



*ugr* | Universidad  
de **Granada**

TRABAJO FIN DE GRADO  
INGENIERÍA INFORMÁTICA

**Estimación de la distancia cámara-sujeto  
en fotografías faciales mediante técnicas  
de aprendizaje profundo**

---

**Autor**  
Iván Salinas López

**Directores**  
Enrique Bermejo Nievas  
Pablo Mesejo Santiago



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍAS INFORMÁTICA Y DE  
TELECOMUNICACIÓN

—  
Granada, Junio de 2024



# **Estimación de distancia cámara-sujeto en fotografías faciales usando aprendizaje profundo**

Iván Salinas López

**Palabras clave:** palabra\_clave1, palabra\_clave2, palabra\_clave3, .....

## **Resumen**

Poner aquí el resumen.



# **Camera-subject distance estimation in facial photographs using deep learning**

Iván Salinas López

**Keywords:** Keyword1, Keyword2, Keyword3, ....

## **Abstract**

Write here the abstract in English.



---

Yo, **Iván Salinas López**, alumno de la titulación Grado en Ingeniería Informática de la **Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación de la Universidad de Granada**, con DNI 78026145W, autorizo la ubicación de la siguiente copia de mi Trabajo Fin de Grado en la biblioteca del centro para que pueda ser consultada por las personas que lo deseen.

Fdo: Iván Salinas López

Granada a X de Junio de 2024



---

D. **Enrique Bermejo Nievas**, Investigador Senior en Panacea Cooperative Research y miembro del Instituto Andaluz Interuniversitario en Ciencia de Datos e Inteligencia Computacional.

D. **Pablo Mesejo Santiago**, Profesor del Área de Ciencias de la Computación e Inteligencia Artificial del Departamento Ciencias de la Computación e Inteligencia Artificial de la Universidad de Granada.

**Informan:**

Que el presente trabajo, titulado *Estimación de la distancia cámara-sujeto en fotografías faciales mediante técnicas de aprendizaje profundo*, ha sido realizado bajo su supervisión por **Iván Salinas López**, y autorizamos la defensa de dicho trabajo ante el tribunal que corresponda.

Y para que conste, expiden y firman el presente informe en Granada a X de Junio de 2024.

**Los directores:**

**Enrique Bermejo Nievas**

**Pablo Mesejo Santiago**



# Agradecimientos

Poner aquí agradecimientos...



# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. Definición del problema . . . . .	1
1.2. Motivación . . . . .	3
1.3. Objetivos . . . . .	4
1.4. Planificación del proyecto . . . . .	5
<b>2. Fundamentos teóricos</b>	<b>7</b>
2.1. Aprendizaje automático . . . . .	7
2.2. Aprendizaje profundo . . . . .	8
2.2.1. Redes neuronales . . . . .	8
2.2.2. Redes neuronales convolucionales . . . . .	10
2.2.3. Transferencia de aprendizaje . . . . .	13
2.3. Parámetros de la cámara y perspectiva . . . . .	13



# Índice de figuras

1.1. Efectos de la distorsión de perspectiva en características faciales de fotografías realizadas a diferentes SCD: 0.5 m, 1 m y 3 m. Estos efectos varían en relación a la distancia y son independientes de la longitud focal [3] . . . . .	2
1.2. Ejemplo de superposición craneofacial [29]. . . . .	3
1.3. A la izquierda la fotografía distorsionada, a la derecha la fotografía sin distorsión [39]. . . . .	4
2.1. Esquema de una red neuronal [20]. . . . .	8
2.2. Modelo neuronal para una neurona $k$ [16]. . . . .	9
2.3. Ejemplo de CNN [30]. . . . .	10
2.4. Ejemplo de convolución en CNN [37]. . . . .	11
2.5. Funciones de activación comúnmente aplicadas en CNN [37].	11
2.6. Tipos de pooling comúnmente utilizados en CNN [37]. . . .	12
2.7. Relación entre punto nodal y longitud focal [15] . . . .	13
2.8. Tamaños del sensor expresados según el factor de recorte . .	14
2.9. Ejemplo de longitud focal equivalente según el tamaño del sensor [32]. . . . .	14
2.10. Distancia desde la cámara al sujeto. . . . .	15



# Índice de cuadros

1.1. Planificación inicial del proyecto . . . . .	6
---	---



# Capítulo 1

## Introducción

### 1.1. Definición del problema

Durante la última década, la identificación facial ha ganado una relevancia significativa gracias a la revolución del aprendizaje profundo y los sistemas automáticos de reconocimiento facial. Este avance ha conducido a una expansión de las posibles aplicaciones en el mercado actual desde los campos del cumplimiento de la ley o la ciencia forense, hasta áreas del sector privado como el comercio minorista, las aplicaciones multimedia o la seguridad. Además, el desarrollo tecnológico en el ámbito de la imagen ha mejorado tanto la calidad como la disponibilidad de datos fotográficos, lo cual facilita un análisis más exhaustivo de los factores que influyen en las imágenes fotográficas.

Las técnicas de identificación facial normalmente son realizadas por expertos con o sin la ayuda de sistemas automáticos. Los expertos analizan los datos y evalúan las características anatómicas de un individuo desconocido para compararlas con las de uno o varios individuos conocidos. Actualmente existen cuatro métodos de comparación facial reconocidos: análisis morfológico, superposición, foto-antropometría y comparación holística [11].

Para que este análisis sea confiable y concluyente, los datos (fotografías faciales), deben estar en unas condiciones adecuadas (calidad, resolución, enfoque o iluminación) y la escena (punto de vista de la cámara, pose de la cabeza, expresión facial) debe ser lo más neutral y representativa posible [10, 31]. Estos requisitos nos aseguran que los rasgos faciales sean lo más fieles posible a las características anatómicas del individuo, y por tanto, permiten que las técnicas de identificación sean más robustas.

Muchos estudios han identificado limitaciones en los actuales métodos de reconocimiento facial automático. Los principales factores desafiantes son la pose, la iluminación, la expresión y la variación en la edad [19, 23]. Sin

embargo también existen otros factores importantes como la oclusión, el género o la etnia [7, 13].

Uno de los factores más importantes en las fotografías faciales es la distorsión de perspectiva, la cual puede provocar deformaciones en los rasgos faciales, como en las orejas, la nariz o la forma general del rostro, especialmente cuando la cámara está muy cerca del sujeto al momento de tomar la fotografía [26] (ver Figura 1.1). Esta alteración en la perspectiva repercute negativamente en los sistemas de reconocimiento automático [9, 28, 33], lo cual puede obstaculizar la identificación precisa de individuos.



Figura 1.1: Efectos de la distorsión de perspectiva en características faciales de fotografías realizadas a diferentes SCD: 0.5 m, 1 m y 3 m. Estos efectos varían en relación a la distancia y son independientes de la longitud focal [3]

La distorsión de perspectiva está estrechamente relacionada con la distancia cámara-sujeto (subject-to-camera distance, SCD en adelante), de hecho, la relación es de decremento logarítmico, es decir, valores pequeños de la SCD corresponden con una mayor distorsión, mientras que la distorsión disminuye conforme la SCD aumenta [29]. Conocer la SCD en fotografías faciales permite cuantificar la cantidad de distorsión presente en la imagen, así como las diferencias en la distorsión entre dos conjuntos de imágenes. Esta información es útil para reproducir con fiabilidad las condiciones originales de una escena, especialmente en el ámbito forense, y para facilitar el desarrollo de técnicas que permitan la corrección precisa de dicha distorsión

La SCD, a diferencia de otros parámetros de la cámara como la longitud focal o el tamaño del sensor, no puede obtenerse directamente desde los metadatos de la fotografía [34]. Por tanto, se necesita un método preciso para su estimación.

En los últimos años se han utilizado varios métodos que combinan técnicas manuales y automatizadas basadas en puntos de referencia o en características anatómicas de la cara[12, 8]. Sin embargo, no se han obtenido resultados favorables debido a la dificultad para obtener estimaciones precisas en largas distancias por la diversa fisionomía de la cara, y a los problemas

relacionados con los parámetros de la cámara, como el recorte de imágenes o la combinación de diferentes longitudes focales en el mismo conjunto de datos.

El único método totalmente automatizado para estimar la SCD en fotografías faciales, hasta la fecha, se conoce como FacialSCDnet [3]. Este método utiliza una arquitectura basada en aprendizaje profundo para procesar fotografías faciales y estimar la SCD con precisión. Sin embargo, las soluciones basadas en aprendizaje profundo tienen sesgos sobre los datos y las muestras deben ser lo más representativas posible.

Considerando todos estos aspectos, el presente Trabajo de Fin de Grado (TFG) consiste en mejorar el método actual del estado del arte en la estimación automática de la distancia cámara-sujeto en fotografías faciales mediante el uso de técnicas de aprendizaje profundo.

## 1.2. Motivación

En el ámbito forense, el principal foco recae sobre la determinación de la identidad humana cuando se dispone de información esquelética [5]. En las últimas décadas, los antropólogos han centrado su atención en mejorar las técnicas para realizar una identificación más precisa. En este contexto, la estimación de la SCD juega un papel crucial, ya que, si estimamos la SCD con fiabilidad, podemos recrear la imagen con los restos esqueléticos (aplicando el valor de la SCD) [6]. A continuación, realizamos las comparaciones anatómicas mediante la superposición craneofacial [1] para identificar si se corresponde con la misma persona (ver Figura 1.2). Si los parámetros de adquisición entre ambas imágenes (normal y esquelética) son distintos entonces dificultaría el análisis morfológico [17].

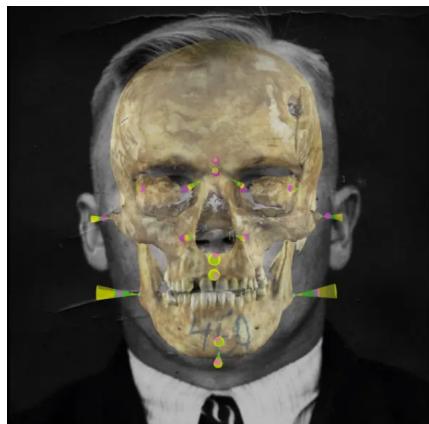


Figura 1.2: Ejemplo de superposición craneofacial [29].

En el ámbito de la biometría y el reconocimiento facial, conocer la SCD es fundamental para comprender y corregir las distorsiones de perspectiva que pueden afectar a la precisión de la identificación. Esta distancia influye directamente en la representación facial capturada, afectando la geometría y proporciones del rostro (ver Figura 1.3). Un entendimiento preciso de esta medida permite aplicar técnicas de corrección de distorsión [36, 39] para garantizar una representación fiel y coherente de las características faciales, mejorando de esta manera la precisión y fiabilidad de los sistemas biométricos y de reconocimiento facial.



Figura 1.3: A la izquierda la fotografía distorsionada, a la derecha la fotografía sin distorsión [39].

En el campo del aprendizaje profundo, es común encontrar sesgos [35] en los datos que pueden afectar a la precisión y generalización de las soluciones. Dichos sesgos pueden surgir de diversos factores, como la falta de diversidad o el ruido excesivo en el conjunto de datos utilizado para entrenar los modelos. Esto también ocurre en los métodos actuales para la estimación automática de la SCD. Es por ello que, una forma de mejorar el conjunto de datos podría ser incluir una variedad más amplia de grupos étnicos, edades, géneros, expresiones faciales, condiciones de iluminación y fondos, así como modelos tanto faciales como de cuerpo completo. Al hacerlo, se busca mitigar los sesgos inherentes a los datos y desarrollar modelos más robustos y equitativos. Esta mejora implica una mayor capacidad de generalización y adaptación a una amplia gama de situaciones y contextos.

Por todo ello, la estimación de la SCD de forma precisa y generalizable tiene el potencial de impulsar avances significativos en diversos campos de investigación.

### **1.3. Objetivos**

El objetivo general de este TFG consiste en desarrollar un modelo de aprendizaje profundo adecuado para mejorar la estimación de la distancia cámara-sujeto en fotografías faciales. Para el desarrollo del proyecto, dividi-

remos el objetivo general en una serie de objetivos parciales:

1. Realizar un análisis exhaustivo del estado del arte para la estimación de la SCD en fotografías faciales.
2. Generar un conjunto de datos sintético de mayor calidad (utilizando modelos 3D más completos de cuerpo entero y poses distintas, con fondos e iluminación más realistas).
3. Realizar un estudio experimental que permita validar los enfoques propuestos y extraer conclusiones sobre su aplicabilidad al problema.
4. Desarrollar y entrenar con nuevas arquitecturas que mejoren los resultados.
5. Usar tecnologías más recientes que mejoren los tiempos de aprendizaje y los resultados obtenidos.

#### 1.4. Planificación del proyecto

Para abordar el desarrollo de este proyecto, es esencial considerar que el TFG tiene asignados 12 créditos ECTS, lo que equivale a aproximadamente 300 horas de trabajo. Dada la distribución temporal del segundo cuatrimestre, con unas 20 semanas disponibles, se estima que se requerirá dedicar al TFG unas 20 horas semanales, equivalentes a 4 horas diarias durante 5 días a la semana. Se reservan así 4 semanas como margen para posibles retrasos o imprevistos que puedan surgir durante el desarrollo del proyecto.

En cuanto a la metodología de desarrollo, se ha optado por seguir un enfoque basado en el ciclo de vida en cascada [27], aunque con una variante que permite retroalimentación. Aunque el proyecto presenta requisitos y objetivos claros, se reconoce la posibilidad de ajustes menores durante su desarrollo, especialmente a medida que se obtenga más información sobre el problema y los métodos. Esta flexibilidad se considera crucial para adaptarse a posibles cambios en el contexto o los requisitos del proyecto.

Las fases del ciclo de vida del proyecto son las siguientes:

A continuación se describen las fases del ciclo de vida del proyecto:

- Análisis de Requisitos: Consiste en las reuniones iniciales con los clientes, en este caso los directores del TFG. Se realiza un análisis del problema y un estudio detallado de la bibliografía existente
- Diseño: Consiste en la exploración y selección de los métodos apropiados basados en el análisis previo, tanto para la resolución como para la

validación de la solución propuesta. Además, se llevarán a cabo pruebas preliminares y se elaborará el diseño del software experimental.

- Implementación: Consiste en la adaptación del código de los modelos investigados, la implementación de nuevas funcionalidades y la generación de un conjunto de datos sintético junto con su posterior preprocesado.
- Pruebas: Consiste en la realización de diversos experimentos para validar el funcionamiento del software desarrollado, utilizando los modelos y datos previamente definidos.

Tarea	Semanas	Febrero				Marzo				Abril				Mayo			Junio				
	- Horas	5	12	19	26	4	11	18	25	1	8	15	22	29	6	13	20	27	3	10	17
Análisis de Requisitos	3 - 60																				
Diseño	3 - 60																				
Implementación	5 - 90																				
Pruebas	5 - 90																				

Tabla 1.1: Planificación inicial del proyecto

## Capítulo 2

# Fundamentos teóricos

### 2.1. Aprendizaje automático

El aprendizaje automático (Machine Learning, ML) [18, 6] es una rama de la IA y de las ciencias de la computación centrada en el uso de datos y algoritmos para imitar la forma en la que los humanos aprenden, detectando patrones o regularidades para realizar predicciones.

Existen 3 tipos de aprendizaje dentro del ML [1, 29]:

El **aprendizaje supervisado** consiste en entrenar con datos de los que se saben sus etiquetas (una etiqueta nos indica qué es cada dato). Por ejemplo, los datos de entrada podrían ser imágenes de animales y sus etiquetas podrían ser 'perro' o 'gato'. A partir de los datos y sus etiquetas, el modelo aprende una función que dado un nuevo dato, predice su etiqueta. Es el tipo de aprendizaje más utilizado, los datos vienen ya 'preparados' para su uso. Es el tipo de aprendizaje que utilizaremos en este TFG.

En el **aprendizaje no supervisado** el agente aprende los patrones de los datos de entrada sin ninguna realimentación, es decir, los datos no están etiquetados. La herramienta más usada en el aprendizaje no supervisado es el agrupamiento, que consiste en detectar potenciales grupos en los datos de entrada. Este enfoque requiere un mayor número de datos.

En el **aprendizaje por refuerzo** el agente aprende mediante una serie de recompensas o castigos. El agente intentará realizar acciones que le proporcionen mejores recompensas en el futuro. Este tipo de aprendizaje es muy utilizado para enseñar a jugar a juegos.

## 2.2. Aprendizaje profundo

### 2.2.1. Redes neuronales

Las redes neuronales (ANNs) [14, 16, 4] son redes computacionales que intentan, a groso modo, simular el proceso de decisión de las neuronas del sistema nervioso central de los animales o humanos. Las ANNs poseen unidades de procesamiento de información llamadas neuronas, que están conectadas entre sí mediante capas. La red se compone de (ver Figura 2.1):

- Una capa de entrada, que tendrá tantos inputs como características o variables tenga el problema
- Una o varias capas ocultas, compuestas por neuronas. El número de capas ocultas define la profundidad de la red neuronal.
- Una capa de salida, que representa el valor o valores predichos

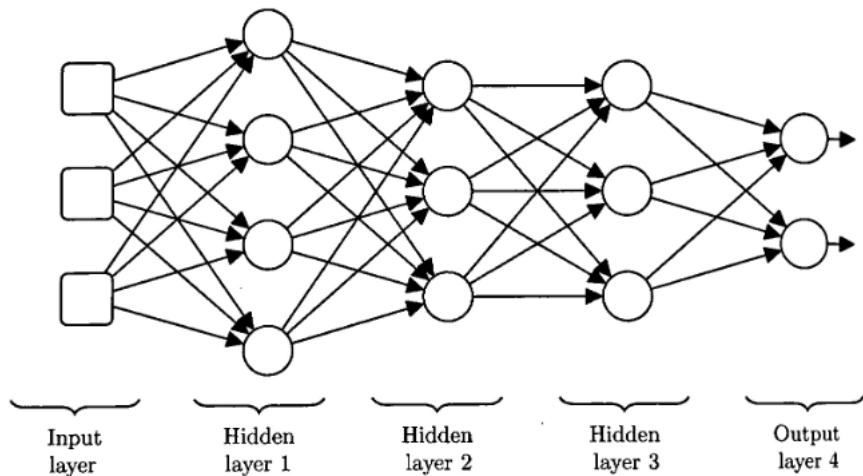


Figura 2.1: Esquema de una red neuronal [20].

Las neuronas son la unidad fundamental de cómputo, tienen varios valores de entrada y un valor de salida que se conecta con las neuronas de la siguiente capa. Los elementos básicos del modelo neuronal son (ver Figura 2.2):

- Un conjunto de conexiones con las señales de entrada. Cada conexión tiene su propio peso/fuerza.
- Una función de suma de las señales de entrada, ponderadas cada una con su peso. Estas operaciones constituyen una combinación lineal.

- Una función de activación, para limitar la amplitud de la salida de la neurona. Normalmente, el rango de salida está en el intervalo [0,1], o alternativamente en [-1,1]. Existen muchos tipos de funciones de activación pero, se suelen utilizar cuatro: la función signo, la función logística, la función arco-tangente y la función ReLU.

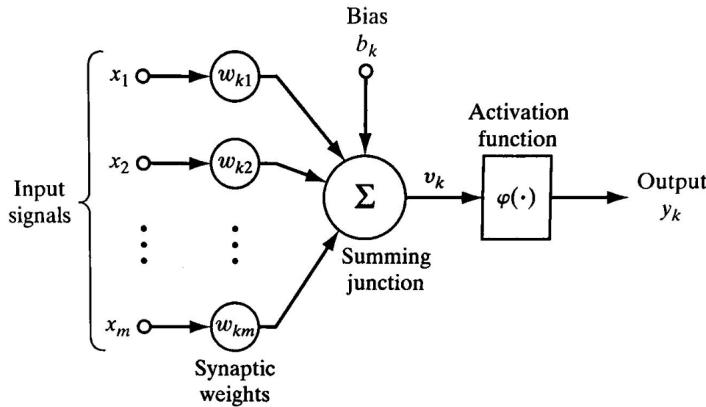


Figura 2.2: Modelo neuronal para una neurona  $k$  [16].

En términos matemáticos, podemos describir la salida de una neurona como:

$$y = \phi\left(\sum_{j=1}^m w_j x_j + b\right) \quad (2.1)$$

siendo  $\phi$  la función de activación,  $m$  el número de señales de entrada,  $w_j$  el peso de cada entrada  $x_j$ , y  $b$  el sesgo.

El algoritmo de aprendizaje de la red neuronal consiste en ir modificando los pesos y el sesgo, iterativamente, hasta alcanzar el resultado deseado. Este proceso iterativo se conoce como entrenamiento, y permite, a través de las modificaciones de los pesos, reconocer y extraer las características más relevantes de los datos.

El objetivo del entrenamiento es minimizar el error de predicción de la salida de la red neuronal, para ello, se define una función de pérdida. Existen numerosas funciones de pérdida, algunas de las más conocidas son: el error cuadrático medio (MSE), el error absoluto medio (MAE) o la entropía cruzada. La información de la función de pérdida se transmite desde la salida a la capa inicial, con el fin de adecuadamente los pesos para generar una mejor estimación de la predicción.

El sobreentrenamiento es un factor importante a evitar. Sobreentrenar el modelo de aprendizaje, significa, ajustarlo demasiado al conjunto de da-

tos de entrenamiento, de manera que, al recibir nuevos datos no utilizados para entrenar, se estime un mal resultado debido a la poca capacidad de generalización ante nuevos datos.

### 2.2.2. Redes neuronales convolucionales

Las Redes Neuronales Convolucionales (Convolutional Neural Network, CNN) [21, 22, 38] son un tipo de red neuronal profunda que trabaja con patrones de cuadrícula, como pueden ser imágenes (ver Figura 2.3).

Este tipo de redes neuronales incluyen dos tipos de capas adicionales: capas convolucionales y capas de pooling. Estas capas están en la primera parte de la red y son las encargadas de extraer las características relevantes de la entrada. Esto nos permite automatizar el proceso de extracción de características, a la vez que mejorarlo tanto en tiempo como en rendimiento.

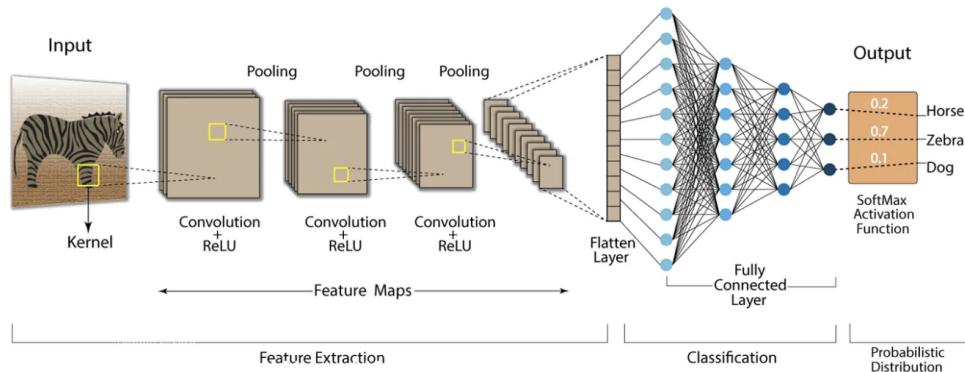


Figura 2.3: Ejemplo de CNN [30].

A continuación, se describen los posibles tipos de capas presentes en una CNN [35, 37]:

#### Capa de convolución

La capa convolucional es un componente fundamental de las CNN que se utiliza para extraer características de una imagen o conjunto de datos. Funciona aplicando operaciones de convolución a través de un conjunto de filtros o kernels sobre la entrada para detectar patrones visuales o características específicas en los datos (ver Figura 2.4).

Estos kernels pueden tener distintos tamaños y sus valores se aprenden a lo largo del entrenamiento de la red. Sin embargo, hay que tener en cuenta que la operación de convolución es lineal y no permite aprender patrones complejos. En este contexto es donde entra la capa de activación.

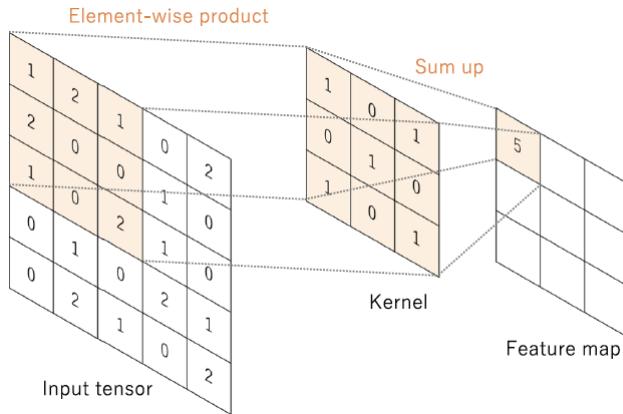


Figura 2.4: Ejemplo de convolución en CNN [37].

### Capa de activación

La capa de activación en una CNN sigue a la capa convolucional y se encarga de introducir no linealidades en el modelo mediante una función de activación. Esta función aumenta la capacidad de la red para aprender relaciones no lineales en los datos, lo que es fundamental para capturar patrones más complejos. Algunas de las funciones de activación comunes utilizadas son la función ReLU (Rectified Linear Unit), la función sigmoide y la función tangente hiperbólica (ver Figura 2.5).

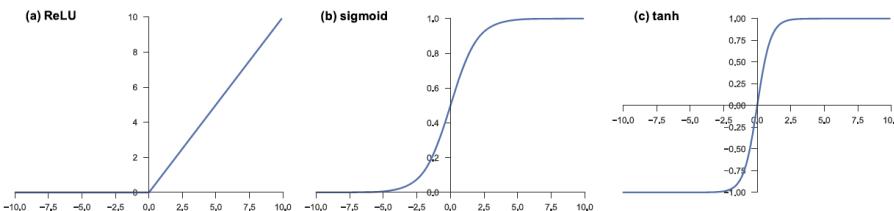


Figura 2.5: Funciones de activación comúnmente aplicadas en CNN [37].

### Capa de pooling

La capa de pooling también es específica de las CNN y se encarga de reducir la dimensionalidad de las características conservando la información más relevante.

Esta capa resume la información en regiones locales mediante una operación de downsampling en las características de entrada. Al reducir la dimensionalidad de las características, la capa de pooling disminuye el número de parámetros aprendibles en la red, lo que puede ayudar a prevenir el sobre-

ajuste y mejorar la eficiencia computacional del modelo. Además, la capa de pooling también ayuda a introducir invariancia a pequeñas traslaciones y distorsiones en los datos de entrada, lo que permite a la red reconocer patrones incluso si están ligeramente desplazados en la imagen.

Los dos tipos más comunes de pooling son el max pooling, que selecciona el valor máximo de una región local en las características de entrada, y el average pooling, que calcula el promedio de los valores en una región local (ver Figura 2.6).

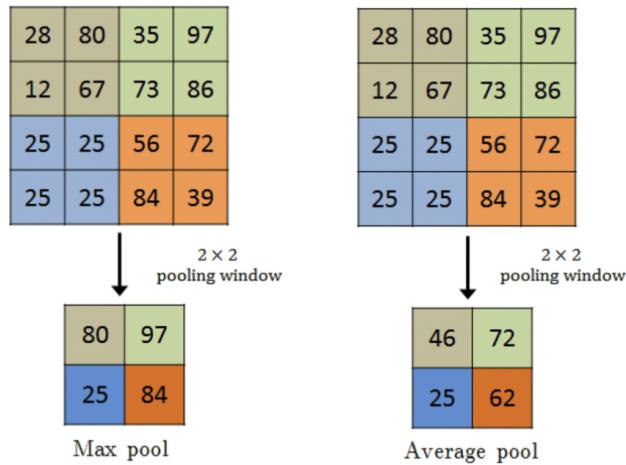


Figura 2.6: Tipos de pooling comúnmente utilizados en CNN [37].

Los hiperparámetros de la capa de pooling incluyen el tamaño del filtro, el stride y el tipo de padding. Estos hiperparámetros afectan la forma en que se realiza el downsampling en las características.

### Capa totalmente conectada

La capa totalmente conectada sigue a las capas de convolución y pooling. En esta capa, las características extraídas por las capas anteriores se transforman en un formato unidimensional (vector) antes de conectarse a una o más capas totalmente conectadas, también conocidas como capas densas.

Esta capa, al igual que en las redes neuronales clásicas, tiene la responsabilidad de combinar y procesar las características extraídas para producir la salida final de la red.

Cada neurona en una capa totalmente conectada está conectada a todas las neuronas de la capa anterior a través de pesos aprendibles. Estos pesos determinan la contribución de cada neurona de entrada a la neurona de salida correspondiente en la capa totalmente conectada. Durante el entrenamiento, estos pesos se ajustan mediante algoritmos de optimización como

backpropagation y descenso de gradiente para minimizar la diferencia entre las salidas predichas y las etiquetas reales.

Es importante destacar que la capa totalmente conectada suele estar seguida por una función de activación no lineal, como ReLU, para introducir no linealidades en el modelo y permitir la representación de patrones complejos en los datos. Además, la última capa de activación de la CNN, generalmente se selecciona según la naturaleza de la tarea que se está abordando.

### 2.2.3. Transferencia de aprendizaje

## 2.3. Parámetros de la cámara y perspectiva

### Longitud focal

La longitud focal [15] mide la distancia, en milímetros, entre el 'punto nodal' (punto donde la luz converge en una lente) y el sensor de la cámara (ver Figura 2.7).

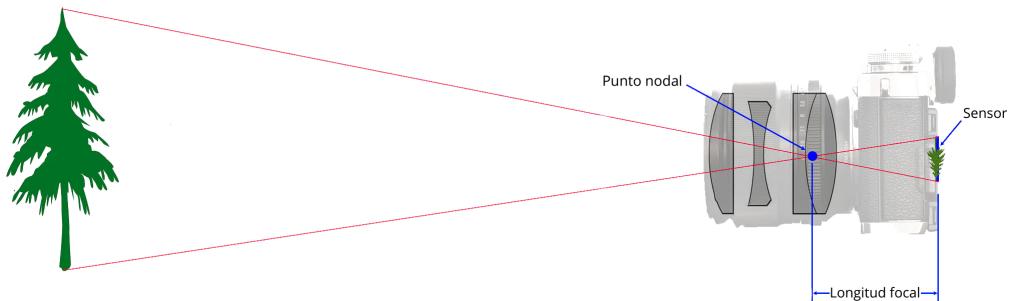


Figura 2.7: Relación entre punto nodal y longitud focal [15]

La longitud focal es importante porque se relaciona con el campo de visión de una lente, es decir, cuánta escena capturarás. También explica qué tan grande o pequeño aparecerá un sujeto en la fotografía.

Valores más altos de longitud focal (como 500 mm) se ven más 'cerca', mientras que valores más bajos (como 20 mm) están más 'lejos'.

### Sensor de la cámara

El sensor de la cámara [2, 24] es el componente que captura la luz y la convierte en una imagen digital. El tamaño del sensor afecta a la calidad de la imagen y a la cantidad de luz que puede capturar. El tamaño estándar es de 36mm x 24mm, también conocido como Full frame o 35mm.

Otra forma equivalente de referirnos al tamaño del sensor es mediante el factor de recorte, que se define como la relación existente entre el tamaño de un sensor de 35mm y el sensor de nuestra cámara (ver Figura 2.8).



Figura 2.8: Tamaños del sensor expresados según el factor de recorte

Uno de los aspectos más importantes de utilizar un sensor con factor de recorte distinto a 1, es que la longitud focal se ve afectada, de forma que hablamos de longitud focal equivalente.

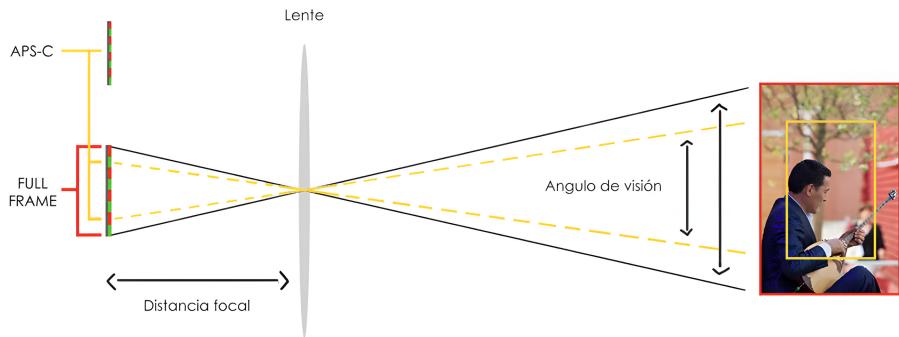


Figura 2.9: Ejemplo de longitud focal equivalente según el tamaño del sensor [32].

### Distancia cámara-sujeto

La distancia cámara-sujeto se define como la separación física entre la cámara y el sujeto que está siendo fotografiado (ver Figura 2.10). Alterar esta distancia provoca variaciones en la apariencia visual del rostro en la fotografía obtenida [25]. Este fenómeno se conoce como distorsión de perspectiva.



Figura 2.10: Distancia desde la cámara al sujeto.

En este TFG se utilizarán los ojos del sujeto como punto de referencia para medir la SCD, esto se debe a su importancia y su posición central en el rostro.

### Distorsión de perspectiva



# Bibliografía

- [1] Yaser S. Abu-Mostafa, M. Magdon-Ismail y H.T. Lin. *Learning from Data: A Short Course*. AMLBook, 2012.
- [2] Adorama. *What Is Crop Factor And How Do You Calculate It?* Accedido el 17 de Marzo de 2024. 2022. URL: <https://www.adorama.com/alc/what-is-crop-factor-everything-you-need-to-know/>.
- [3] Enrique Bermejo et al. «FacialSCDnet: A deep learning approach for the estimation of subject-to-camera distance in facial photographs». En: *Expert Systems with Applications* 210 (2022), pág. 118457. ISSN: 0957-4174.
- [4] C.M. Bishop. *Neural networks for pattern recognition*. Oxford University Press, 1995.
- [5] Soren Blau y Luis Fondebrider. «Human Skeletal Remains: Identification of Individuals». En: *Encyclopedia of Global Archaeology*. 2014, págs. 3556-3558.
- [6] Sara Brown. *Machine learning, explained*. Accedido el 11 de Marzo de 2024. 2021. URL: <https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained>.
- [7] Joy Buolamwini y Timnit Gebru. «Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification». En: *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*. Vol. 81. 2018, págs. 77-91.
- [8] Xavier P. Burgos-Artizzu, Matteo Ruggero Ronchi y Pietro Perona. «Distance Estimation of an Unknown Person from a Portrait». En: *Computer Vision – ECCV 2014*. Vol. 8689. 2014, págs. 313-327.
- [9] Naser Damer et al. «Deep Learning-based Face Recognition and the Robustness to Perspective Distortion». En: *24th International Conference on Pattern Recognition*. 2018, págs. 3445-3450.
- [10] Gary Edmond et al. «Law's Looking Glass: Expert Identification Evidence Derived from Photographic and Video Images». En: *Current Issues in Criminal Justice* 20 (2009), págs. 337-377.

- [11] FISWG. *Facial Comparison Overview and Methodology Guidelines V2.0.* Accedido el 7 de Febrero de 2024. 2022. URL: [https://fiswg.org/fiswg\\_facial\\_comparison\\_overview\\_and\\_methodology\\_guidelines\\_V2.0\\_2022.11.04.pdf](https://fiswg.org/fiswg_facial_comparison_overview_and_methodology_guidelines_V2.0_2022.11.04.pdf).
- [12] Arturo Flores et al. «Camera Distance from Face Images». En: *Advances in Visual Computing*. Vol. 8034. 2013, págs. 513-522.
- [13] Nicholas Furl, P.Jonathon Phillips y Alice J O'Toole. «Face recognition algorithms and the other-race effect: computational mechanisms for a developmental contact hypothesis». En: *Cognitive Science* 26 (2002), págs. 797-815. ISSN: 0364-0213.
- [14] Daniel Graupe. *Principles of artificial neural networks 3rd edition.* World Scientific, 2007.
- [15] Elizabeth Gray. *What Is Focal Length in Photography? A Beginner's Guide.* Accedido el 17 de Marzo de 2024. 2023. URL: <https://photographylife.com/what-is-focal-length-in-photography>.
- [16] Simon Haykin. *Neural Networks and Learning Machines 3rd edition.* Pearson, 2009.
- [17] Sean Healy y Carl Stephan. «Perspective distortion tolerances and skull-face registration in craniofacial superimposition: an analytical review». En: *International Journal of Legal Medicine* 137 (2023).
- [18] IBM. *What is machine learning?* Accedido el 11 de Marzo de 2024. 2019. URL: <https://www.ibm.com/topics/machine-learning>.
- [19] Anil K. Jain, Brendan Klare y Unsang Park. «Face Matching and Retrieval in Forensics Applications». En: *IEEE MultiMedia* 19 (2012), págs. 20-20.
- [20] John D. Kelleher. *Deep learning.* MIT Press, 2019.
- [21] Y. LeCun et al. «Backpropagation Applied to Handwritten Zip Code Recognition». En: *Neural Computation* 1 (1989), págs. 541-551.
- [22] Y. Lecun et al. «Gradient-based learning applied to document recognition». En: *Proceedings of the IEEE* 86 (1998), págs. 2278-2324.
- [23] Zhifeng Li, Unsang Park y Anil K. Jain. «A Discriminative Model for Age Invariant Face Recognition». En: *IEEE Transactions on Information Forensics and Security* 6 (2011), págs. 1028-1037.
- [24] Javier Lucas. *Qué Es El Factor de Recorte de tu Sensor y Cómo Influye en la Focal de tus Objetivos.* Accedido el 17 de Marzo de 2024. 2022. URL: <https://www.dzoom.org/que-es-el-factor-de-recorte-de-tu-sensor-y-como-influye-en-la-focal-de-tus-objetivos/>.

- [25] Eilidh Noyes y Rob Jenkins. «Camera-to-subject distance affects face configuration and perceived identity». En: *Cognition* 165 (2017), págs. 97-104.
- [26] Bo Peng et al. «Position Determines Perspective: Investigating Perspective Distortion for Image Forensics of Faces». En: *IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2017, págs. 1813-1821.
- [27] R.S. Pressman. *Software Engineering: A Practitioner's Approach*. McGraw-Hill, 2005.
- [28] Zahid Riaz. y Michael Beetz. «On the effect of perspective distortions in face recognition». En: *Proceedings of the International Conference on Computer Vision Theory and Applications*. Vol. 1. 2012, págs. 718-722.
- [29] Stuart Russell y Peter Norvig. *Artificial Intelligence: A Modern Approach 4<sup>th</sup> edition*. Pearson, 2021.
- [30] Nafiz Shahriar. *What is Convolutional Neural Network — CNN (Deep Learning)*. Accedido el 12 de Marzo de 2024. 2023. URL: <https://nafizshahriar.medium.com/what-is-convolutional-neural-network-cnn-deep-learning-b3921bdd82d5>.
- [31] Nicole Spaun. «Facial Comparisons by Subject Matter Experts: Their Role in Biometrics and Their Training». En: *International Conference on Biometrics*. 2009, págs. 161-168.
- [32] Nicholas Tinelli. *Sensores y factor de recorte: en palabras fáciles*. Accedido el 17 de Marzo de 2024. 2020. URL: <https://nicholastinelli.com/es/sensores-y-factor-de-recorte-en-palabras-faciles/>.
- [33] Joachim Valente y Stefano Soatto. «Perspective distortion modeling, learning and compensation». En: *IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2015, págs. 9-16.
- [34] Andrea Valsecchi. *Comprensión de los parámetros de la cámara*. Accedido el 17 de Febrero de 2024. 2019. URL: <https://skeleton-id.com/investigaciones/comprehension-de-los-parametros-de-la-camara>.
- [35] Gal Vardi. «On the Implicit Bias in Deep-Learning Algorithms». En: *Communications of the ACM* 66 (2022), págs. 86-93.
- [36] Zhixiang Wang et al. «DisCO: Portrait Distortion Correction with Perspective-Aware 3D GANs». En: *arXiv preprint arXiv:* (2023).
- [37] Rikiya Yamashita et al. «Convolutional neural networks: an overview and application in radiology». En: *Insights into Imaging* 9 (2018), págs. 611-629.

- [38] Guangle Yao, Tao Lei y Jiandan Zhong. «A review of Convolutional-Neural-Network-based action recognition». En: *Pattern Recognition Letters* 118 (2019), págs. 14-22.
- [39] Yajie Zhao et al. «Learning Perspective Undistortion of Portraits». En: *17th IEEE/CVF International Conference on Computer Vision*. 2019, págs. 7848-7858.

## Búsquedas en Scopus

### Estimación de la distancia en fotografías faciales (430)

TITLE-ABS-KEY ( ( distance OR depth ) AND estimation AND ( photographs OR images ) AND facial ) AND ( LIMIT-TO ( SUBJAREA , “COMP” ) OR LIMIT-TO ( SUBJAREA , “ENGI” ) )

### Estimación de la distancia en fotografías faciales mediante IA (218)

TITLE-ABS-KEY ( ( deep AND learning ) OR ( machine AND learning ) OR ( artificial AND intelligence ) OR ( computer AND vision ) OR ( soft AND computing ) AND ( ( distance OR depth ) AND estimation AND ( photographs OR images ) AND facial ) ) AND ( LIMIT-TO ( SUBJAREA , “COMP” ) OR LIMIT-TO ( SUBJAREA , “ENGI” ) )

### Estimación de la distancia en fotografías faciales mediante deep learning (126)

TITLE-ABS-KEY ( ( deep AND learning ) AND ( ( distance OR depth ) AND estimation AND ( photographs OR images ) AND facial ) ) AND ( LIMIT-TO ( SUBJAREA , “COMP” ) OR LIMIT-TO ( SUBJAREA , “ENGI” ) )