syndrome in chinese elderly population. *Lipids in health and disease*, 17(1): 289, 2018.
- Masaharu Kagawa, CB Binns, and Andrew P Hills. Body composition and anthropometry in japanese and australian caucasian males and japanese females. *Asia Pac J Clin Nutr*, 16(Suppl 1):31–36, 2007.
- Alok Madan, Olga G Archambeau, Vanessa A Milsom, Rachel L Goldman, Jeffery J Borckardt, Anouk L Grubaugh, Peter W Tuerk, and B Christopher Frueh. More than black and white: differences in predictors of obesity among native hawaiian/pacific islanders and european americans. *Obesity*, 20(6):1325–1328, 2012.

- Pernille F Rønn, Gregers S Andersen, Torsten Lauritzen, Dirk L Christensen, Mette Aadahl, Bendix Carstensen, and Marit E Jørgensen. Ethnic differences in anthropometric measures and abdominal fat distribution: a cross-sectional pooled study in inuit, africans and europeans. *J Epidemiol Community Health*, 71(6):536–543, 2017.
- Anthony J Koehl and Jeffrey C Long. The contributions of admixture and genetic drift to diversity among post-contact populations in the americas. *American journal of physical anthropology*, 165(2):256–268, 2018.
- Emily T Norris, Lu Wang, Andrew B Conley, Lavanya Rishishwar, Leonardo Mariño-Ramírez, Augusto Valderrama-Aguirre, and I King Jordan. Genetic ancestry, admixture and health determinants in latin america. *BMC genomics*, 19(8):861, 2018.
- Andrés Ruiz-Linares, Kaustubh Adhikari, Victor Acuña-Alonzo, Mirsha Quinto-Sanchez, Claudia Jaramillo, William Arias, Macarena Fuentes, María Pizarro, Paola Everardo, Francisco De Avila, et al. Admixture in latin america: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. *PLoS genetics*, 10(9):e1004572, 2014.
- Philipp Gunz and Philipp Mitteroecker. Semilandmarks: A method for quantifying curves and surfaces. *Hystrix*, 24(1), 2013. ISSN 03941914. doi: 10.4404/hystrix-24.1-6292.
- Manisha Chandalia, Ping Lin, Thanalakshmi Seenivasan, Edward H Livingston, Peter G Snell, Scott M Grundy, and Nicola Abate. Insulin resistance and body fat distribution in south asian men compared to caucasian men. *PloS one*, 2(8):e812, 2007.
- Dympna Gallagher, Steven B Heymsfield, Moonseong Heo, Susan A Jebb, Peter R Murgatroyd, and Yoichi Sakamoto. Healthy percentage body fat ranges: an approach for developing guidelines based on body mass index. *The American journal of clinical nutrition*, 72(3):694–701, 2000.
- EC Rush, JH Goedecke, C Jennings, L Micklesfield, L Dugas, EV Lambert, and LD Plank. Bmi, fat and muscle differences in urban women of five ethnicities from two countries. *International journal of obesity*, 31(8):1232, 2007.

- Elaine C Rush, Ismael Freitas, and Lindsay D Plank. Body size, body composition and fat distribution: comparative analysis of european, maori, pacific island and asian indian adults. *British Journal of Nutrition*, 102(4):632–641, 2009.
- SN Wulan, KR Westerterp, and G Plasqui. Ethnic differences in body composition and the associated metabolic profile: a comparative study between asians and caucasians. *Maturitas*, 65(4):315–319, 2010.
- Jose Medina-Inojosa, Virend K Somers, Taiwo Ngwa, Ling Hinshaw, and Francisco Lopez-Jimenez. Reliability of a 3d body scanner for anthropometric measurements of central obesity. *Obesity, open access*, 2(3), 2016.
- Andreas Kuehnapfel, Peter Ahnert, Markus Loeffler, Anja Broda, and Markus Scholz. Reliability of 3d laser-based anthropometry and comparison with classical anthropometry. *Scientific reports*, 6:26672, 2016.
- Andreas Kuehnapfel, Peter Ahnert, Markus Loeffler, and Markus Scholz. Body surface assessment with 3d laser-based anthropometry: reliability, validation, and improvement of empirical surface formulae. *European journal of applied physiology*, 117(2):371–380, 2017.
- Xingguo Liu, Jianwei Niu, Linghua Ran, and Taijie Liu. Estimation of human body volume (bv) from anthropometric measurements based on three-dimensional (3d) scan technique. *Aesthetic plastic surgery*, 41(4):971–978, 2017.
- Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik Learned-Miller. Multi-view convolutional neural networks for 3d shape recognition. In *Proceedings of the IEEE international conference on computer vision*, pages 945–953, 2015.
- Xinwei He, Yang Zhou, Zhichao Zhou, Song Bai, and Xiang Bai. Triplet-center loss for multi-view 3d object retrieval. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1945–1954, 2018.
- Haibin Huang, Evangelos Kalogerakis, Siddhartha Chaudhuri, Duygu Ceylan, Vladimir G Kim, and Ersin Yumer. Learning local shape descriptors from part correspondences with

- multiview convolutional networks. *ACM Transactions on Graphics (TOG)*, 37(1):1–14, 2017.
- Zhaoliang Lun, Matheus Gadelha, Evangelos Kalogerakis, Subhransu Maji, and Rui Wang. 3d shape reconstruction from sketches via multi-view convolutional networks. In *2017 International Conference on 3D Vision (3DV)*, pages 67–77. IEEE, 2017.
- Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- Fang Wang, Le Kang, and Yi Li. Sketch-based 3d shape retrieval using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1875–1883, 2015.
- Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015.
- Chao-Yuan Wu, R Manmatha, Alexander J Smola, and Philipp Krahenbuhl. Sampling matters in deep embedding learning. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2840–2848, 2017.
- Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- Mathias Eitz, Ronald Richter, Tamy Boubekeur, Kristian Hildebrand, and Marc Alexa. Sketch-based shape retrieval. *ACM Trans. Graph.*, 31(4):31, 2012.
- David Ha and Douglas Eck. A neural representation of sketch drawings. *arXiv preprint arXiv:1704.03477*, 2017.
- Johanna Delanoy, Mathieu Aubry, Phillip Isola, Alexei A Efros, and Adrien Bousseau. 3D Sketching using Multi-View Deep Volumetric Prediction. *arXiv preprint arXiv:1707.08390*, 2017.

- Haibin Huang, Evangelos Kalogerakis, Ersin Yumer, and Radomir Mech. Shape synthesis from sketches via procedural models and convolutional networks. *IEEE Transactions on Visualization and Computer Graphics*, 2, 2016.
- Wanchao Su, Dong Du, Xin Yang, Shizhe Zhou, and Hongbo Fu. Interactive sketch-based normal map generation with deep neural networks. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 1(1):22, 2018.
- Takafumi Saito and Tokiichiro Takahashi. Comprehensible rendering of 3-D shapes. In *ACM SIGGRAPH Computer Graphics*, volume 24, pages 197–206. ACM, 1990.
- Forrester Cole, Aleksey Golovinskiy, Alex Limpaecher, Heather Stoddart Barros, Adam Finkelstein, Thomas Funkhouser, and Szymon Rusinkiewicz. Where do people draw lines? *Communications of the ACM*, 55(1):107–115, 2012.
- K Eissen and R Steur. Sketching: the basics (ed. 2012) Amsterdam. *Google Scholar*, 2011.
- Frederic Cordier, Karan Singh, Even Etem, Marie-Paule Cani, and Yotam Gingold. Sketch-based modeling. In *Proceedings of the 37th Annual Conference of the European Association for Computer Graphics: Tutorials*, page 7. Eurographics Association, 2016.
- Doug DeCarlo, Adam Finkelstein, Szymon Rusinkiewicz, and Anthony Santella. Suggestive contours for conveying shape. *ACM Transactions on Graphics (TOG)*, 22(3):848–855, 2003.
- Srikrishna Karanam, Yang Li, and Richard J Radke. Person re-identification with discriminatively trained viewpoint invariant dictionaries. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4516–4524, 2015.
- David Kenneth Wagg and Mark S Nixon. On automated model-based extraction and analysis of gait. In *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, pages 11–16. IEEE, 2004.
- Serge Belongie, Jitendra Malik, and Jan Puzicha. Shape matching and object recognition using shape contexts. Technical report, California Univ. San Diego La Jolla Dept. of Computer Science and Engineering, 2002.

- Jianbo Shi and Carlo Tomasi. Good features to track. Technical report, Cornell University, 1993.
- Kfir Aberman, Jing Liao, Mingyi Shi, Dani Lischinski, Baoquan Chen, and Daniel Cohen-Or. Neural Best-Buddies: Sparse Cross-Domain Correspondence. *arXiv preprint arXiv:1805.04140*, 2018.
- Edgar Simo-Serra, Satoshi Iizuka, and Hiroshi Ishikawa. Mastering sketching: adversarial augmentation for structured prediction. *ACM Transactions on Graphics (TOG)*, 37(1): 11, 2018.
- Bruno Alfredo Pazos, José Pablo Navarro, Arturo Leonardo Morales, Celia Cintas, Alexandra Trujillo, Soledad de Azevedo, Rolando González-José, and Claudio Delrieux. Detección automática de tejido blando nasal en ct-scan y mri utilizando random forests. In *IX Congreso Argentino de Informática y Salud (CAIS)-JAIIO 47 (CABA, 2018)*, 2018.
- Celia Cintas, Claudio Delrieux, Pablo Navarro, Mirsha Quinto-Sanchez, Bruno Pazos, and Rolando Gonzalez-Jose. Automatic ear detection and segmentation over partially occluded profile face images. *Journal of Computer Science & Technology*, 19, 2019.
- Jian Zhou and Olga G Troyanskaya. Predicting effects of noncoding variants with deep learning-based sequence model. *Nature methods*, 12(10):931–934, 2015.
- Wenlong Ma, Zhixu Qiu, Jie Song, Jiajia Li, Qian Cheng, Jingjing Zhai, and Chuang Ma. A deep convolutional neural network approach for predicting phenotypes from genotypes. *Planta*, 248(5):1307–1318, 2018.
- Zhengping Che, Yu Cheng, Shuangfei Zhai, Zhaonan Sun, and Yan Liu. Boosting deep learning risk prediction with generative adversarial networks for electronic health records. In *2017 IEEE International Conference on Data Mining (ICDM)*, pages 787–792. IEEE, 2017.
- 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, et al. Facial asymmetry and genetic ancestry in Latin American admixed populations. *American journal of physical anthropology*, 157(1):58–70, 2015.

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, et al. Socioeconomic status is not related with facial fluctuating asymmetry: evidence from latin-american populations. *PLoS one*, 12(1): e0169287, 2017.

Kaustubh Adhikari, Javier Mendoza-Revilla, Anood Sohail, Macarena Fuentes-Guajardo, Jodie Lampert, Juan Camilo Chacón-Duque, Malena Hurtado, Valeria Villegas, Vanessa Granja, Victor Acuña-Alonzo, et al. A gwas in latin americans highlights the convergent evolution of lighter skin pigmentation in eurasia. *Nature communications*, 10(1):1–16, 2019.

Jerry Wei, Arief Suriawinata, Louis Vaickus, Bing Ren, Xiaoying Liu, Jason Wei, and Saeed Hassanpour. Generative image translation for data augmentation in colorectal histopathology images. *arXiv preprint arXiv:1910.05827*, 2019.