UNIVERSIDAD DE MÁLAGA

E.T.S. INGENIERÍA INFORMÁTICA
UNIVERSIDAD DE MÁLAGA

Graduada en Ingeniería de la Salud

Detección de iris de personas en movimiento para identificación biométrica mediante redes neuronales

Iris detection of people in motion for biometric identification using neuronal networks

Realizado por
Paula Mariam Lozano Soria

Tutorizado por
Antonio Jesús Bandera Rubio
Camilo Andrés Ruiz Beltrán

Departamento
Tecnología Electrónica

UNIVERSIDAD DE MÁLAGA

MÁLAGA, septiembre de 2025

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA INFORMÁTICA
GRADO EN INGENIERÍA DE LA SALUD

Detección de iris de personas en movimiento para identificación biométrica mediante redes neuronales

Iris detection of people in motion for biometric identification using neural networks

Realizado por
Paula Mariam Lozano Soria

Tutorizado por
Antonio Jesús Bandera Rubio
Camilo Andrés Ruiz Beltrán

Departamento
Tecnología Electrónica

UNIVERSIDAD DE MÁLAGA
MÁLAGA, SEPTIEMBRE DE 2025

Fecha defensa: Septiembre de 2025

Resumen

El reconocimiento de iris es uno de los métodos de identificación más fiables y precisos que se encuentran en la actualidad, se basa en el análisis de los patrones que presenta la parte coloreada que rodea la pupila, este patrón es único para cada individuo y se mantiene estable a lo largo de la vida, lo que reduce considerablemente la tasa de error y minimiza el riesgo de suplantación de identidad. Pero tradicionalmente este proceso requiere imágenes capturadas en condiciones controladas con sujetos inmóviles y a corta distancia. Este Trabajo Fin de Grado consistirá en el desarrollo e implementación de un sistema automático de detección de iris en imágenes digitales. Dichas imágenes se extraen de secuencias de videos de personas en movimiento, haciendo un recorrido caminando en línea recta acercándose al sensor. Al problema del movimiento se le suma la dificultad del enfoque y la nitidez de las imágenes. En este proyecto se aborda el desafío de adaptar esta tecnología a escenarios más realistas como accesos automáticos o entornos de seguridad. La solución propuesta entrena un modelo de aprendizaje profundo, una versión modificada de la red neuronal convolucional (CNN) YOLOX, entrenada para realizar la detección y segmentación del iris en tiempo real.

Palabras clave: **detección de iris, identificación biométrica, redes neuronales convolucionales, YOLOX.**

Abstract

Iris recognition is one of the most reliable and accurate identification methods currently available, based on the analysis of patterns in the colored part surrounding the pupil. This pattern is unique to everyone and remains stable throughout life, which considerably reduces the error rate and minimizes the risk of identity theft. But traditionally this process requires images captured under controlled conditions with stationary subjects and at close range. This Final Degree Project will consist of the development and implementation of an automatic iris detection system for digital images. These images are extracted from video sequences of people in motion, walking in a straight line approaching the sensor. The problem of movement is compounded by the difficulty of focusing and sharpness of the images. This project addresses the challenge of adapting this technology to more realistic scenarios such as automatic access or security environments. The proposed solution trains a deep learning model, a modified version of the YOLOX convolutional neural network (CNN), trained to perform real-time iris detection and segmentation.

Keywords: **Iris detection, biometric identification, convolutional neural networks, YOLOX.**

Índice

1

Introducción

1.1. Motivación

El reconocimiento biométrico se ha consolidado como una tecnología fundamental en el diseño de sistemas de seguridad, acceso y autenticación, debido a su capacidad para identificar de forma precisa y única a las personas a partir de características fisiológicas. Dentro de las distintas modalidades biométricas, el reconocimiento de iris destaca por su alta fiabilidad y bajo índice de falsos positivos, gracias a la complejidad y singularidad del patrón de iris de cada individuo.

No obstante, la mayoría de los sistemas de identificación por reconocimiento de iris existentes están diseñados para operar bajo condiciones controladas, que implican que el usuario permanezca estático y a una distancia corta del sensor. Además, el usuario debe colaborar en su propia identificación, moviéndose levemente hasta conseguir que el ojo quede correctamente centrado en el sensor de imagen. Estas restricciones limitan significativamente el despliegue de

tecnologías de reconocimiento de iris en escenarios del mundo real, como en puntos de control masivos, accesos rápidos o entornos con alta circulación de personas.

La motivación principal de este trabajo radica en abordar estas limitaciones mediante el desarrollo de un sistema capaz de realizar la detección del iris en sujetos que se encuentran en movimiento, caminando hacia el sensor, y desde una distancia considerable. Este enfoque pretende avanzar hacia soluciones biométricas más flexibles, no invasivas y eficientes que permitan una identificación rápida y precisa sin necesidad de que el usuario detenga su marcha o se coloque en una posición específica. La detección y segmentación del iris en estas condiciones dinámicas supone un reto técnico importante ya que las variaciones de iluminación, enfoque, movimiento y ángulo de visión complican la captura y procesamiento de imágenes. También es un reto relevante el integrar esta solución en un hardware de tamaño y peso relativamente reducido y con bajo consumo. Sin embargo, esto conseguiría que el dispositivo se pudiera ubicar en prácticamente cualquier localización.

El diseño de una solución capaz de detectar el iris de la persona en movimiento y a distancia (*Iris At A Distance*, IAAD) constituye una línea de investigación del grupo de Ingeniería de Sistemas Integrados de la Universidad de Málaga [5]. En su última versión, la solución implementada empleaba la red neuronal convolucional (CNN) YOLO v3 Tiny para la detección de ojos. Esta solución se integra en un MPSoC (*MultiProcessor System-On-Chip*) de AMD/Xilinx, la UltraScale+ ZU4EV, montada en el micromodulo TE0820-03-4DE21FA de Trenz[1]. La propuesta desarrollada en este TFG se integrará en esta misma plataforma.

[1] https://shop.trenz-electronic.de/en/Products/Trenz-Electronic/TE08XX-Zynq-UltraScale/

1.2. Objetivos

El objetivo principal de este Trabajo Fin de Grado es entrenar, evaluar y validar una red neuronal convolucional (CNN) basada en la arquitectura YOLOX. El diseño se construye sobre la propuesta de sistema de detección de ojos en un escenario de reconocimiento de iris a distancia llevada a cabo por el cotutor de este trabajo Camilo A. Ruiz Beltrán. En este caso, en lugar de detectar una única clase, el ojo, se detectarán tanto los ojos como las zonas de iris/pupila, con la idea de robustecer el proceso de detección eliminando falsos positivos. Además, se modifica la CNN empleada, con el objetivo de reducir el consumo de recursos en la parte programable del MPSoC. Es importante destacar que el sistema se centra sólo en el paso de la detección de las zonas de iris/pupila en un escenario IAAD. Para ello, se ha desarrollado un proceso iterativo que incluye la creación de una base de datos, múltiples ciclos de entrenamiento y validación experimental en el laboratorio.

Con el fin de alcanzar este objetivo general, se plantean los siguientes objetivos específicos:

- Diseñar y construir una base de datos propia de imágenes de rostros de personas en posición frontal, capturadas en condiciones no controladas de iluminación con sujetos caminando hacia el sensor. En las imágenes capturadas se etiquetarán tanto ojos como zonas de iris/pupila. Es importante resaltar que no existe base de datos conocida de estas características que sirva como conjunto de entrenamiento y validación para el diseño de modelos de redes neuronales que trabajen en entornos IAAD.

- Preparar y comprender el entorno de trabajo, incluyendo la estructura de la red, el tratamiento de los datos de entrada, y las condiciones técnicas necesarias para el entrenamiento de modelos de aprendizaje profundo.

- Entrenar la CNN mediante sucesivos ciclos, analizando el impacto de distintos parámetros sobre el rendimiento del modelo y revisando los resultados obtenidos con imágenes de la base de datos antes de llevar a cabo las pruebas en tiempo real.

- Realizar pruebas prácticas en laboratorio tras cada ciclo de entrenamiento, evaluando visual y cuantitativamente los resultados, con el fin de validar la capacidad del modelo para detectar el iris de personas en movimiento.

- Identificar posibles limitaciones del sistema desarrollado y proponer mejoras o futuras líneas de trabajo en el ámbito del reconocimiento biométrico en condiciones no ideales.

- Contextualizar los resultados obtenidos dentro del campo de la biometría ocular, analizando su aplicabilidad en entornos reales como el control de accesos o la identificación en tiempo real.

1.3. Estructura de la memoria

En este apartado se explican brevemente los capítulos que componen la memoria de este proyecto.

- Capítulo 1 Introducción. En esta sección se presenta la motivación, un resumen de los objetivos del proyecto y la estructura de la memoria.

- Capítulo 2 Identificación biométrica. Se aborda el concepto de biometría, un breve resumen de su historia, junto a sus principios y fundamentos, las diferentes clasificaciones y se exponen algunas de las técnicas de identificación biométrica más utilizadas en la actualidad.

- Capítulo 3 Biometría ocular. Este capítulo se centra en el reconocimiento de iris, explicando el origen de la biometría ocular, los diferentes parámetros que se pueden medir y sus aplicaciones. Centrándose posteriormente en el iris como principal rasgo biométrico.

- Capítulo 4 Redes neuronales. En este apartado se expone una visión general de las redes neuronales artificiales, comenzando por sus fundamentos inspirados en el funcionamiento del cerebro. Se describe la estructura en capas y el funcionamiento de una red básica, para luego centrarse en las rede neuronales convolucionales, dado que son la base del sistema de detección de iris.

- Capítulo 5 Desarrollo. En esta fase se detalla todo el proceso práctico del proyecto, desde la creación de la base de datos, el procesamiento y etiquetado de imágenes y el entrenamiento de la red neuronal, hasta la validación del modelo. Incluyendo una descripción detallada de todo el sistema hardware utilizado.

- Capítulo 6 Resultados y discusión. Se presentan y comparan los resultados obtenidos con la red neuronal entrenada, analizando los parámetros de rendimiento y se evalúa la eficiencia del sistema propuesto en el reconocimiento del iris.

- Capítulo 7 Conclusiones y líneas futuras. Trata de las conclusiones sacadas tras la finalización del proyecto, conteniendo las dificultades y posibles mejoras que se han encontrado en el desarrollo de éste.

2

Identificación biométrica

Este apartado comprende un breve resumen de los acontecimientos más relevantes en la historia de la biometría, junto a sus fundamentos, formas de clasificación y las diferentes técnicas de identificación que se encuentran en la actualidad.

2.1. Historia de la biometría

El primer concepto de biometría surgió en 1882, en el Servicio de Identificación de Policía de Paris, donde Alphonse Bertillon llevó a cabo el primer sistema de identificación de personas basado en rasgos físicos, con el objetivo de identificar criminales, al cual denominó antropometría. El sistema de Bertillon consistía en recoger once medidas de todo el cuerpo, desde la altura del sujeto hasta la longitud del dedo meñique izquierdo [6].

En las Figuras 2.1 y 2.2 se muestran imágenes extraídas del libro de Bertillon donde explicaba las once medidas y como realizarlas, en la primera imagen se muestra la altura del sujeto sentado y en la segunda imagen aparece una clasificación de iris diferenciándolos por color.

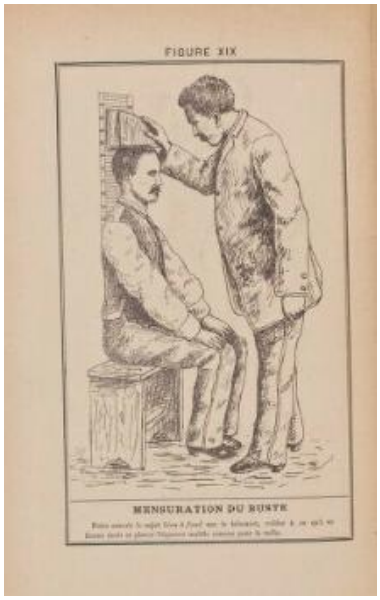

Figura 2.1. Procedimiento para la toma de medidas del método Bertillonage. Extraída de A. Bertillon, *Identification anthropométrique: Instructions signalétiques*, 1885.

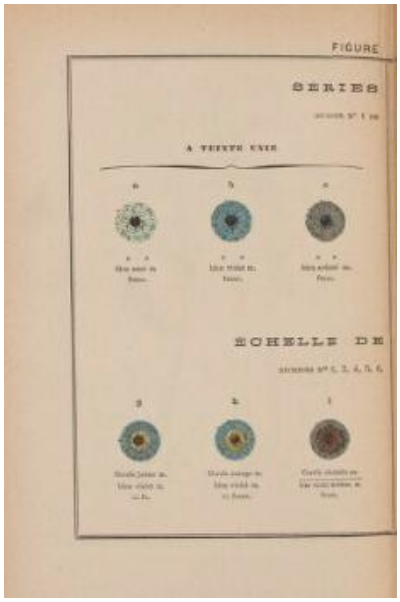

Figura 2.2. Clasificación de iris según el patrón y color del método Bertillonage. Extraída de A. Bertillon, *Identification anthropométrique: Instructions signalétiques*, 1885.

El método Bertillonage fue muy utilizado, pero aun así tenía fallos debido a los errores que se podían cometer en la medida y a que dichas medidas podían cambiar con el paso del tiempo. Años más tarde, el jefe de policía de Buenos Aires, Juan Vucetich, resolvió el crimen de Francisca Rojas tras encontrar el rastro de una huella dactilar en sangre [6]. Aunque hay evidencias de que las huellas dactilares ya se usaban en la antigüedad, no es hasta 1892 cuando Sir Francis Galton, científico británico, publicó "*Fingerprints*". Su objetivo era usarlas como ayuda para la determinación de la herencia. Aunque no encontró esas características en las huellas dactilares, sí que pudo afirmar que estas eran únicas y que no cambian en el transcurso de la vida de un individuo [8] [10]. Galton también proporcionó el primer método de clasificación de huellas dactilares, que fue puesto en funcionamiento por Edward Henry en Scotland Yard [6]. Hoy en día la huella dactilar sigue siendo un método ampliamente utilizado para la identificación de los ciudadanos, en el caso de España gracias al Sistema Automático de Identificación Biométrico (ABIS) con una base de datos gestionada por el Estado.

En las últimas décadas, se han desarrollado diversas técnicas de identificación, muchas de las cuales forman parte de nuestra vida cotidiana. Por ejemplo, en dispositivos móviles, como los teléfonos inteligentes, es frecuente emplear el reconocimiento facial para desbloquear la pantalla principal o acceder a aplicaciones. Esta tecnología está presente en la mayoría de los smartphones actuales. En el caso de dispositivos de la marca Apple, éstos llevan incorporado el asistente Siri, que identifica y responde únicamente a la voz del propietario, mejorando así la personalización del sistema.

2.2. Principios y fundamentos de la biometría

La palabra biometría proviene del griego bio- *"bios",* que significa vida, y -metría (en griego "*metron*"), que se traduce como medida. La biometría se define como la ciencia relacionada con las tecnologías que analizan las características humanas para reconocer o verificar la identidad del individuo de manera automática [9]. Estos sistemas constan de varias etapas esenciales: la captura de la muestra biométrica, el procesamiento para extraer las características relevantes, el almacenamiento en bases de datos, y la comparación con registros previos para la autenticación [6].

Las técnicas biométricas se clasifican según las funciones y el tipo de características que analiza.

2.2.1. Clasificación según sus funciones

Según el tipo de aplicación, los sistemas biométricos son capaces de verificar o identificar. El modo de verificación, también conocido como autenticación, consiste en comprobar si una persona es realmente quien afirma ser. Esto se puede hacer contrastando sus datos biométricos con un registro. Por otro lado, el modo de identificación se emplea para determinar la identidad de una persona dentro de un conjunto de registros [11].

2.2.2. Clasificación según la característica a analizar

Dependiendo de la característica que se quiera analizar podemos distinguir:

- Biometría fisiológica. Se basa en los rasgos físicos y biológicos propios de cada individuo, los cuales son generalmente permanentes y únicos a lo largo del tiempo. En este grupo se encuentran atributos altamente fiables, debido

a su dificultad para ser modificados, lo que hace que también sean los más utilizados.

- Biometría conductual. Se compone del análisis de patrones de comportamiento. También se consideran características propias de cada persona, pero estas pueden estar sujetas a variaciones según factores externos o emocionales.

En la Tabla 2.1 se muestran diferentes ejemplos de rasgos dependiendo si son fisiológicos o conductuales.

TIPO DE BIOMETRÍA	**EJEMPLOS DE CARACTERÍSTICAS**
BIOMETRÍA FISIOLÓGICA	- Huellas dactilares - Rostro - Iris - Retina - Geometría de la mano - Patrones de venas - ADN
BIOMETRÍA CONDUCTUAL	- Firma - Voz - Patrón de la marcha - Manera de teclear

Tabla 2.1. Ejemplos de características según su clasificación. Elaboración propia.

La gráfica de la Figura 2.3 refleja el porcentaje de ingresos asociados a cada tipo de rasgo biométrico, el análisis de estos ingresos ofrece una perspectiva sobre qué

métodos son los más populares y extendidos en la actualidad. En primer lugar, se encuentra la huella dactilar siendo el rasgo más antiguo y el que más ingresos genera.

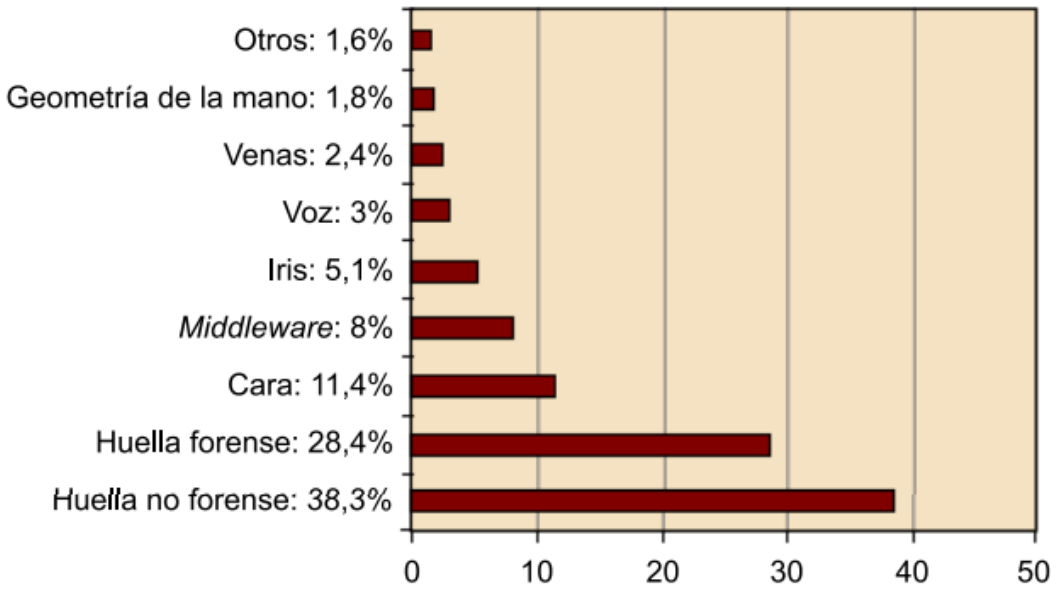

Figura 2.3. Porcentaje de los ingresos generados por las diferentes características biométricas. Extraída de F. Serratosa, *La biometría para la identificación de las personas*. Universitat Oberta de Catalunya, 2008.

2.3. Técnicas de identificación biométrica

Las técnicas de identificación biométricas permiten verificar o determinar la identidad de un individuo a través de análisis de rasgos únicos, como los que se ejemplificaban en la Tabla 2.1.

Dentro de las técnicas de identificación se encuentran:

- Reconocimiento de huellas dactilares: se basa en el análisis de los patrones únicos presentes en las crestas y surcos de la piel en la punta de los dedos. Esta es una técnica de alta precisión y facilidad de captura, aunque necesite

contacto. Se emplea tanto en sistemas de control de acceso físico como digital. Se puede encontrar comúnmente en dispositivos móviles para desbloqueo seguro, aunque está siendo reemplazado por el reconocimiento facial. Está presente en el DNI electrónico, donde se almacena la imagen de la huella dactilar en el chip del documento.

- Reconocimiento facial: estos sistemas clasifican la apariencia del sujeto e intentan medir algunos puntos específicos, como la distancia entre los ojos, el ancho de la nariz, la distancia del ojo a la boca o la forma de la mandíbula. Aunque no tiene el mismo grado de exactitud que el reconocimiento de huellas dactilares, presenta la ventaja de funcionar sin contacto físico, lo que lo hace especialmente útil y no invasivo. Esta técnica se utiliza ampliamente en sistemas de seguridad para el control de acceso en aeropuertos y dispositivos móviles.

- Firma: se analiza la manera en la que una persona firma su nombre, considerando factores como la velocidad, presión y ritmo del trazo. Se utiliza principalmente en aplicaciones financieras.

- Identificación por ADN: es una técnica biométrica molecular que analiza la secuencia genética única de cada individuo. Es especialmente utilizada en medicina forense y pruebas de paternidad. Pese a ser altamente fiable presenta varias desventajas. Por un lado, se considera una técnica invasiva debido al procedimiento necesario para la recolección de muestra biológicas, y, por otro lado, implica costes elevados y tiempos de procesamiento superiores a otras técnicas biométricas [6].

- Reconocimiento de la retina: esta técnica biométrica se basa en el análisis del patrón único de los vasos sanguíneos situados en la parte posterior del ojo, concretamente la retina. Este patrón es altamente distintivo, incluso comparando el ojo izquierdo con el derecho de una misma persona, y permanece prácticamente inalterable a lo largo de la vida. Para llevar a cabo el escaneo es necesario hacer uso de una fuente de luz infrarroja de baja intensidad, y el sujeto no podrá portar gafas ni lentes de contacto que impidan el paso de la luz. A pesar de la precisión, se necesita la interacción del usuario y el coste del hardware es bastante elevado [11].

Tras haber realizado este breve recorrido introductorio por el ámbito de la biometría, sus aplicaciones, técnicas y características principales, nos enfocaremos en la identificación biométrica mediante el iris ocular.

3

Biometría ocular

Este capítulo se centrará en estudiar el iris como principal elemento de identificación biométrica, comenzando por un resumen sobre el origen del reconocimiento de iris. A continuación, se presentan los fundamentos anatómicos con el objetivo de contextualizar. Por último, se analizan las ventajas y limitaciones asociadas al uso del iris en sistema de reconocimiento.

3.1. El origen del reconocimiento de iris

Dentro del campo de la biometría ocular se encuentran varios rasgos que resultan únicos en cada individuo y han sido utilizados para la identificación. Entre los más relevantes se pueden destacar el patrón del iris, la estructura vascular de la retina o la disposición de las venas que componen la esclera. La primera patente fue lanzada por Robert Hill en 1978. Se trataba de un método de identificación mediante el patrón vascular de la retina. Básicamente consistía en escanear el ojo del sujeto con una fuente de luz, detectando la disposición de vasos sanguíneos que se reflejaba en la retina y compararlo con un modelo de referencia para determinar la identidad de la persona [14].

Tal y como se mencionó en el apartado anterior, Alphonse Bertillone ya utilizaba el color del iris para identificación, como se muestra en su libro *Identification anthropométrique : Instructions signalétiques*. Fue en 1936 cuando el oftalmólogo Frank Burch propuso por primera vez el concepto de autenticación de un sujeto mediante el patrón del iris [1]. Posteriormente F. H. Adler escribió: "De hecho, las marcas del iris son tan distintivas que se ha propuesto utilizar fotografía como medio de identificación, en lugar de huellas dactilares" [2]. Pero la idea de utilizar el iris como rasgo biométrico no cobró relevancia hasta 1987, con la patente de los doctores oftalmólogos Leonard Flom y Aran Safir. En esta patente se describen técnicas y herramientas para identificar el ojo de una persona basándose en las características visibles del iris y la pupila [15]. En 1994, con la ayuda del profesor John Daugman de la Universidad de Cambridge, consiguieron desarrollar un algoritmo para automatizar la detección del iris humano [3]. Daugman introdujo un algoritmo basado en transformadas de onda Gabor bidimensional, que permitía extraer un código único de cada patrón de iris, llamado IrisCode.

El físico explica que los patrones del iris poseen una variabilidad tan elevada entre individuos (incluso entre gemelos monocigóticos, o entre los dos ojos de una misma persona) que permiten lograr tasas extremadamente bajas de falsas coincidencias [4]. Todos estos avances dieron lugar a que poco tiempo después, el sistema de reconocimiento fuese implementado progresivamente en contextos como centros penitenciarios y puntos de control en aeropuertos [16].

En la Figura 3.1 se presentan dos ejemplos de imágenes monocromáticas oculares utilizadas por Daugman en el desarrollo de su algoritmo IrisCode con la finalidad de codificar la secuencia de fase correspondiente a los patrones del iris [4].

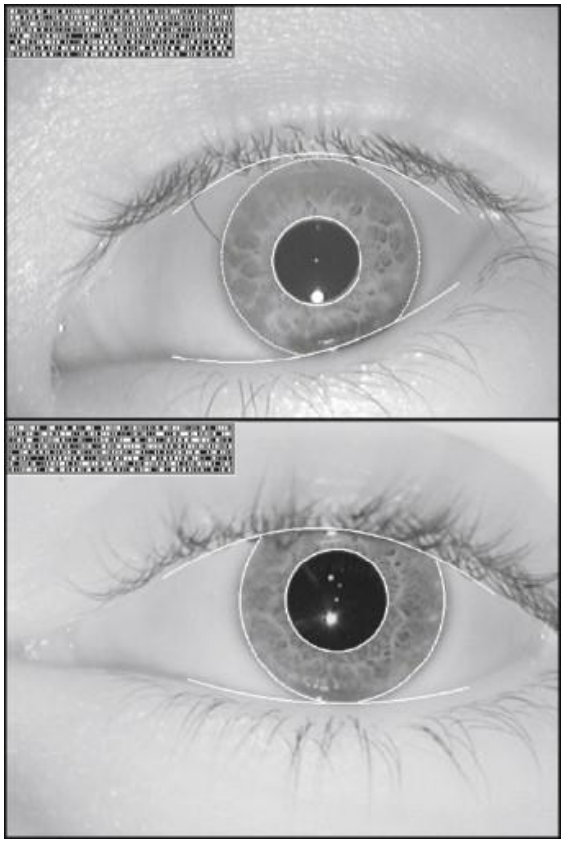

Figura 3.1. Ejemplo de patrón de iris. Extraído de J. Daugman, "How iris recognition works," IEEE Transactions on Circuits and Systems for Video Technology, vol. 14, no. 1, pp. 21–30, 2004.

En la actualidad la investigación en este campo ha cambiado significativamente, centrándose en abordar nuevos desafíos como el uso de gafas o lentes de contacto (correctivas transparentes o estéticas de color), la distancia entre el sensor y el sujeto, el ángulo de inclinación de la cabeza o la captura de datos de personas caminando como es nuestro caso. El desarrollo de estas nuevas mejoras hace que el sistema biométrico se aproxime a un entorno más real, en el que las pautas para la identificación del sujeto no sean tan estrictas.

3.2. Anatomía y fisiología ocular

Con el fin de comprender en profundidad el funcionamiento de los sistemas de reconocimiento basados en el iris, resulta imprescindible conocer los aspectos

anatómicos y fisiológicos más importantes del ojo humano. Este apartado se centra en describir brevemente la morfología del ojo, destacando las funciones y particularidades del iris.

El ojo, sitiado en la cavidad orbitaria, es el órgano especializado en la percepción visual [18], compuesto por múltiples estructuras anatómicas que trabajan de forma coordinada para captar y procesar la luz. Está conformado por tres capas:

1. La capa externa fibrosa, y por lo tanto la más resistente, está formada por la córnea y la esclerótica. Sirve de protección para las siguientes capas.

2. La capa media o vascular, llamada úvea, está compuesta por el iris, el cuerpo ciliar y la coroides.

3. La capa interna neurosensorial, conocida como la retina, se encarga de recibir los estímulos luminosos y transformarlos en señales nerviosas. Estas señales son transmitidas a través del nervio óptico hacia el cerebro, donde se interpretan como imágenes visuales.

Además de la clasificación del globo ocular en capas, también puede considerarse su división en dos segmentos [19]: el segmento anterior (SA) y el segmento posterior (SP), que presentan una disposición asimétrica. El SA comprende todas las estructuras situadas por delante del cristalino, incluyendo la córnea, el iris y el cuerpo capilar. A su vez se subdivide en la cámara anterior (CA), ubicada entre la córnea y el iris, y la cámara posterior (CP), localizada entre el iris y la cara anterior del cristalino. Por otro lado, el SP abarca desde la cara posterior de

cristalino hasta la retina, incluyendo el humor vítreo, la retina, la coroides y el nervio óptico.

3.2.1. Capa externa

La esclerótica también conocida como la zona blanca del ojo, ocupa la región posterior abarcando las tres cuartas partes de la superficie ocular, y contiene el humor vítreo, sustancia transparente y viscosa que ocupa el espacio entre la retina y el cristalino aportando estructura. La córnea se encuentra en la parte anterior del globo ocular, es transparente y permite la entrada de luz en el ojo, a su vez consta de cinco capas: epitelio, capa de Bowman, estroma, membrana de Descemet y endotelio corneal [21]. Estas dos estructuras están unidas por el limbo y ambas están compuestas por el mismo material fibroso. La diferencia entre ellas se debe a la falta de paralelismo en la disposición entre las fibras de colágeno, lo que le confiere ese aspecto blanquecino a la esclerótica. En el caso de la córnea, dichas fibras tienen una orientación paralela y uniforme [19].

3.2.2. Capa vascular

La capa media es un tejido con gran cantidad de vasos sanguíneos que aportan nutrientes a la retina [19], compuesta por la úvea, que a su vez se compone por el iris (la parte coloreada del ojo), el cuerpo ciliar y la coroides. El cuerpo ciliar, formado por el músculo ciliar y el epitelio ciliar, es el encargado de dividir la cámara posterior y el cuerpo vítreo. La coroides consiste en una densa red de capilares que tienen como función principal nutrir y oxigenar la estructura del ojo [21].

3.2.3. Capa interna

Se trata de la capa nerviosa/sensorial. En ella se encuentra la retina, considerada la superficie más interna del globo, y la responsable de la función fotorreceptiva.

La retina está compuesta por múltiples capas de células, entre las que destacan los fotorreceptores, bastones y conos, responsables de captar la luz y convertirla en señales eléctricas. Estas señales son procesadas a través de la red de interneuronas que forman el nervio óptico, siendo éste el encargado de transmitir la información visual al cerebro.

En la imagen que se muestra en la Figura 3.2 se pueden identificar las diferentes estructuras explicadas.

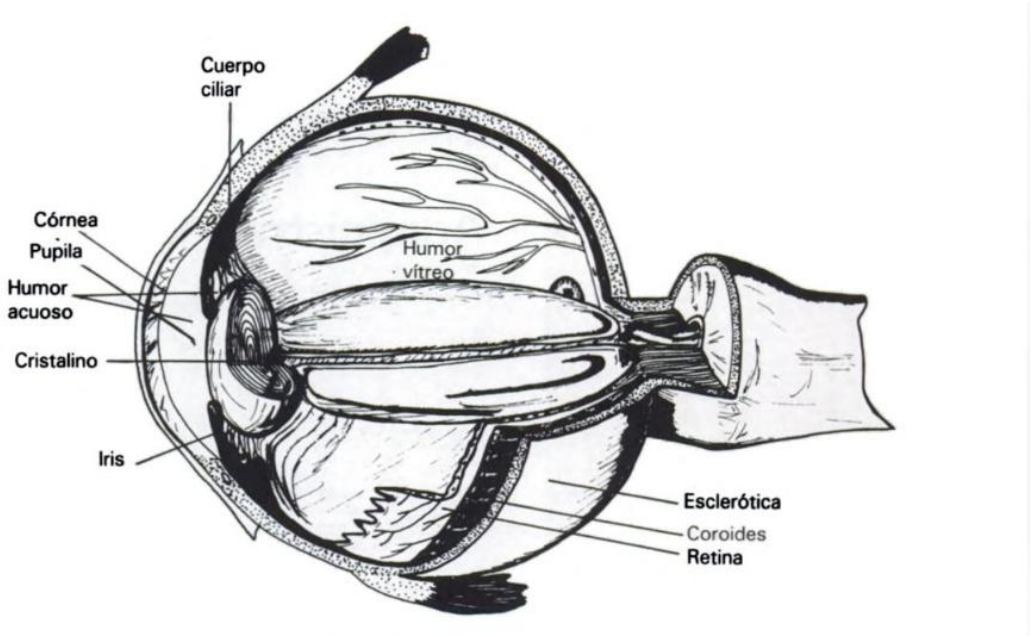

Figura 3.2. Corte sagital del globo ocular. Extraído de M. Piña, Vía oftálmica: Anatomía y fisiología ocular. Administración de medicamentos: teoría y práctica, p. 99, 1994.

3.2.4. El iris

Mientras que el resto de estructuras oculares se describen brevemente para contextualizar el funcionamiento del ojo, en este subapartado se procede a analizar en mayor profundidad el iris.

El iris es una estructura circular ubicada en la SA del globo ocular, específicamente entre la córnea y el cristalino, formando parte de la capa media vascular. Anatómicamente, el iris está compuesto por cuatro capas [22]:

1. El epitelio posterior es la capa más posterior del iris, no visible, y en contacto con la retina ciega. Constituye el fondo de la estructura del iris. Este epitelio también es conocido como epitelio pigmentario.

2. La siguiente capa abarca una pequeña zona muscular, formada por dos músculos antagonistas, el músculo dilatador y el músculo esfínter. Ambos están compuestos por fibras musculares lisas.

    - El músculo dilatador está inervado por el sistema nervioso simpático. Tal y como indica su nombre es el encargado de dilatar la pupila.

    - Al contrario que el anterior, el músculo esfínter provoca la contracción del agujero pupilar. Bordea la pupila y puede ser observado con facilidad en algunos iris. Está controlado por el sistema parasimpático.

    En la siguiente imagen se puede observar el funcionamiento de estos músculos. Cuando el ojo se expone a una luz brillante, actúa el músculo esfínter, contrayendo la pupila, a este proceso se le conoce como miosis. Por lo contrario, cuando el entorno es oscuro, se lleva a cabo la midriasis, o abertura de la pupila, por el músculo dilatador [25]. Por esta razón se optó por utilizar luz infrarroja, evitando la dilatación pupilar.

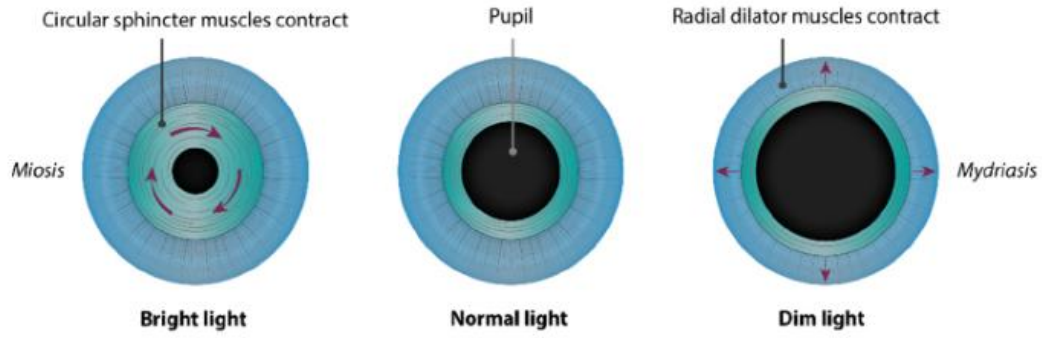

Figura 3.3. Adaptación de la apertura pupilar a diferentes intensidades luminosas. Extraído de C. Angée, B. Nedelec, E. Erjavec, J. Rozet y L. F. Taie, "Congenital Microcoria: Clinical Features and Molecular Genetics," Genes, vol. 12, no. 5, p. 624, 2021.

3. Por encima del endotelio se encuentra el estroma, un tejido conjuntivo fibrovascular que contiene melanocitos. Éstos son los responsables de la producción de melanina, el pigmento que determina el color del ojo. El color del iris no está determinado por pigmentos de color azul o marrón, sino por la interacción de la luz con las capas que lo componen. En el caso de los ojos oscuros (marrones o negros), existe una alta concentración de melanina en el estroma, lo que provoca una mayor absorción de luz. En cambio, en los ojos claros (azules o verdes), hay menor cantidad de melanina, por lo que la luz se dispersa (fenómeno de dispersión de Rayleigh) [19] [21] [22].

4. El endotelio o epitelio anterior, se sitúa por encima de las anteriores, es una capa prácticamente unicelular y discontinua, presentando grandes perforaciones donde se sitúan las lagunas y criptas, denominadas estomas de Fuchs.

Fisiológicamente se considera un obturador muscular con alta pigmentación que se cohesiona periféricamente al cuerpo ciliar. Su principal función es actuar como diafragma, modulando la entrada de luz en función del estímulo luminoso [2].

En la Figura 3.4 se pueden distinguir con detalle cada una de las capas del iris, mostrando las estructuras internas relevantes para su funcionamiento fisiológico.

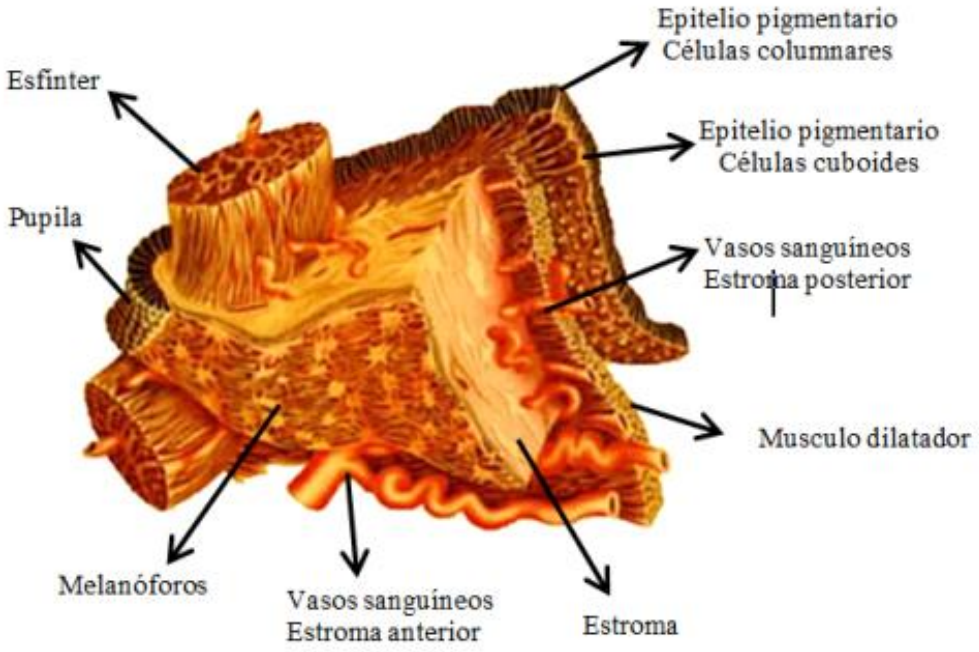

Figura 3.4. Estructura anatómica del iris organizada por capas. Extraído de L. Brusi, *El iris. Exploración con biomicroscopio ocular: técnicas y protocolo de intervención*, pp. 338–351. Editorial de la Universidad Nacional de La Plata (EDULP), 2014.

Durante la etapa prenatal, el iris comienza su desarrollo en el tercer mes de gestación, alcanzando su estructura prácticamente definitiva en el octavo mes [25]. El proceso de pigmentación sin embargo continúa hasta los primeros años de vida, pudiendo variar de tonalidad hasta conseguir el color definitivo. Esto se debe a que la melanina sigue acumulándose durante los primeros meses después de nacimiento. Este tejido presenta un patrón altamente complejo, en el que se

combinan múltiples características distintivas, como pueden ser ligamentos arqueados, surcos, anillos, corona, pecas o crestas. Estos elementos pueden observarse en las Figuras 3.1 y 3.5.

La imagen presentada en la Figura 3.5 muestra una vista frontal del iris, destacándose sus principales estructuras anatómicas externas, las cuales se encuentran en la cara visible. A simple vista el iris está dividido en dos anillos concéntricos. El anillo mayor, ubicado en la parte periférica, próximo al cuerpo ciliar, es la zona más amplia del iris y contiene el músculo dilatador. Por otro lado, el anillo menor se encuentra en la parte central, próxima a la pupila, en la zona en que se localiza el músculo esfínter pupilar. Los pliegues, situados en la zona ciliar, se originan como resultado de las contracciones de dichos músculos. El borde pupilar delimita la pupila y marca la inserción del músculo esfínter, mientras que el borde ciliar constituye el límite externo del iris.

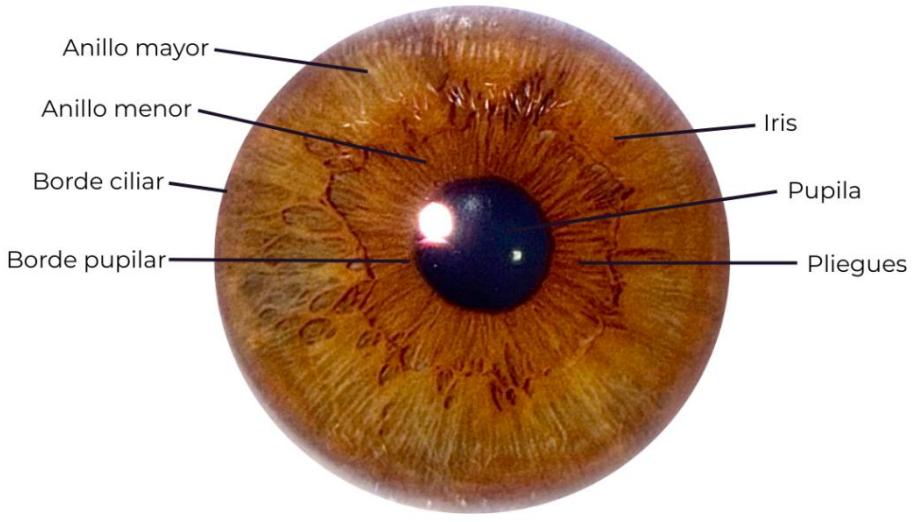

Figura 3.5. Segmento anterior del ojo, donde se muestra la cara frontal del iris. Elaboración propia.

3.3. El iris como rasgo biométrico

Tras haber analizado la anatomía y fisiología del iris, este apartado se centrará en su estudio como característica biométrica, dado que constituye el elemento principal de análisis en el presente proyecto.

Una de las principales fortalezas del iris como rasgo biométrico es la alta variabilidad entre individuos, lo que hace que la probabilidad de encontrar falsas coincidencias sea prácticamente imposible. Gracias a las diferentes estructuras que lo componen (mostradas en el apartado anterior), el patrón del iris es complejo y fiable. Así, se estima que existen alrededor de 266 puntos diferenciables en un iris, mientras que la gran mayoría de sistemas biométricos sólo poseen entre 13 y 60 características únicas [8]. Además, a diferencia de otros rasgos que pueden verse afectados por el envejecimiento, lesiones, cambios físicos o condiciones ambientales, como es el caso de las huellas dactilares o el rostro, el tejido del iris se mantiene estable con el paso del tiempo. Esta estabilidad se debe a que las estructuras que componen el patrón se forman en la etapa fetal [25] y, salvo daños oculares graves, no sufren modificaciones a lo largo de los años. La probabilidad de deterioro externo se reduce significativamente ya que se encuentra protegido dentro del ojo, por lo que no está expuesto directamente al entorno.

Otra ventaja destacable es la dificultad de falsificación o suplantación. Al encontrarse protegido por la córnea, es difícil de replicar sin un sistema de alto nivel tecnológico. Asimismo, se trata de un procedimiento no invasivo, donde la interacción del usuario con el sistema es muy reducida. Sin embargo, en la mayoría de los sistemas actuales, la persona debe moverse ligeramente frente al sistema hasta conseguir que el ojo quede enfocado. Este aspecto se trata actualmente de minimizar, tratando que el sistema sea lo menos intrusivo posible. Por otra parte,

es un método higiénico, al no requerir contacto físico directo con el sensor. Este aspecto difiere, por ejemplo, del uso de la huella dactilar, que implica una limpieza del sensor tras cada uso.

Aunque puede presentar algunas desventajas como la sensibilidad a la calidad de la imagen, ya que se requieren imágenes con alta calidad, o incluso el rechazo por parte de los usuarios por temas de privacidad, el reconocimiento de iris sigue siendo una de las tecnologías biométricas más precisas y fiables.

3.4. Mejoras en la detección de iris

Con el objetivo de crear sistemas de reconocimiento de iris más agiles, en los que la necesidad de interacción del usuario sea mínima, surgen iniciativas como el que se presenta en este TFG. El objetivo es avanzar hacia tecnologías más robustas y adaptables, capaces de ajustarse a entornos menos estructurados, donde factores como la iluminación variable, el movimiento del sujeto o la distancia a la cámara representan desafíos importantes (*Iris At A Distance*, IAAD). Este enfoque tiene como propósito que los sistemas de reconocimiento no se limiten a aplicaciones en los que se pueda capturar el iris con la colaboración del usuario, sino que pueden integrarse eficazmente en aplicaciones cotidianas como el control de acceso o la identificación en espacios públicos.

Algunas de las mejoras para conseguir sistemas de reconocimiento de iris más robusto en entornos IAAD son:

1. Captura del sensor a distancia. En algunos sistemas el usuario debe colaborar e interactuar con el equipo, colocarse cerca del sensor en una posición determinada o en una marca concreta. El sistema espera a que el

usuario se mueva hasta conseguir que la imagen de iris capturada esté enfocada.

2. Tolerancia al movimiento. Distinguiendo dos tipos de movimiento: caminar y movimientos con la cabeza. Por un lado, la detección de iris de una persona caminando representa un gran avance hacia sistemas biométricos que permiten la identificación del usuario sin que este se detenga. Por otro, los movimientos de la cabeza son gestos naturales e, incluso a veces, involuntarios (rotaciones, inclinaciones). Estos movimientos pueden provocar que los ojos no sean frontales, lo que hace que restricciones como la circularidad de la pupila o iris no se cumplan.

3. Compatibilidad con gafas y lentillas. Que el sistema cuente con la capacidad de funcionar eficazmente cuando el usuario utiliza gafas o lentes de contacto. En el caso de las gafas es importante considerar que pueden producir reflejos o destellos. En algunos casos, el sistema exige al usuario que se quite las gafas para el reconocimiento.

4. Adaptación a cambios de iluminación. Aunque en este tipo de sistemas se necesita una fuente de iluminación potente y controlada, es también frecuente que no ésta no pueda verse interferida por fuentes de iluminación naturales no controladas.

5. Aplicables en entorno no cooperativos. Algunas de las condiciones ya nombradas hacen que esta situación de no colaboración activa del usuario con el sistema pueda ser posible. Actualmente lo normal es que se exija esta colaboración.

Por otra parte, y ajena al entorno IAAD, la visibilidad del iris puede verse afectada por factores como las pestañas o la anatomía del párpado, especialmente en las personas asiáticas. Estos elementos pueden ocultar parcialmente el iris, haciendo que su detección sea más compleja.

3.4.1. Detección de iris a distancia

Dentro de las mejoras que se destacan anteriormente, el presente proyecto aborda en gran medida dos de ellas: la captura de la imagen a distancia y que el sujeto camine hacia el sensor.

A diferencia de los sistemas tradicionales que requieren una alineación precisa del ojo a corta distancia, los enfoques IAAD buscan capturar imágenes a mayor distancia, incluso mientras el individuo está en movimiento, normalmente caminando hacia el sensor. El objetivo es minimizar la interacción del usuario con el equipo, disminuyendo situaciones que puedan incomodar al sujeto como permanecer inmóvil durante un tiempo prolongado, posicionarse a una distancia muy corta del dispositivo de captura, fijar la mirada durante cierto tiempo, entre otras. Esto implica desafíos técnicos adicionales, pues la imagen capturada no puede verse afectada por ruidos por movimiento y tener la resolución suficiente para permitir el reconocimiento.

Por otra parte, la posibilidad de realizar la captura del iris sin la intervención activa del usuario amplía considerablemente el rango de aplicaciones, pudiendo integrarlos en entornos de control de accesos a espacios públicos, estaciones o aeropuertos, lo que permitiría que el flujo de personas fuera continuo y más eficiente.

4

Redes neuronales

El presente capítulo introduce las redes neuronales artificiales, comenzando por su inspiración biológica y su estructura básica. Seguidamente, se profundiza en la redes neuronales convolucionales, explicando su arquitectura y funcionamiento.

4.1. Introducción a las redes neuronales

La inteligencia artificial (IA) se ha convertido en una de las tecnologías más influyentes en los últimos años. Desde asistentes virtuales capaces de mantener una conversación hasta sistemas de diagnóstico de enfermedades a partir de imágenes médicas. Según la *Encyclopedia of artificial intelligence* [54]:

"La IA es un campo de la ciencia y la ingeniería que se ocupa de la comprensión, desde el punto de vista informático, de los que denomina comúnmente comportamiento inteligente. También se ocupa de la creación de artefactos que exhiben este comportamiento".

Una de las subáreas más relevantes dentro del campo de la IA es el aprendizaje automático (*machine learning*) [29]. La capacidad de aprendizaje es fundamental para simular el comportamiento inteligente de los seres humanos, ya que esta les permite adquirir conocimientos y adaptarse a variaciones del entorno [30]. Sin aprendizaje una IA estaría limitada a ejecutar instrucciones, sin poder adaptarse ni mejorar con la experiencia.

Dentro del campo del aprendizaje automático las redes neuronales artificiales (RNA) representan una de las técnicas más eficientes y funcionales. Se trata de modelos computacionales inspirados en la estructura y el funcionamiento del cerebro humano. Su objetivo es emular el comportamiento de los seres humanos, desde el aprendizaje hasta la toma de decisiones. La idea de simular una red de neuronas artificiales surge en 1943, con el artículo "*A logical calculus of the ideas immanent in nervous activity*". En este artículo los autores, Warren McCulloch y Walter Pitts, propusieron un modelo lógico de neurona que podía realizar operaciones booleanas básicas [31].

4.2. Fundamento biológicos

Con el fin de emular los comportamientos del ser humano, las RNA están inspiradas en la estructura y el funcionamiento del sistema nervioso, concretamente del cerebro humano.

Las células que componen el sistema nervioso se denominan neuronas [32] [33]. Las neuronas son células excitables, que se enlazan entre sí, formando un complejo sistema neuronal. Cada neurona se compone principalmente de tres elementos: el soma (cuerpo celular), las dendritas y el axón (Figura 4.1). Estos elementos le permiten recibir, transmitir, y almacenar o procesar información. El proceso de

interacción es básico, y se fundamenta en la denominada sinapsis. Gracias a la sinapsis las neuronas son capaces de transmitir información a otras neuronas o a órganos efectores. Desde el punto de vista fisiológico, la sinapsis es el lugar donde el impulso nervioso se transmite de una neurona presináptica a otra postsináptica mediante neurotransmisores. Se pueden distinguir dos tipos de sinapsis: química y eléctrica.

Por lo tanto, y de forma muy resumida, las neuronas reciben señales eléctricas de otras neuronas a través de las dendritas, procesan estas señales en el cuerpo celular o soma, y las trasmiten a otras neuronas a través del axón. Este proceso de interconexión y procesamiento tratará de ser imitado por las RNA.

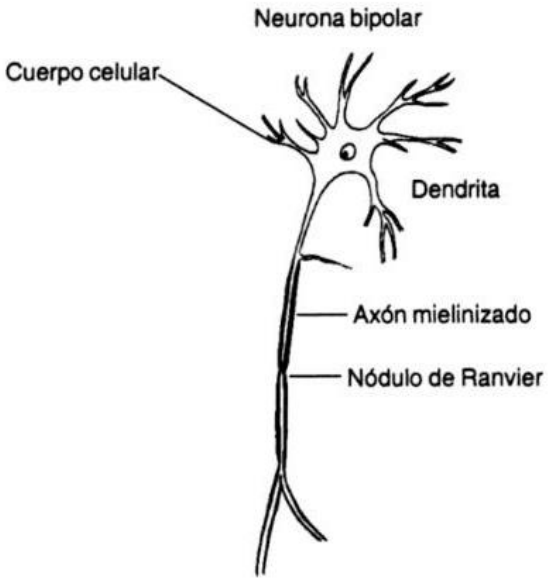

Figura 4.1. Esquema de una neurona bipolar. Extraído de V. M. Alcaraz, *Estructura y función del sistema nervioso*. UNAM, 2000.

4.3. Estructura básica

Una RNA está constituida por neuronas artificiales o nodos, que a su vez están organizadas en capas que procesan la información de forma secuencial. Se pueden distinguir tres tipos de capas [34] [28]:

1. Capa de entrada. Esta capa recibe los datos que entran en el sistema y transmite la información a la siguiente capa.

2. Capas ocultas. Son las capas intermedias. Es donde ocurre el procesamiento real de la información. Cada neurona de una capa oculta recibe señales de todas las neuronas de la capa anterior y de este modo se forma la red de nodos interconectados. Cuantas más capas ocultas tiene una red, más compleja es.

3. Capa de salida, es la encargada de generar la salida de la red. El número de nodos dependerá del tipo de problema.

En la Figura 4.2 se puede observar la distribución de los tres tipos de capas de una RNA.

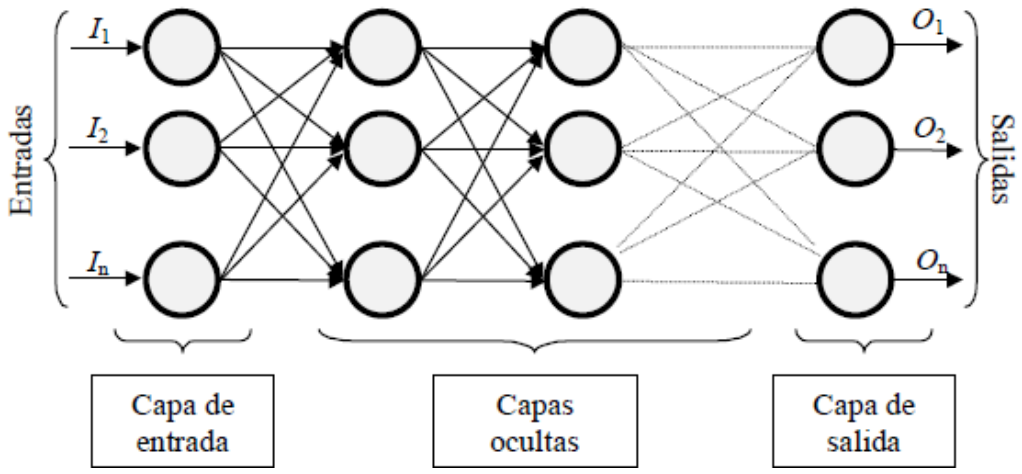

Figura 4.2. Ejemplo de la estructura de una red neuronal organizada en capas. Extraído de D. J. Matich, Redes Neuronales: Conceptos básicos y aplicaciones. Universidad Tecnológica Nacional, México, 2001, pp. 12–16.

Cada neurona artificial simula el comportamiento de una neurona biológica básica y realiza una operación matemática sobre los datos que recibe. Cada conexión entre neuronas tiene un peso, que indica la importancia que se le da a cada dato de entrada. Cuando una neurona recibe entradas (por ejemplo, los valores de intensidad de un píxel), multiplica cada entrada por su peso correspondiente. Si la neurona recibe varias entradas, realiza una suma ponderada de todas ellas, multiplicando cada valor de entrada por su respectivo peso, y sumando los resultados junto a un término adicional llamado sesgo (*bias*). A continuación, este valor pasa por la función de activación que, dependiendo del umbral, se encarga de determinar si la neurona se activa o no. Si se activa, la información continuará propagándose por la red [34].

4.4. Redes neuronales convolucionales

Las redes neuronales convolucionales o CNN (*Convolutional Neural Networks*) son un tipo de arquitectura de RNA especialmente diseñada para el procesamiento y análisis de imágenes. Las RNA presentaban muchas limitaciones a la hora de trabajar con imágenes completas, por eso es que las CNN están diseñadas para que la información de entrada corresponda a imágenes representadas como vectores organizados espacialmente. Este enfoque permite simplificar la estructura de la red y disminuir su complejidad. A diferencia de las redes neuronales tradicionales, las neuronas de una CNN están organizadas tridimensionalmente, considerando el ancho, la altura y la profundidad [37].

4.4.1. Arquitectura básica de una CNN

De acuerdo con Sakib [39], una CNN típica incluye tres bloques principales: convolución, activación, y pooling.

- Capa de convolución: su función principal es extraer automáticamente características relevantes de las imágenes de entrada, permitiendo a la red identificar patrones visuales. Aplica pequeños filtros (kernels) deslizantes que detectan rasgos locales como líneas, bordes o texturas. Al desplazarse por toda la imagen, el filtro realiza una operación que consiste en multiplicar sus valores por los píxeles de cada posición y sumar los resultados (convolución). Estos filtros no son predefinidos, sino que aprenden automáticamente a detectar distintos patrones durante el proceso de entrenamiento, permitiendo a la red identificar patrones cada vez más complejos.

- Capa de activación: después de la operación de convolución, la capa de activación decide qué valores deben mantenerse y cuáles descartar. Para ello utilizan funciones matemáticas también conocidas como funciones de activación, una de las más utilizadas es la ReLU (*Rectified Linear Unit*), esta permite pasar directamente los valores positivos y convierte en cero los negativos, mejorando el rendimiento del modelo y acelerando el proceso de aprendizaje. Se define, por tanto, como

$$f(x) = \max(0, x)$$

  donde x es la entrada de una neurona [40].

- Capa de *pooling*: esta capa toma la información ya procesada y la resume, reduciendo su tamaño y manteniendo la información más importante. Gracias a esta reducción, la red procesa la información de forma más rápida sin centrarse en detalles irrelevantes. El tipo más utilizado es el *max pooling*, una técnica de submuestreo no lineal que reduce la resolución de los mapas de activación al seleccionar el valor máximo en pequeñas

regiones [41]. Dicho de forma más sencilla, divide la imagen en regiones (por ejemplo, de 2x2 píxeles) y selecciona el valor máximo dentro de ese bloque.

Según el artículo [38], además de las capas de convolución, activación y *pooling*, las redes neuronales convolucionales incorporan una capa completamente conectada (*fully connected layer*). Esta capa actúa como una etapa final de procesamiento, donde se combinan las características extraídas previamente para generar la salida final del modelo. Tal y como indica su nombre, en esta capa cada neurona está conectada a todas las neuronas de la capa anterior. Se comporta como una red neuronal tradicional, aplicando la suma ponderada de las entradas, seguida de una función de activación para producir los resultados finales.

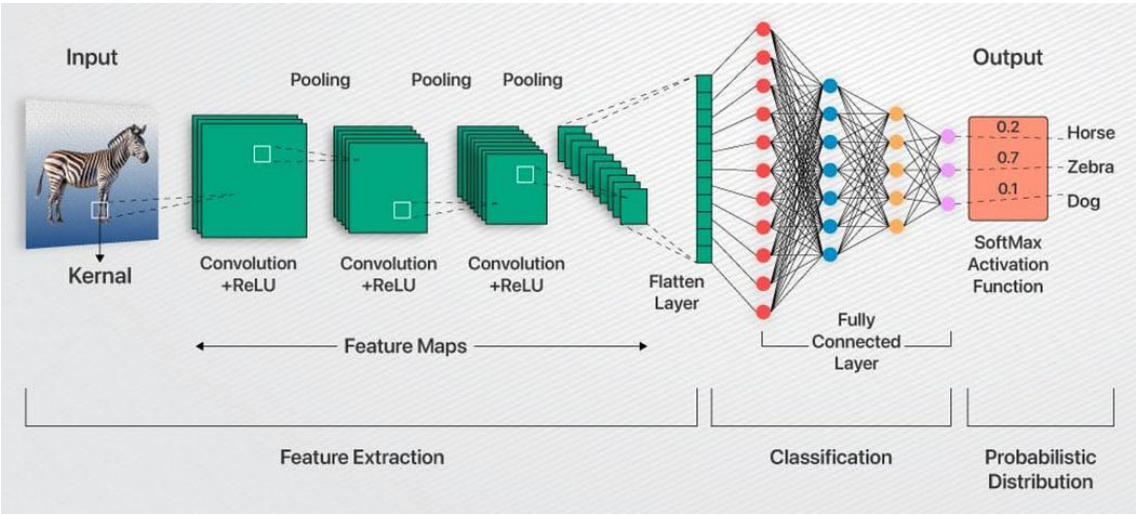

Figura 4.3. Estructura en capas de una red neuronal convolucional. Extraído de R. Singh, "Decoding CNNs: A Beginner's Guide to Convolutional Neural Networks and their Applications," Medium, 4 feb. 2025.

En la Figura 4.3 se puede observar de forma gráfica la estructura general de una CNN básica, identificando las diferentes capas desde la imagen de entrada hasta la capa completamente conectada.

5

Desarrollo

Este capítulo recoge en detalle todo el proceso llevado a cabo en el laboratorio para el entrenamiento del sistema de reconocimiento de iris de personas en movimiento, además de una descripción completa del sistema y la red neuronal YOLOX.

5.1. Tecnologías y herramientas utilizadas

5.1.1. Python

Python fue utilizado para el desarrollo de *scripts* de entrenamiento. Se trata de un lenguaje de programación de código abierto, interpretado e interactivo, ampliamente utilizado en el ámbito tecnológico por disponer de una sintaxis simple y clara, con la posibilidad de utilizarse en diferentes plataformas [35].

5.1.2. Pytorch

Framework de *deep learning* que permite definir y entrenar redes neuronales de manera flexible y eficientes, proporcionando herramientas avanzadas para trabajar con GPUs.

5.1.3. Roboflow

Roboflow es una plataforma en línea diseñada para facilitar todo el proceso de desarrollo de modelos de visión artificial, especialmente aquellos basados en CNN. Optimizando tareas como la detección de objetos, segmentación, clasificación de imágenes, entre otras [45].

5.1.4. AMD Vitis AI - Vivado

Se trata de una plataforma de desarrollo especializada en la implementación de aplicaciones de inteligencia artificial. Vitis AI proporciona un entorno completo para el desarrollo de modelos de aprendizaje profundo sobre hardware especializado, como FPGAs y SoCs desarrollados por AMD/Xilinx [36].

5.1.5. Sensor de imagen EMERAL 16MP

El sistema utiliza el sensor de imagen CMOS EMERAL de Teledyne e2v, el cual es capaz de capturar imágenes con una resolución de hasta 16 millones de píxeles, lo que permite obtener un alto nivel de detalle en cada captura [46].

5.1.6. Zynq UltraScale+ MPSoC

El procesador Zynq UltraScale+ de AMD/Xilinx se encuentra integrado en el sistema hardware empleado. Es un sistema en chip multiprocesador (MPSoC), es decir, un sistema con una arquitectura heterogénea, en la que se combinan núcleos ARM de alto rendimiento, GPU, núcleos ARM de tiempo real, y lógica programable [47].

5.2. Descripción del sistema

El sistema propuesto tiene como finalidad la detección de ojos e iris humanos en sujetos en movimiento, orientado a su posterior uso en procesos de identificación mediante reconocimiento de iris a distancia (*Iris At A Distance*, IAAD). Está diseñado para funcionar en entornos reales, donde el sujeto se desplaza de forma natural sin la necesidad de detenerse o interaccionar con el sistema, implementando técnicas de visión artificial, inteligencia artificial y procesamiento integrado.

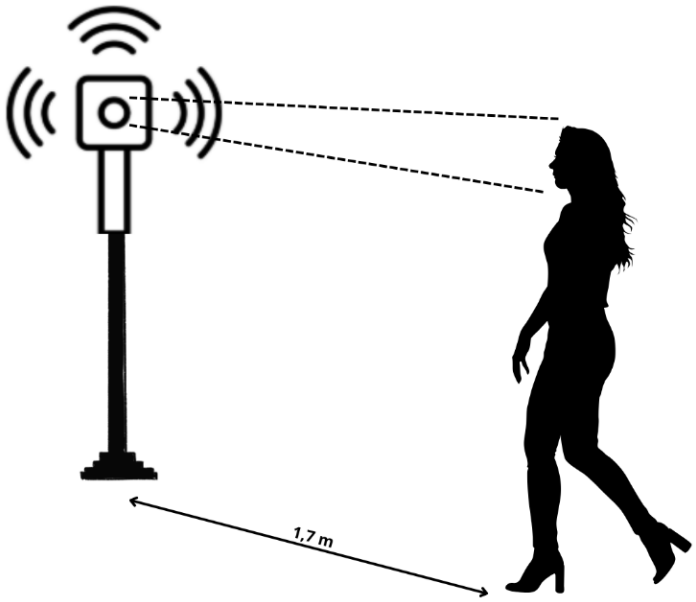

Figura 5.1. Esquema de la disposición del sistema en el escenario de aplicación. Elaboración propia.

El modelo consiste en una unidad de captura y procesamiento de imágenes que se encuentra fija en un soporte vertical, a una altura que garantiza una correcta captación de la región ocular (en versiones más robustas, el sistema consta de dos cámaras, para asegurar cubrir un mayor rango en alturas). Permite capturar imágenes a una distancia de 1,7 m aproximadamente, donde el sujeto camina a una velocidad media de 1-2 m por segundo. En la Figura 5.1 se presenta un

esquema de la disposición del sistema donde se puede observar el escenario de aplicación.

5.2.1. Arquitectura hardware del sistema para captura y procesamiento

Para la adquisición de imágenes se ha utilizado un sensor de imagen digital CMOS de 16 megapíxeles, en concreto un Teledyne e2v EMERALD, capaz de capturar 47 fotogramas por segundo. El tamaño de las imágenes de ojo detectadas será de 640 x 480 píxeles. El sensor se monta con una lente de 100mm. La iluminación la proporciona una fuente de luz infrarroja de 51 W con LED de alta potencia, que se dispara únicamente durante la captura del sensor. En el sistema se emplea iluminación infrarroja cercana en lugar de luz visible, ya que permite obtener imágenes del iris con mayor detalle. Además, al ser una luz no perceptible para el ojo humano, no resulta incómoda para el usuario (lo cual es beneficioso pero también perjudicial, pues el ojo no reacciona y puede dañarse). Por otro lado, la luz visible puede provocar reflejos en la córnea, lo que dificulta la correcta detección del iris y afecta a la calidad de las imágenes.

En este tipo de sistemas, el tiempo de exposición y la profundidad de campo juegan un papel fundamental, que determina la calidad de las imágenes capturadas. El tiempo de exposición lo establecemos nosotros, y deberá ser pequeño para que la imagen capturada del ojo, en una persona que se está moviendo, no se vea afectado por ruido de movimiento. En nuestro caso se usa un tiempo de 1 ms. Por otra parte, la profundidad de campo se define como el rango de distancia en el que el ojo se encuentra enfocado con nitidez. Debido a lo reducido del tiempo de exposición entra poca iluminación al sensor, y esto genera uno de los grandes inconvenientes del sistema: la limitación en la profundidad de campo (estimada en unos 10 - 15 cm). Para asegurar que se obtenga al menos una

imagen del iris correctamente enfocada, es necesario capturar una alta cantidad de fotogramas por segundo (el máximo que permite el sensor son 47 fps). Esto hace que el sistema pueda llegar a generar más de 250 imágenes de ojos por cada secuencia de paso, lo que implica una gran cantidad de datos que el sistema debe procesar, almacenar y analizar, esto puede aumentar el tiempo de procesamiento, afectando a la eficiencia del proyecto. En la plataforma desarrollada no se implementa ningún mecanismo de filtrado para descartar las imágenes desenfocadas. No obstante, este inconveniente podría mitigarse añadiendo un filtro de calidad de imagen, similar al propuesto en [5], que permite descartar automáticamente aquellas capturas que no cumplan con un nivel mínimo de nitidez.

El sistema está desarrollado sobre el micromódulo comercial TE0820-03-4DE21FA fabricado por Trenz Electronic, que integra en su interior el procesador MPSoc Zynq UntraScale+ ZU4EV perteneciente a la familia AMD/Xilinx UntraScale+. Como se ha comentado, este tipo de dispositivos cuentan, principalmente, con una parte de procesamiento (PS), formada fundamentalmente por un ARM Cortex-A53, y una parte programable (PL), implementada con una FPGA. Con estos recursos, este procesador heterogéneo ofrece un alto rendimiento computacional para procesar imágenes y gran flexibilidad en la FPGA, que facilita la integración de unidades o *cores* específicos [44].

Como núcleo más relevante, en la FPGA del MPSoC se integra una DPU (*Deep Processing Unit*). Se trata de un bloque de propiedad intelectual (*IP core*) desarrollado por AMD/Xilinx específicamente para acelerar las operaciones asociadas a redes neuronales profundas. La DPU se implementa en la parte lógica de la FPGA mediante el uso de las herramientas de Vivado y Vitis-AI, que

permiten generar y cargar la configuración necesaria en el dispositivo. Los detalles de este proceso se describen en el apartado 5.5.

La Figura 5.2 muestra la unidad de captura y procesamiento utilizada para este proyecto, en la que se distingue la fuente de iluminación, la placa del sensor EMERALD y el micromódulo TE0820-03-4DE21FA.

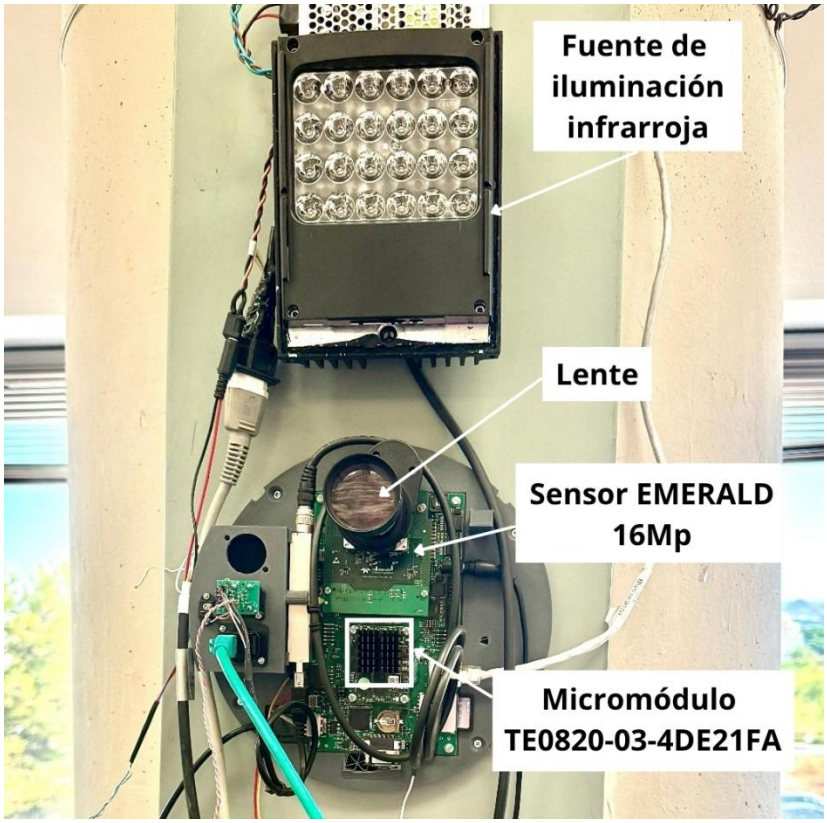

Figura 5.2. Unidad de captura y procesamiento del sistema para la detección de iris. Elaboración propia.

5.2.2. YOLOX para la detección ocular

La detección de ojos en personas en movimiento es una etapa clave dentro del sistema biométrico desarrollado. Para ello, se emplea una CNN basada en la arquitectura YOLO (*You Only Look Once*), un algoritmo de detección de objetos en tiempo real, publicado en julio de 2021 por Megvii Technology [53]. Según se

expone en [48], la detección de objetos se plantea como un problema de regresión en lugar de una tarea de clasificación, separando espacialmente las regiones rectangulares que delimitan cada objeto y asociándole una probabilidad de clase. En concreto se ha optado por emplear YOLOX (*You Only Look eXeceeding*) , debido a su arquitectura moderna, que supera a versiones anteriores de la serie YOLO. Esta red reduce la complejidad computacional gracias a su diseño de ancla-libre. En versiones anteriores, la detección dependía de anclas predefinidas para proponer las regiones de interés. En el caso de YOLOX, la estructura aprende a estimar de forma directa las coordenadas de las cajas delimitadoras y las clases de los objetos, eliminando la dependencia de anclas o formas preestablecidas, sin necesidad de definir previamente los tamaños de los objetos a detectar, lo que simplifica considerablemente la arquitectura y mejora la generalización de diferentes conjuntos de datos [49].

Otro aspecto a destacar del modelo, es el uso de la estrategia SimOTA (*Simple Optional Transport Assigment*), para conseguir una asignación de etiquetas más eficaz [43]. Además, YOLOX destaca por utilizar una cabeza desacoplada, lo que permite separar las tareas de clasificación (qué objeto hay en cada caja) y localización (dónde está ese objeto), mejorando la precisión de la CNN.

Como se ha comentado, en la versión anterior del sistema se utiliza la arquitectura YOLOv3 [5]. En la Figura 5.3 se ilustran las diferencias entre la versión YOLOv3 y YOLOX. En la parte superior se observa la cabeza acoplada (*Coupled Head*) utilizada por las versiones anteriores de YOLO. En este diseño, tras una serie de convoluciones, una única rama genera de forma simultánea las predicciones de clasificación (Cls), regresión de las cajas (Reg) y confianza del objeto (Obj). Esto puede producir que la precisión del modelo disminuya, como se comprobó usando

el modelo Vainilla COCO en [43], demostrando que con el uso de una cabeza desacoplada aumentaba el AP (*Average Precision*).

Por lo contrario, en la parte inferior de la figura se muestra la cabeza desacoplada. En este caso, después de aplicar convoluciones de diferentes tamaños, la arquitectura se divide en dos ramas independientes, la rama de clasificación (Cls) y la rama de regresión (Reg + IoU). Este diseño desacoplado permite que cada rama aprenda características especializadas para su tarea, mejorando la eficiencia del entrenamiento.

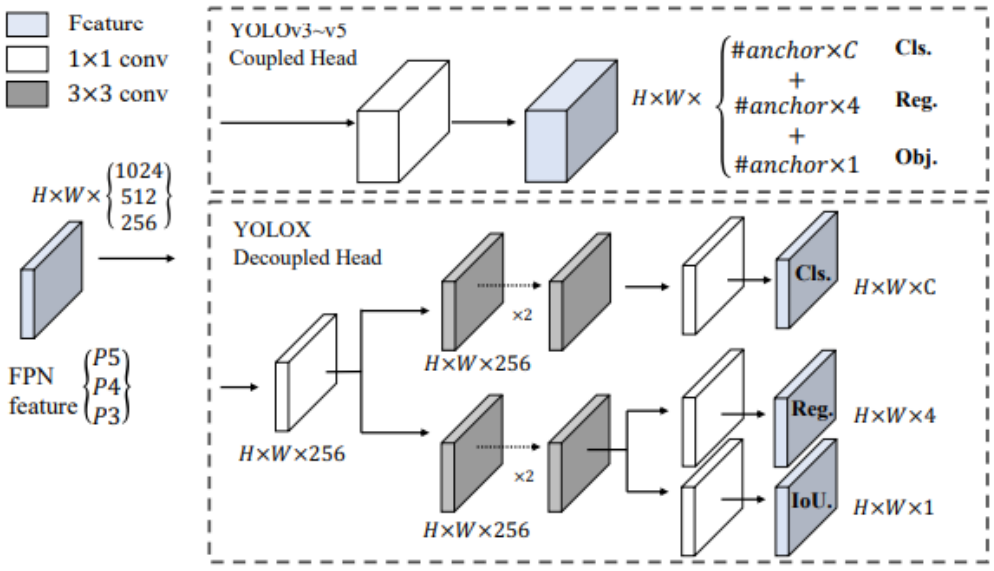

Figura 5.3. Ilustración sobre las diferencias entre el diseño de YOLOv3 basado en cabeza acoplada y la configuración de cabeza desacoplada que presenta YOLOX. Extraído de Z. Ge, S. Liu, F. Wang, Z. Li y J. Sun, "YOLOX: Exceeding YOLO series in 2021," arXiv preprint, arXiv:2107.08430, 2021.

5.2.2.1. Arquitectura interna de YOLOX

YOLOX se estructura en tres bloques principales [48,50]: *Backbone, neck y head.*

- *Backbone* o columna vertebral de la red. Se encarga de extraer las características relevantes de la imagen de entrada. Esta arquitectura utiliza una CSP Darknet53 (red neuronal convolucional preentrenada con un conjunto de datos COCO).

- *Neck*, actúa como puente entre las otras capas. Compuesto por una red piramidal de características (*Feature Pyramid Network*, FPN), la cual genera mapas de características a diferentes escalas. Se ocupa de fusionar los mapas obtenidos por el *backbone*, introduciéndolos como entradas en la cabeza a tres escalas diferentes (1024, 512 y 256).

- *Head*, es la capa final de la red, se trata de una cabeza desacoplada como se ha explicado anteriormente. Es la responsable de realizar las predicciones de los rectángulos delimitadores y las clases de los objetos (ojos e iris).

5.3. Creación de una base de datos propia

Una vez decidida una estructura de CNN, el siguiente paso es entrenarla con patrones de los que se disponga del resultado (*ground truth*). En [5], se utilizó para este entrenamiento de la red YOLOv3 la base de datos CASIA. Formada por capturas de sujetos en posición estática, perfectamente frontales e iluminados, el inconveniente de esta base de datos es que dicho conjunto no representaba adecuadamente las variaciones de postura, movimiento y luz, necesarias para el escenario dinámico que se deseaba abordar. Por ello, en el marco de este trabajo se decidió crear una base de datos específica. En lugar de etiquetar los ojos, como se había utilizado en [5], en este caso se decidió etiquetar ojos e irises. Esto permitiría robustecer la detección: un iris válido sería aquel que estuviera dentro de un ojo. Además, los irises podrían ser la entrada a un sistema que segmentara

el iris. La base de datos se crea con la estudiante de prácticas Ana Barrios, también de Ingeniería de la Salud.

El proceso comenzó con la recopilación y organización de una base de datos que contaba con más de 7000 imágenes que habían sido capturadas de secuencias de videos de personas en movimiento con el propio sistema. En la mayoría de ellas se muestra el rostro completo de la persona, pero en muchas otras solo aparece una parte del rostro debido al movimiento de acercamiento del individuo a la cámara. Este proceso era indispensable para realizar un entrenamiento correcto de la CNN, debido a que no se disponía de ninguna base de datos compuesta por fotogramas de personas en movimiento.

Se tuvieron en cuenta todo tipo de perfiles con diferentes características físicas para conseguir que el sistema aspire a ser robusto y generalizable. Considerando variaciones como personas de distintos rangos de edad, sexo, factores externos como el uso de gafas, lentillas, maquillaje ocular. Esta diversidad en los datos ha sido esencial para entrenar un modelo más preciso. Para garantizar que el sistema solo entrenara a partir de ejemplos válidos, se descartaron manualmente todas las imágenes que no cumplían con un mínimo de enfoque y nitidez, y de esta forma evaluar si puede diferenciar las imágenes desenfocadas.

Uno de los primeros pasos esenciales en el entrenamiento del modelo fue el etiquetado de las imágenes, proceso mediante el cual se marcan de manera manual o automática las regiones de interés de cada una de ellas. En este caso el objetivo era etiquetar los ojos e irises de todas las imágenes de la base de datos. Para realizar esta tarea se empleó la plataforma Roboflow, una herramienta online utilizada para implementar aplicaciones de visión artificial [45]. Dentro de los

tipos de proyectos que podemos identificar en la aplicación encontramos: clasificación de etiqueta única, clasificación de múltiples etiquetas, segmentación de instancias, segmentación semántica, detección de puntos clase y detección de objetos [12]. El proyecto llamado "Detección de iris" se creó con el modo detección de objetos, dado que el propósito era localizar de forma automática la posición de los ojos e irises dentro de la imagen. Esta categoría de proyecto permite generar anotaciones mediante *bounding boxes* (cuadros delimitadores), estos encuadran la región de la imagen que queremos detectar. Según el artículo [13], la interacción con *bounding box* es una de las más naturales y cómodas para el usuario. Esta debe ser usada de tal manera que la segmentación deseada debe quedar lo suficientemente cerca de cada uno de los lados del cuadrado (ver Figura 5.5).

Dentro de la aplicación se pueden crear diferentes categorías para identificar las *bounding boxes.* En nuestro caso se definieron las etiquetas "eyes" e "iris". Como se muestra en la Figura 5.4, estas etiquetas son denominadas en la aplicación como clases, y están ubicadas en el apartado de anotaciones. En la imagen se muestran las dos etiquetas con el número de veces que han sido empleadas en esa imagen. En el caso que se ejemplifica (Figura 5.5) se aprecia el rostro de la persona prácticamente completo, por lo que el número de anotaciones es 4.

Figura 5.4. Clases creadas en el proyecto en Roboflow. Elaboración propia.

Inicialmente se realizó un etiquetado manual de aproximadamente 3000 imágenes, marcando con precisión la ubicación de ojos e irises, como se muestra de color verde y morado en la Figura 5.5. Esta primera fase permitió generar un conjunto inicial de datos, que posteriormente se utilizarían para entrenar al modelo y que él mismo pudiera etiquetar otro lote de imágenes de la base de datos. El etiquetado de este segundo lote deberá ser revisado.

Figura 5.5. Imagen etiquetada mediante *bounding box* en Roboflow. Elaboración propia.

Para optimizar el conjunto etiquetado, Roboflow permite crear nuevas versiones a las que se les pueden añadir *augmentations* (aumentos), trasformaciones artificiales que se aplican al conjunto de datos para aumentar su variedad sin necesidad de recopilar nuevas imágenes. Las ampliaciones que proporciona la aplicación se pueden dividir en dos tipos: a nivel de imagen y a nivel de cuadro delimitador (véase Figura 5.6).

Dentro de las transformaciones que ofrece la aplicación las utilizadas fueron rotación entre -15°/+15°, lo que significa que cada imagen puede rotar aleatoriamente dentro de ese rango. También se añadió la opción de corte ±10° horizontal y vertical, añadiendo la condición que por cada imagen original se generan 3 versiones aumentadas.

Figura 5.6. Tipos de aumentos que proporciona Roboflow. Elaboración propia.

Una vez finalizado el proceso de etiquetado manual y añadidos los aumentos en la plataforma Roboflow, se procede a exportar el conjunto de datos en el formato COCO JSON, ampliamente utilizado para almacenar información sobre las anotaciones de las imágenes. Este formato incluye las coordenadas de las cajas delimitadoras junto otros datos relevantes que son utilizados por la CNN para interpretar correctamente las regiones de interés. Este proceso se llevó a cabo en reiteradas ocasiones debido al gran volumen de imágenes disponibles.

Como se ha comentado, con la red ya entrenada se procedió a etiquetar un segundo conjunto de datos. Aunque la red neuronal realiza un etiquetado automático, fue necesaria una supervisión para revisar y corregir los cuadros

delimitadores de todo este segundo conjunto de datos. Durante las primeras etapas, el modelo presentó algunas dificultades al identificar correctamente ojos e iris. Así, en ocasiones, confundía regiones de la barba (en el caso de los hombres) o el cabello (Figura 5.7). Este tipo de errores fueron corregidos manualmente, para evitar falsos positivos a la hora de utilizar el sistema en tiempo real. Se ha sido muy cuidadoso con estos aspectos, pues la calidad del entrenamiento de una red neuronal depende en gran medida de la validez y precisión de los datos utilizados. Disponer de un conjunto de datos correctamente etiquetados es esencial para que el modelo aprenda a distinguir las características de interés. De lo contrario, la presencia de anotaciones incorrectas puede comprometer la fiabilidad del sistema.

Figura 5.7. Ejemplo de fallo en el etiquetado automático, la red neuronal identifica iris en zonas no oculares. Elaboración propia.

Siguiendo con la línea de la preparación del conjunto de datos para el entrenamiento, otro preprocesamiento que se realizó con la herramienta de Roboflow, fue redimensionar todas las imágenes a un tamaño 256x256 píxeles,

esta operación de *resize* es común en procesos de entrenamiento de CNN. Reducir la resolución permite disminuir el coste computacional y la memoria necesaria durante el entrenamiento. En [51] se demuestra que la reducción de la resolución no afecta a la detección de ojos, consiguiendo un 100% de detecciones positivas. La base de datos está disponible en Roboflow (https://app.roboflow.com/practicasana/eye-iris-dataset-dndqh/1).

5.4. Fase de entrenamiento de la CNN

Una vez completado el proceso de preparación de la base de datos, se procedió a la fase de entrenamiento del modelo de detección de iris. Esta etapa se desarrolló utilizando PyTorch, una biblioteca ampliamente utilizada en el ámbito del aprendizaje profundo. El código empleado para el entrenamiento del modelo se obtuvo del repositorio de YOLOX en GitHub [52].

Antes de iniciar el entrenamiento, fue necesario realizar una fase de preparación que incluyó la configuración del entorno de trabajo. Comenzando por la descarga del repositorio en GitHub y la instalación de las dependencias necesarias. Posteriormente, se creó un archivo de configuración (yolox_voc) donde se definieron los parámetros clave para el entrenamiento, como el número de clases, el número de épocas, el tamaño de entrada, el tamaño del batch, la tasa de aprendizaje, así como otros ajustes necesarios para la correcta ejecución del modelo. Asimismo, se organizaron las carpetas necesarias para almacenar los conjuntos de datos, configuraciones, resultados del entrenamiento y pesos preentrenados.

Los parámetros utilizados para el entrenamiento fueron los siguientes:

- Número de clases: 2

- Número de épocas: 300
- Tamaño de entrada: 256 x 256
- Tamaño del batch: 16

Una vez completado el entrenamiento, este se puede evaluar con el parámetro de precisión media (AP), el cual permite medir el rendimiento del modelo en la detección de objetos. A medida que se avanzaba en la iteraciones de entrenamiento, este valor fue incrementando progresivamente, lo que indica que el modelo iba mejorando su capacidad para detectar y segmentar con precisión los ojos e iris en las imágenes. Finalmente, tras completar el proceso de entrenamiento, el modelo alcanzó un AP del **71,04%**, un resultado que puede considerarse satisfactorio teniendo en cuenta la complejidad del problema abordado.

Es importante destacar que al tratarse de regiones de interés muy pequeñas y con gran variabilidad, como el iris, resulta más difícil obtener valores de AP extremadamente altos en comparación con objetos de mayor tamaño. Aun así, el 71,04% alcanzado demuestra un rendimiento sólido y la viabilidad del enfoque propuesto para la detección ocular y segmentación del iris en movimiento.

5.5. Cuantización del modelo

Tras el entrenamiento de la red YOLOX, el modelo obtenido se encuentra en formato de punto flotante de 32 bits. Este formato es el estándar durante la fase de entrenamiento, y permite que el modelo aprenda ajustando los pesos con valores muy precisos. Sin embargo, la DPU que se sintetiza en la PL del MPSoC no admite esta precisión, estando optimizada para trabajar con valores enteros de

8 bits (requieren menos memoria y pueden ser procesados de forma más eficientes en hardware).

Por este motivo, antes de implementar el modelo en el hardware es necesario realizar un proceso de cuantización, que permite convertir estos valores aritméticos de punto flotante de 32 bits a valores de punto fijo de 8 bits, lo que reduce el tamaño del modelo y lo hace compatible con la DPU.

Para ellos se han evaluado dos estrategias de cuantización con el objetivo de comparar cuál ofrecía mejores resultados de precisión y rendimiento.

5.5.1. *Post-Training Quantization* (PQT)

La cuantización posterior al entrenamiento (PQT) [56], se trata de una técnica que se aplica una vez finalizado el entrenamiento del modelo con valores en punto flotante. En este método el modelo se convierte directamente a un formato de menor precisión, los pesos y las activaciones se evalúan en un conjunto de datos representativo para determinar el rango de valores que toman estos parámetros.

La principal ventaja que presenta este método es su rapidez, ya que no requiere modificar el entrenamiento original ni realizar un proceso de ajuste adicional. Sin embargo, esta técnica puede ocasionar una pérdida de precisión en ciertos modelos.

5.5.2. *Quantization-Aware Training* (QAT)

La cuantización durante el entrenamiento (QAT) [56] es una técnica que optimiza un modelo previamente entrenado incorporando las operaciones de cuantización (escalado, recorte y redondeo) durante un segundo proceso de entrenamiento. Esto

permite que la red neuronal aprenda a trabajar con valores de menor precisión desde la fase de entrenamiento, reduciendo la pérdida de exactitud que normalmente ocurre al cuantizar un modelo ya entrenado (PQT).

El método QAT ofrece como principal ventaja la capacidad de mantener una alta precisión en el modelo final, debido a que este aprende a compensar las pérdidas de precisión asociadas al uso de valores de menor resolución. No obstante, este método también presenta algún inconveniente. Requiere realizar un nuevo entrenamiento, lo que implica mayor tiempo computacional.

5.5.3. Análisis comparativo de PQT y QAT

Con el objetivo de evaluar qué técnica de cuantización ofrece mejor rendimiento en la implementación del modelo YOLOX sobre la plataforma MPSoC, se aplicaron ambos métodos al mismo modelo base previamente entrenado. Para ello se evaluó la precisión de cada modelo cuantizado utilizando el parámetro AP.

Los resultados obtenidos, mostrados en la Tabla 5.1, reflejan el impacto de cada técnica sobre la precisión media del modelo para las clases "ojo" e "iris":

Técnica de cuantización	Clase	AP	Clase	AP
PQT	Ojo	57,226	Iris	53,100
QAT	Ojo	65,680	Iris	69,080

Tabla 5.1. Comparación de los resultados obtenidos tras la cuantización del modelo mediante las técnicas PQT y QAT. Elaboración propia.

Como se puede observar, el modelo cuantizado mediante QAT logra un rendimiento significativamente superior respecto al obtenido con PQT en ambas clases. Esto se debe a que en el caso del QAT, el modelo es reentrenado considerando las limitaciones introducidas por la cuantización, lo que le permite adaptarse mejor a estos cambios y conservar una mayor precisión.

Aunque PQT es una técnica más rápida y sencilla de aplicar, su impacto sobre el rendimiento del modelo es más notable, especialmente en tareas más sensibles como la segmentación del iris.

Después de evaluar ambas técnicas de cuantización, se observa una diferencia significativa en el rendimiento. Finalmente, se entiende que QAT es la técnica más adecuada para este proyecto, con un AP del **69,08%**, dado que la tarea de reconocimiento de iris requiere mantener una alta precisión en las características extraídas y minimizar la pérdida de información. El hecho de que el entrenamiento sea más lento no es problema en nuestra aplicación.

Comparando este valor con el mostrado en el apartado 5.4 (AP tras el entrenamiento 71,04%), podemos comprobar una pequeña reducción del **1,96%** del AP tras el proceso de cuantización. Esta disminución era esperable dado que el cuantizado transforma los parámetros del modelo a formatos más compactos para optimizar recursos, lo cual era necesario para la implementación del modelo en la FPGA.

5.6. Implementación dentro del MPSoC

En la fase de implementación dentro del MPSoC, se ha empleado el IP DPU (*Deep Learning Processing Unit*) de AMD/Xilinx para acelerar la ejecución del

modelo de red neuronal optimizado. Esta IP permite desplegar modelos previamente cuantizados sobre la FPGA integrada, aprovechando la capacidad de procesamiento paralelo del hardware para ejecutar inferencias de forma mucho más eficientes que en la CPU del sistema. La DPU está específicamente diseñada para ejecutar operaciones típicas de *deep learning*, como convoluciones, activaciones y normalizaciones, de forma altamente eficiente. Se implementa como un IP core configurable dentro del dispositivo MPSoC con FPGA mediante la herramienta de Vivado [55].

5.6.1. Implementación mediante Vivado

La implementación del sistema se llevó a cabo empleando Vivado Desing Suite, herramienta de diseño hardware para FPGAs y SoCs de AMD/Xilinx, que permitió integrar el IP DPU dentro de la arquitectura del MPSoC, concretamente el Zynq UltraScale+ de Xilinx.

En este proyecto, Vivado desempeña un papel clave al permitir la creación de un diseño jerárquico de bloques mediante la herramienta IP Integrator, donde se presenta gráficamente cómo se conectan los distintos módulos del sistema dentro del MPSoC. A través de este entorno es posible construir una arquitectura de hardware personalizada que aprovecha la lógica programable de la FPGA para ejecutar en paralelo modelos de redes neuronales, en este caso YOLOX.

En la Figura 5.8 se presenta el diagrama de bloques del sistema para la detección de ojos e iris basado en CNN. El diagrama representa la arquitectura implementada en el MPSoC Zynq UltraScale+, donde se integran módulos de captura de imagen, preprocesamiento, transferencia de datos y aceleración mediante la DPU.

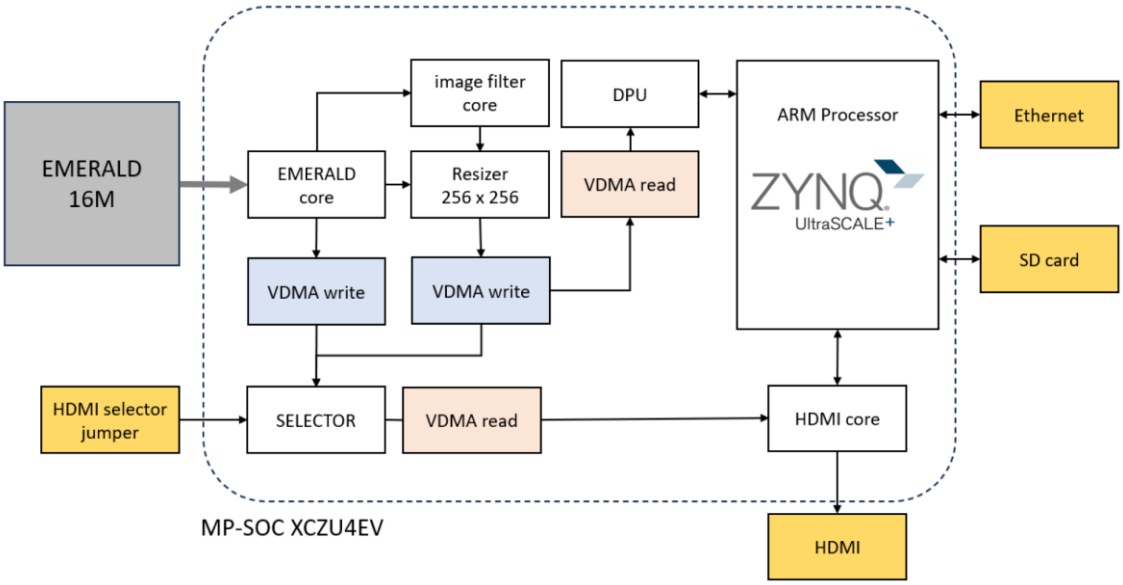

Figura 5.8. Diagrama de bloques del sistema para la detección de ojos e iris basado en CNN. Extraído de C. A. Ruiz-Beltrán, A. Romero-Garcés, M. González-García, R. Marfil y A. Bandera, "FPGA-Based CNN for Eye Detection in an Iris Recognition at a Distance System," Electronics, vol. 12, p. 4713, 2023.

El flujo comienza con la cámara EMERALD 16M encargada de capturar las imágenes en tiempo real, las cuales serán almacenadas posteriormente en la memoria DDR, y cuya señal es gestionada por el bloque EMERALD core. A continuación, la imagen será sometida a un filtro digital y redimensionada a una resolución de 256x256 píxeles, que la adapta al formato requerido por la red neuronal desplegada en la DPU. Para la transferencia eficiente entre bloques y memoria se utilizan controladores VDMA (*Video Direct Memory Access*), que permiten almacenar y recuperar datos de manera flexible. La DPU es el núcleo encargado de ejecutar el modelo de inteligencia artificial, procesando las imágenes en paralelo y acelerando significativamente la inferencia respecto a la CPU. El procesador ARM integrado en el Zynq UltraScale+ coordina el funcionamiento general del sistema: recibe los resultados de la DPU y se encarga de enviarlos al Ethernet, guardarlos en la tarjeta SD o dirigirlos al bloque HDMI core para su visualización en pantalla. Finalmente, el HDMI selector jumper permite elegir si

la salida mostrada corresponde a la señal directa de la cámara o a la imagen procesada por el sistema.

5.6.2. Integración del modelo mediante Vitis AI

Para la carga del modelo de red neuronal a la FPGA, se emplea Vitis AI, una herramienta integral de Xilinx diseñada para optimizar y desplegar modelos de redes neuronales en plataformas adaptables [55]. Esta herramienta facilita la integración de modelos entrenados en *frameworks* como PyTorch, permitiendo su ejecución en dispositivos como FPGAs.

El proceso de carga del modelo en Vitis AI [55] se inicia con la conversión del modelo entrenado al formato .xmodel, mediante el uso del Vitis AI Compiler. Este archivo contiene una representación optimizada del modelo, adaptada para su ejecución en la DPU del dispositivo. Una vez generado el archivo .xmodel, se procede a su carga en el dispositivo utilizando la API de Vitis AI Runtime, que gestiona la ejecución de inferencias de manera eficiente.

Esta herramienta ha sido utilizada para transformar el modelo YOLOX entrenado al formato .xmodel, de modo que pueda ejecutarse de manera eficiente en la DPU, disminuyendo la carga sobre el procesador ARM.

6

Resultados y discusión

En el presente apartado se analizan los resultados obtenidos tras el entrenamiento e implementación del modelo.

Inicialmente se realizaron varias sesiones de entrenamiento. Parte del conjunto de datos utilizado había sido etiquetado por la propia CNN y revisado manualmente, aunque a simple vista las imágenes revisadas estuvieran bien etiquetadas, el resultado del entrenamiento mostraba los cuadros delimitadores en las zonas de interés correctas. El problema surgió cuando se hizo la prueba en tiempo real. El resultado se visualiza en un ordenador externo que se encuentra conectado al sistema. Se podía observar que algunos de los recortes no incluían la totalidad del iris, recortándolo lateral o inferiormente, como se ilustra en la figura 6.1. Esto

suponía una dificultad relevante, aumentando la probabilidad de errores o falsos positivos en la siguiente etapa de reconocimiento de iris.

Para resolver este problema, fue necesario llevar a cabo una revisión y reetiquetado del conjunto de datos en la plataforma de Roboflow, aumentado la zona de interés que señalaba al iris, tras comprobar que un etiquetado demasiado ajustado a los bordes del objeto a detectar generaba recortes incompletos.

Esto permite comprobar que el proceso de entrenamiento de la CNN es muy sensible incluso a errores mínimos, lo que resalta la importancia de disponer de una base de datos precisa y correctamente etiquetada.

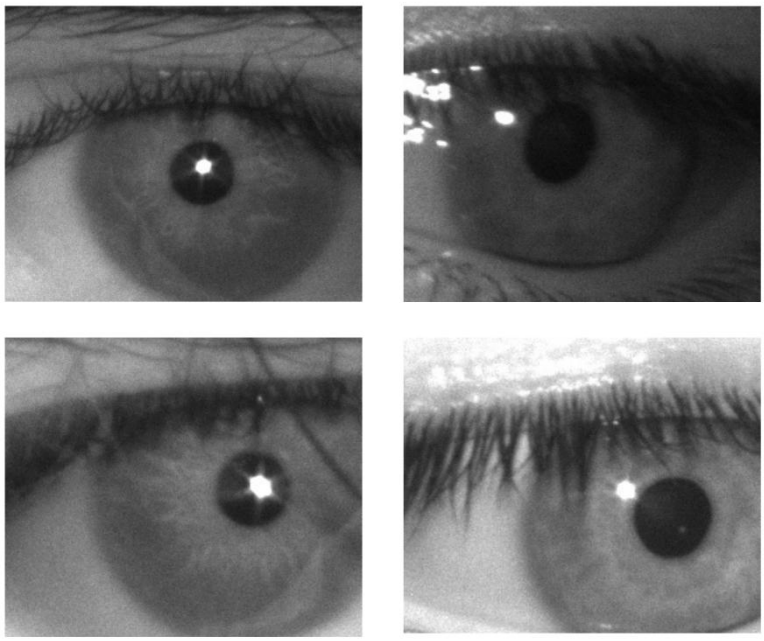

Figura 6.1. Resultados de los recortes de iris generados por el sistema de dos sujetos. Las imágenes de la parte superior muestran los iris correctamente segmentados. Por el contrario, las imágenes de la parte inferior omiten una pequeña zona del iris (lateral e inferiormente). Elaboración propia.

Tras corregir los errores y entrenar nuevamente la CNN con un conjunto de datos más preciso, se implementó el modelo convertido a formato .xmodel en la placa. Los resultados obtenidos tras esta segunda implementación fueron los esperados. El sistema era capaz de realizar la segmentación de ojos e iris de forma correcta, mostrando una gran precisión en la detección y delimitación de estas regiones de interés. La Figura 6.2 muestra cuatro imágenes de dos sujetos diferentes (sujeto 1: izquierda; sujeto 2: derecha). Las imágenes ubicadas en la zona superior muestran el recorte del ojo, mientras las que aparecen en la parte inferior presentan una correcta segmentación del iris.

Una vez conseguido el objetivo, se logró obtener recortes completos y adecuados, aptos para su posterior uso en tareas de reconocimiento de iris.

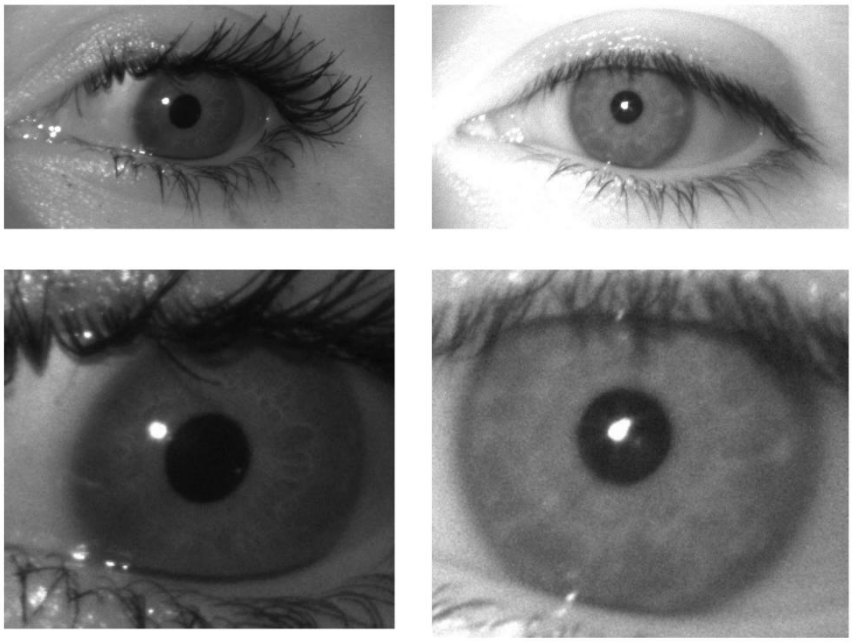

Figura 6.2. Resultados de los recortes de ojos e iris generados por el sistema en tiempo real de dos sujetos diferentes. Elaboración propia.

Sin embargo, el sistema continúa presentando ciertas limitaciones. Como se comentó anteriormente, es capaz de generar un flujo de entre 200 y 300 imágenes

de ojos en un intervalo de 2 a 3 segundos, que es el tiempo que tarda una persona en hacer el recorrido hacia en sensor. De éstas, aproximadamente el 98% de las capturas son imágenes no válidas (con baja resolución, borrosas o desenfocadas), debido principalmente a la corta profundidad de campo que presenta el sistema (10 - 15 cm). Con el objetivo de mitigar dicha limitación, se entrenó la CNN con un conjunto de imágenes nítidas, intentando que la red aprendiera a descartar los fotogramas de baja calidad generados durante la captura, mejorando así la precisión del modelo. Tras varias pruebas y análisis de los resultados obtenidos, se pudo comprobar que este requisito no se ha cumplido como se esperaba.

La Figura 6.3 ilustra una secuencia de imágenes capturadas en tiempo real correspondiente a un único sujeto. En la primera fase cuando el sujeto se encuentra a gran distancia (primeras capturas), las imágenes obtenidas son borrosas. A medida que este se aproxima al sensor, se observa una mejora progresiva en la nitidez de las capturas, que posteriormente vuelve a empeorar (debido a la limitada profundidad de campo). Las columnas de la izquierda muestran la segmentación de ojos, mientras que las de la derecha corresponden al iris.

Se puede apreciar que el sistema no logra filtrar correctamente los fotogramas con baja resolución, por lo que el entrenamiento con un conjunto de imágenes nítidas no ha sido suficiente para abordar esta limitación. Tal y como explican en [5], la idea de utilizar como entrada del sistema no solo la imagen capturada, sino que también una versión de la mismas con un filtro paso bajo, podría ayudar a descartar gran parte de las imágenes desenfocadas.

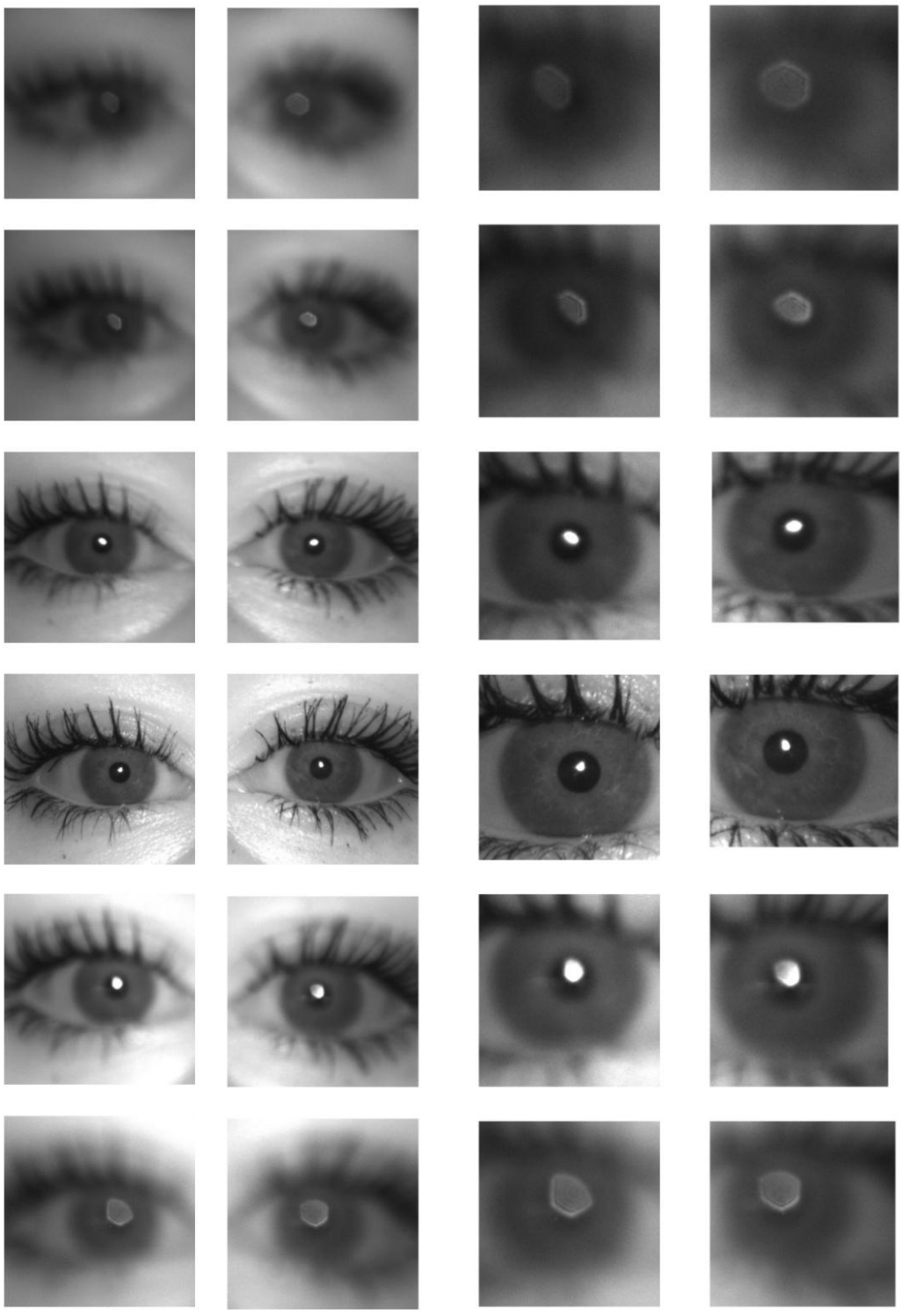

Figura 6.3. Secuencia de imágenes capturas por el sistema en tiempo real. Las dos primeras columnas corresponden a segmentaciones de ojos, mientras que las columnas situadas en la derecha muestran segmentaciones de iris. Elaboración propia.

Además, se evaluó el rendimiento del sistema en sujetos que utilizaban gafas. Aunque la presencia de este accesorio fue considerada en la base de datos empleada para el entrenamiento del modelo, durante las pruebas en tiempo real se observaron limitaciones asociadas a los reflejos generados en las lentes. Estos reflejos no siempre pueden ser controlados en condiciones reales y, en algunos casos, provocan imágenes con pérdida de información relevante en la región ocular, lo que impide un reconocimiento de iris fiable. En la Figura 6.4 se puede observar una secuencia de segmentaciones de ojos de un sujeto con gafas.

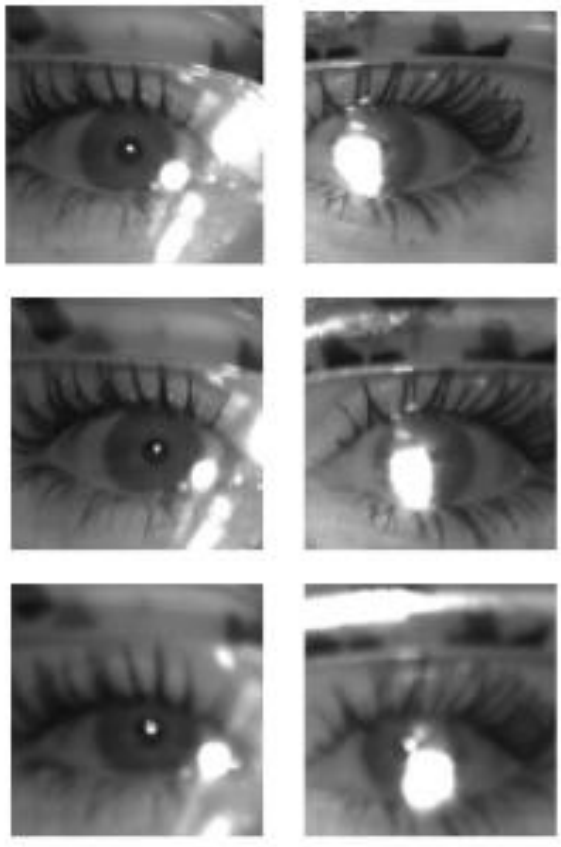

Figura 6.4. Secuencia de imágenes capturadas por el sistema en tiempo real de un sujeto con gafas, donde se observa la presencia de reflejos en las lentes. Elaboración propia.

7

Conclusiones y líneas futuras

El desarrollo de este Trabajo Fin de Grado ha permitido implementar y evaluar un sistema de detección de ojos e iris de personar en movimiento basado en CNN y desplegado sobre una plataforma hardware con FPGA. Para ello, se ha empleado la red neuronal YOLOX, entrenada y posteriormente adaptada mediante Vitis AI para su ejecución en la DPU integrada en la FPGA del MPSoC Zynq UltraScale+. Tras la preparación de la base de datos en Roboflow y el entrenamiento de la red neuronal, los resultados obtenidos en este proyecto han sido satisfactorios, alcanzando una precisión media del 71,04%, un resultado notable teniendo en cuenta que el iris es una región extremadamente pequeña y compleja de detectar dentro de la imagen. Este valor se redujo al 69,08% tras el proceso de cuantización mediante QAT, paso imprescindible para su ejecución

eficiente en la FPGA. Aun tras la reducción este valor sigue siendo competitivo y demuestra la viabilidad del sistema en condiciones reales.

El sistema propuesto ha demostrado su capacidad para detectar y segmentar con precisión tanto el iris como los ojos de los sujetos en movimiento, presentando los resultados en tiempo real y cumpliendo de manera satisfactoria el objetivo principal del proyecto. Este trabajo demuestra que es posible combinar modelos de visión artificial con hardware embebido especializado para obtener soluciones más compactas.

A pesar de los avances logrados, el sistema presenta ciertas limitaciones que abren la puerta a futuras mejoras. En particular, el entrenamiento con una base de datos compuesta por imágenes nítidas no ha sido suficiente para que el sistema aprenda a descartar las capturas con baja calidad. La inclusión de este tipo de imágenes en el proceso de detección no solo puede afectar a la precisión del reconocimiento de iris, sino que también incrementa la carga de memoria y el tiempo de procesamiento al tratarse de un volumen elevado de datos.

Como línea futura de trabajo, sería conveniente incorporar un módulo de preprocesamiento que actúe como filtro de calidad de imagen, capaz de detectar y excluir automáticamente aquellas capturas que no cumplan con un nivel mínimo de nitidez. Esta estrategia no solo aumentaría la robustez del sistema, sino que además optimizaría el rendimiento del mismo, reduciendo la complejidad asociada al manejo de un gran número de imágenes.

Referencias

[1] J. Daugman and C. Downing, "Epigenetic randomness, complexity and singularity of human iris patterns," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 268, no. 1477, pp. 1737–1740, 2001.

[2] P. L. Kaufman y A. Alm, *Adler. Fisiología del ojo: Aplicación clínica*, 10ª ed. Elsevier España, 2003.

[3] J. Daugman, "Recognizing people by their iris patterns," *Information Security Technical Report*, vol. 3, no. 1, pp. 33–39, 1998, doi: 10.1016/s1363-4127(98)80016-2.

[4] J. Daugman, "How iris recognition works," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 21–30, 2004, doi: 10.1109/TCSVT.2003.818350.

[5] C. A. Ruiz-Beltrán, A. Romero-Garcés, M. González-García, R. Marfil y A. Bandera, "FPGA-Based CNN for Eye Detection in an Iris Recognition at a Distance System," *Electronics*, vol. 12, p. 4713, 2023, doi: 10.3390/electronics12224713.

[6] F. Serratosa, *La biometría para la identificación de las personas*. Universitat Oberta de Catalunya, 2008.

[7] A. Bertillon, *Identification anthropométrique: Instructions signalétiques*, 1885.

[8] J. A. C. Osorio, F. A. M. Aguirre y J. A. M. Escobar, "Sistemas de seguridad basados en biometría," *Scientia et Technica*, vol. 17, no. 46, pp. 98–102, 2010.

[9] F. Serratosa y A. S. Ribalta, *Biometría*. Universitat Oberta de Catalunya, 2014.

[10] F. Galton, *Fingerprints*. London: Macmillan and Co., 1892.

[11] R. Sanchiz Redondo, "Segmentación de iris mediante contornos activos," 2011.

[12] "Create a Project | Roboflow Docs." Roboflow Docs. [En línea]. Disponible en: https://docs.roboflow.com/datasets/create-a-project.

[13] V. Lempitsky, P. Kohli, C. Rother y T. Sharp, "Image segmentation with a bounding box prior," en *Proc. 2009 IEEE 12th International Conference on Computer Vision*, pp. 277–284, 2009.

[14] R. B. Hill, "Patente de EE. UU. n.º 4.109.237," Washington, DC: Oficina de Patentes y Marcas de EE. UU., 1978.

[15] L. Flom y A. Safir, "Sistema de reconocimiento de iris," Patente estadounidense n.º 4.641.349, 1987. [En línea]. Disponible en: http://www.google.com/patents/US4641349

[16] J. Daugman, "Iris recognition at airports and border crossings," en *Encyclopedia of Biometrics*, pp. 998–1004. Springer, Boston, MA, 2015.

[17] M. Piña, *Vía oftálmica: Anatomía y fisiología ocular. Administración de medicamentos: teoría y práctica*, p. 99, 1994.

[18] F. L. Villar, *Anatomía ocular. Oftalmología*, pp. 1–9, 2000.

[19] A. Lens, S. C. Nemeth y J. K. Ledford, *Anatomía y fisiología ocular*. Slack Incorporated, 2008.

[20] J. V. Forrester, A. D. Dick, P. G. McMenamin, F. Roberts y E. Pearlman, *The Eye: Basic Sciences in Practice*. Elsevier Health Sciences, 2015.

[21] T. Pradeep, D. Mehra y P. H. Le, "Histology, eye," 2019.

[22] J. L. Berdonces, *El Gran libro de la iridología: el iris de los ojos refleja la salud*. RBA Libros, 2015.

[23] L. Brusi, *El iris. Exploración con biomicroscopio ocular: técnicas y protocolo de intervención*, pp. 338–351. Editorial de la Universidad Nacional de La Plata (EDULP), 2014.

[24] J. W. Rohen, Y. Chihiro y E. Lütjen-Drecoll, *Atlas de anatomía humana*. Elsevier España, 2003.

[25] C. Angée, B. Nedelec, E. Erjavec, J. Rozet y L. F. Taie, "Congenital Microcoria: Clinical Features and Molecular Genetics," *Genes*, vol. 12, no. 5, p. 624, 2021, doi: 10.3390/genes12050624.

[26] M. S. García-Vázquez y A. Á. Ramírez-Acosta, "Avances en el reconocimiento del iris: perspectivas y oportunidades en la investigación de algoritmos biométricos," *Computación y Sistemas*, vol. 16, no. 3, pp. 267–276, 2012.

[27] J. R. Matey et al., "Iris on the move: Acquisition of images for iris recognition in less constrained environments," *Proceedings of the IEEE*, vol. 94, no. 11, pp. 1936–1947, 2007.

[28] J. A. Freeman y D. M. Skapura, *Neural Networks: Algorithms, Applications, and Programming Techniques*. Addison Wesley, 1991.

[29] S. C. Shapiro, *Encyclopedia of Artificial Intelligence*, 2nd ed. New Jersey: A Wiley Interscience Publication, 1992.

[30] R. P. Díez, A. G. Gómez y N. de Abajo Martínez, *Introducción a la inteligencia artificial: sistemas expertos, redes neuronales artificiales y computación evolutiva*. Universidad de Oviedo, 2001.

[31] D. Andrés, R. Leal, J. Pauline, V. Porto, C. Arturo y R. Algarín, *El camino a las redes neuronales artificiales*. Editorial Unimagdalena, 2021.

[32] V. M. Alcaraz, *Estructura y función del sistema nervioso*. UNAM, 2000.

[33] M. I. Escobar y H. J. Pimienta, *Sistema nervioso*. Universidad del Valle, 2003.

[34] D. J. Matich, *Redes Neuronales: Conceptos básicos y aplicaciones*. Universidad Tecnológica Nacional, México, 2001, pp. 12–16.

[35] "General Python FAQ," Python Documentation. [En línea]. Disponible en: https://docs.python.org/3/faq/general.html#what-is-python.

[36] "AMD AI Solutions," AMD. [En línea]. Disponible en: https://www.amd.com/en/solutions/ai.html.

[37] K. O'Shea y R. Nash, "An introduction to convolutional neural networks," *arXiv preprint*, arXiv:1511.08458, 2015.

[38] N. Aloysius y M. Geetha, "A review on deep convolutional neural networks," en *Proc. 2017 International Conference on Communication and Signal Processing (ICCSP)*, pp. 588–592, 2017.

[39] S. Sakib, N. Ahmed, A. J. Kabir y H. Ahmed, "An overview of convolutional neural network: Its architecture and applications," 2019.

[40] J. He, L. Li, J. Xu y C. Zheng, "ReLU deep neural networks and linear finite elements," *arXiv preprint*, arXiv:1807.03973, 2018.

[41] H. Gholamalinezhad y H. Khosravi, "Pooling methods in deep neural networks, a review," *arXiv preprint*, arXiv:2009.07485, 2020.

[42] R. Singh, "Decoding CNNs: A Beginner's Guide to Convolutional Neural Networks and their Applications," *Medium*, 4 feb. 2025. [En línea]. Disponible en: https://ravjot03.medium.com/decoding-cnns-a-beginners-guide-to-convolutional-neural-networks-and-their-applications-1a8806cbf536.

[43] Z. Ge, S. Liu, F. Wang, Z. Li y J. Sun, "YOLOX: Exceeding YOLO series in 2021," *arXiv preprint*, arXiv:2107.08430, 2021.

[44] "MPSOC AMD Zynq UltraScale+," AMD. [En línea]. Disponible en: https://www.amd.com/es/products/adaptive-socs-and-fpgas/soc/zynq-ultrascale-plus-mpsoc.html.

[45] "About," Roboflow. [En línea]. Disponible en: https://roboflow.com/about.

[46] "Teledyne e2v | Teledyne Vision Solutions," [En línea]. Disponible en: https://www.teledynevisionsolutions.com/company/about-teledyne-vision-solutions/teledyne-e2v/.

[47] "AMD Technical Information Portal," AMD. [En línea]. Disponible en:

https://docs.amd.com/v/u/en-US/ds891-zynq-ultrascale-plus-overview.

[48] J. Redmon, S. Divvala, R. Girshick y A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," en *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779–788, 2016.

[49] "YOLOX: El Detector de Objetos del Futuro y su Entrenamiento Personalizado." [En línea]. Disponible en: https://www.toolify.ai/es/ai-news-es/yolox-el-detector-de-objetos-del-futuro-y-su-entrenamiento-personalizado-3526273#bar1.

[50] "Getting Started with YOLOX for Object Detection - MATLAB & Simulink," MathWorks. [En línea]. Disponible en: https://es.mathworks.com/help/vision/ug/getting-started-with-yolox-object-detection.html.

[51] C. A. Ruiz-Beltrán et al., "Real-time embedded eye detection system," *Expert Systems with Applications*, vol. 194, p. 116505, 2022.

[52] "Megvii-BaseDetection. YOLOX/tools/train.py at main," GitHub. [En línea]. Disponible en: https://github.com/Megvii-BaseDetection/YOLOX/blob/main/tools/train.py.

[53] J. Terven, D. M. Córdova-Esparza y J. A. Romero-González, "A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS," *Machine Learning and Knowledge Extraction*, vol. 5, no. 4, pp. 1680–1716, 2023.

[54] S. C. Shapiro, *Encyclopedia of Artificial Intelligence*, 2nd ed. New Jersey: A Wiley Interscience Publication, 1992.

[55] "DPU IP Details and System Integration — Vitis™ AI 3.5 documentation," Xilinx. [En línea]. Disponible en: https://xilinx.github.io/Vitis-AI/3.5/html/docs/workflow-system-integration.html.

[56] P. Rathore, “Diving deeper into Quantization Realm: Introduction to PTQ and QAT,” Medium, 30-Aug-2023. [En línea]. Disponible en: https://iprathore71.medium.com/diving-deeper-into-quantization-realm-9c73e3172a3c.

UNIVERSIDAD DE MÁLAGA | uma.es

E.T.S de Ingeniería Informática
Bulevar Louis Pasteur, 35
Campus de Teatinos
29071 Málaga

E.T.S. DE INGENIERÍA INFORMÁTICA