

UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA
Máster Oficial en Sistemas Inteligentes y Aplicaciones Numéricas en
Ingeniería



Trabajo Final de Máster

**Comparación de métodos de re-identificación basados
en AlignedReID++**

Óscar Jesús Díaz de la Fe

Tutores: José Javier Lorenzo Navarro
Modesto Fernando Castrillón Santana

18 de Abril de 2023

Dedicado a todo aquel interesado en la identificación de personas.

Agradecimientos

Mi primer agradecimiento quiero dedicárselo a mis tutores, Javier y Modesto, por estar en todo momento para darme consejos y ánimos a lo largo de todo este trabajo, les agradezco toda la formación que me han prestado tanto en este proyecto como en el curso impartido.

En segundo lugar, y no menos importante, quiero agradecer a mi familia, su apoyo incondicional en mis estudios y sobre todo en este proyecto

Y finalmente a mi compañero Airán por su ayuda a la hora de organizar mis ideas y darme otro punto de vista.

Resumen

Este proyecto tiene como objetivo realizar diferentes experimentos, haciendo uso de un algoritmo diseñado para la identificación de personas, el cual se alimentará a través de un conjunto de datos complejos, consiguiendo con ello, determinar el estado del arte actual. Se tendrá en cuenta la gran cantidad de factores externos que influyen en la dificultad de estos procesos.

Para la elaboración de este proyecto se hará uso de la arquitectura de la re-identificación conocido como AlignedID++, que propone utilizar su método propio de información local de coincidencia dinámica (DMLI) conjuntamente con funciones globales y funciones locales basadas en el mismo, permitiendo así una mayor precisión en la fase de inferencia. Además, se hará uso del conjunto de datos TGC20ReId, adquirido en un contexto deportivo, el cual se compone de 606 corredores y cuyas imágenes se obtendrán desde cinco puntos de registros distintos, conteniendo con ello, una gran cantidad de variaciones entre cada uno de los puntos.

Los resultados obtenidos permitirán determinar la eficacia de estos métodos en la actualidad y abrir la posibilidad de utilizarlos como táctica para evitar trampas en eventos deportivos. En resumen, el proyecto implica la exploración y experimentación de técnicas avanzadas en re-identificación de personas en el ámbito deportivo, con el fin de mejorar la precisión y la confiabilidad de estos procesos en la actualidad.

Abstract

This project aims to carry out different experiments, making use of an algorithm designed for the identification of people, which will be fed through a set of complex data, thereby determining the current state of the art. The large number of external factors that influence the difficulty of these processes will be taken into account.

For the development of this project, it will use a re-identification framework known as AlignedReID++, which proposes to use its own dynamic matching local information (DMLI) method together with global functions and local functions based on it, thus allowing greater precision in the inference phase. . In addition, the dataset TGC20ReId, acquired in a sporting context, will be used, which is made up of 606 runners and whose images will be obtained from five different record points, thus containing a large number of variations between each of the points.

The results obtained will allow to determine the effectiveness of these methods at present and open the possibility of using them as a tactic to avoid cheating in sporting events. In summary, the project implies the exploration and experimentation of advanced techniques for re-identifying people in the sports field, in order to improve the precision and reliability of these processes today.

Índice general

Resumen	III
Abstract	IV
1. Introducción	2
1.1. Hipótesis	3
1.2. Objetivos	4
2. Estado del arte	5
2.1. Re-identificación de personas	5
2.2. Aprendizaje profundo	7
2.2.1. Redes neuronales	8
2.2.2. Redes convolutivas	9
2.2.3. Aprendizaje de características	11
2.3. Transfer learning	12
2.4. Zero Shot Learning	14
2.5. ReID de personas en competiciones deportivas	16
2.6. AlignedReID++	18
3. Conjuntos de datos	23
3.1. Market1501	23
3.2. TGC20ReID	24
3.3. Preparación conjunto de datos	26
4. Herramientas	29
4.1. Python	29
4.1.1. Jupyter Notebook	30
4.1.2. VirtualBox	31
4.1.3. Librería Bob	32
4.1.4. Microsoft Teams	32

5. Experimentos	34
5.1. Zero shot learning	36
5.2. Entrenamiento desde cero	37
5.3. Transfer learning	38
5.4. Resultados de los experimentos	39
6. Conclusión y trabajos futuros	41
6.1. Valoración personal	41
6.2. Mejoras futuras	42
6.2.1. Ampliar conjunto de datos	42
6.2.2. Aprendizaje no supervisado	43
6.2.3. Reducción de ruido	43

Índice de figuras

2.1.	Proceso de un sistema de identificación, mostrando los pasos anteriores (fuente: [1]).	6
2.2.	Ejemplo red neuronal (fuente: [2])	8
2.3.	Esquema red convolutiva (fuente: [3])	10
2.4.	“Fully connected” (fuente: [4])	10
2.5.	Diferentes estrategias para el aprendizaje de características en la ReID de personas (fuente: [1])	12
2.6.	Aprendizaje por transferencia inductiva(Fuente [5]	13
2.7.	Estrategias de “Transfer Learning“ en “Deep Learning“(Fuente [5])	14
2.8.	“Zero-Shot Learning“ (fuente: [6])	15
2.9.	Sistema de control de tiempos en carreras deportivas de AIDE. (fuente: [7])	17
2.10.	Estructura alignedreid++ (fuente: [8])	19
2.11.	Funcionamiento de DMLI (fuente: [8])	20
2.12.	Resultados de alineado y mapa de distancia. (fuente: [8])	20
2.13.	Triplet Loss (fuente: [9])	21
3.1.	Imágenes de consulta de muestra de Market1501. Las “bboxes“ son dibujadas manualmente. Cada persona o identidad en el conjunto de datos tiene un máximo de 6 consultas asociadas, una para cada cámara presente en el escenario. (fuente: [10]) .	25
3.2.	Puntos de registros (fuente: [11])	26
3.3.	Variaciones de un corredor (fuente: [11])	27
3.4.	Imagen tras procesado	28
4.1.	Logo Python	30
4.2.	Logo jupyter	31
4.3.	Logo jupyter	32
4.4.	Librería bob	32
4.5.	Logo jupyter	33

ÍNDICE DE FIGURAS

IV

5.1. Esquema matriz de distancias	35
5.2. CMC zero shot learning	36
5.3. CMC entrenado con TGCReId	37
5.4. CMC entrenado con TGC	38
5.5. CMC comparativa	39
5.6. Resultados comparando distintos métodos (fuente: [8])	40

Índice de cuadros

3.1.	Datos Market-1501	24
3.2.	Características conjunto de datos (fuente: [11])	26
3.3.	Transformación nombres conjunto de datos	27
5.1.	Experimentos	39

Capítulo 1

Introducción

En los últimos años, los eventos deportivos han experimentado un crecimiento exponencial, lo que ha llevado a un aumento en la cantidad de personas participantes, así como en la complejidad de las actividades y en las recompensas ofrecidas por ellos. Este aumento de la popularidad de los eventos deportivos se ha manifestado en muchos lugares del mundo, y España no es una excepción.

En este país, se pueden encontrar una gran variedad de eventos deportivos, desde carreras de bicicletas hasta maratones y carreras de campo a través. Uno de los eventos más populares en España es la Vuelta Ciclista a España, que se celebra anualmente desde 1935 [12]. Desde entonces, la carrera ha ido creciendo en popularidad y en complejidad, hasta llegar a la actualidad, donde la 77^a edición contó con la participación de 168 ciclistas divididos en 21 equipos con 21 etapas [13] [14], con un total de 1.112.640 € en premios repartidos a lo largo de la carrera [15].

Otro ejemplo de evento deportivo popular en España, es la San Silvestre Vallecana, una carrera de 10 kilómetros que se celebra anualmente en Madrid. La carrera tiene más de 50 años de historia y cuenta con la participación de 40.000 corredores que toman las calles de la ciudad para despedir el año. La carrera ha sido galardonada con la Gold Race Running Label de la IAAF, que se otorga a pocas carreras, debido a los rigurosos requisitos que deben cumplir tanto la carrera como el evento en sí [16].

Estos eventos deportivos son gestionados por entidades organizadoras que tienen en cuenta no solo el impacto social, sino también el impacto económico. Para ello, se tiene en cuenta la preparación de los participantes, el equipo necesario y los circuitos a recorrer, y la gente necesaria para planificar, organizar y preparar el evento correctamente y cumplir con los reglamentos. Además, las recompensas económicas entregadas a los vencedores pueden ser significativas, lo que atrae a participantes de todo el mundo a estos eventos.

En la organización de los eventos deportivos, también se tienen en cuenta los requisitos de seguridad necesarios para garantizar la integridad física de los participantes y el público en general. Esto incluye la presencia de servicios médicos en caso de lesiones o emergencias, así como la coordinación con las autoridades locales para garantizar la seguridad de los espectadores y la comunidad en general.

Además, los eventos deportivos tienen un impacto económico. Los participantes, entrenadores, espectadores y el personal del evento pueden gastar dinero en alojamientos, comidas, transportes y otros gastos relacionados con el evento. Esto puede tener un impacto positivo en la economía local, especialmente para las empresas locales que se benefician directamente de la presencia de los participantes y los espectadores en el evento.

Las federaciones procuran asegurar que los participantes no realicen ningún tipo de trampas, ya sea por el uso de dopantes, equipo manipulado, la sustitución de los corredores en medio del recorrido, e incluso ventajas aerodinámicas por factores externos a la carrera.

Tratando más el tema de la sustitución de corredores, uno de los métodos fraudulentos más comunes es la falsificación de dorsales o el traspaso de estos a otras personas. Esto se realiza generalmente para finalizar las competiciones, falseando su identidad con la de otro corredor ya inscrito, siendo esto completamente ilegal, por ir en contra del reglamento deportivo al tratarse de un caso de robo de identidad. Un ejemplo de esto puede ser en el Trail Valle de Tena, donde dos corredores fueron descubiertos haciendo esta tipo de fraude y obligados a devolver sus premios [17].

1.1. Hipótesis

La re-identificación de personas (ReID) es una tarea crítica de la Visión por Computador que busca detectar correctamente a las personas en múltiples cámaras o conjuntos de datos. Aunque la ReID ha logrado avances significativos en los últimos años gracias al progreso en el aprendizaje automático y la Visión por Computador, aún enfrenta importantes desafíos. Uno de los principales obstáculos es la variabilidad en la apariencia de la imagen debido a los cambios en el punto de vista, la iluminación y la resolución de la imagen. Además, las imágenes pueden estar obstruidas por objetos o fondos irrelevantes, lo que dificulta enfocarse en las características importantes. Además, la falta de datos e imágenes anotadas dificulta el entrenamiento de modelos con suficiente precisión.

A pesar de estos desafíos, existen numerosos sistemas que se han demostrado eficaces en situaciones controladas para precisar la identidad de

individuos objetivos. Esta capacidad es útil para confirmar la identidad de los distintos participantes a lo largo del circuito y para detectar con mayor rapidez y eficiencia los casos de artimañas. En resumen, la ReID es una tarea desafiante pero crítica en el campo de la Visión por Computador y ha visto importantes avances en los últimos años que continúan mejorando su capacidad y precisión.

1.2. Objetivos

Continuando con lo mencionado en el apartado anterior, se plantea la hipótesis de que, utilizando las técnicas de ReID ya existentes, se podría llegar a diseñar un sistema robusto para poder detectar el uso de trampas que tengan que ver con el robo de dorsales o el evitar partes del recorrido saliendo del circuito. Para ello se plantean los siguientes objetivos:

- Analizar el actual estado del arte para determinar las técnicas más eficientes de ReID de personas.
- Probar estos métodos con un conjunto de datos complejo, con una gran cantidad de variables.
- Analizar los resultados obtenidos para determinar la robustez y eficacia del método en ambiente no controlados e indicar lo lejano que podemos estar de usar estas técnicas de forma eficaz.

Capítulo 2

Estado del arte

2.1. Re-identificación de personas

Como se ha hablado previamente, la ReID de personas consiste en la identificación y seguimiento de un individuo a través de múltiples cámaras. Es un campo de investigación muy interesante, debido a todas sus posibilidades de aplicación, llegando a ser desde medidas de seguridad, hasta localización para el uso de sistema de vigilancia. Para ello es necesario una gran cantidad de datos en vídeo (o imágenes) que resulta extremadamente poco eficiente ser procesado por un equipo humano, por lo que, a lo largo de los años, se han llegado a diseñar diferentes propuestas para automatizar esta tarea. Sin embargo, aunque se han realizado varios métodos que han dado buenos resultados con respecto a los conjuntos de datos existentes, aún falta mucho para resolver el problema debido a la gran complejidad del desafío que resulta la gran variación que puede haber en los datos, debido a alteraciones de la pose de los individuos, la iluminación, la calidad de la imagen, oclusión. etc. Razón por la que estos diseños se hacen principalmente con conjunto de datos en situaciones controladas, y su eficiencia puede variar al compararlo con una experimentación en entorno real, siendo esto referido como la diferencia de un experimento en un entorno cerrado y uno abierto [1].

En términos generales, se pueden definir una serie de pasos que debe seguir a la hora de desarrollar un sistema de ReID, estos son:

- **Recogida de datos:** consiste en la obtención de los datos a través de una o varias cámaras. Estos datos en bruto, deben ser procesados antes de poder ser utilizados.
- **Generación de las cajas delimitadoras:** este es un punto fundamental, debido a que se crean los cuadros delimitadores que contienen

las imágenes de personas a partir de los datos. Esto se hace manualmente en todas las imágenes, o mediante algoritmos de detección, siendo esto lo más práctico a la hora de trabajar con una gran cantidad de imágenes.

- **Anotación de los datos:** consiste en la anotación de las etiquetas de las imágenes y cámaras. Es indispensable para el aprendizaje de un modelo de ReID.
- **Entrenamiento del modelo:** se trata del entrenamiento de un modelo de ReID, robusto con las imágenes o vídeos de personas anotadas. Se han desarrollado múltiples modelos y técnicas, basados en el aprendizaje de la representación de características, el aprendizaje de la métrica de la distancia o una combinación de ellos.

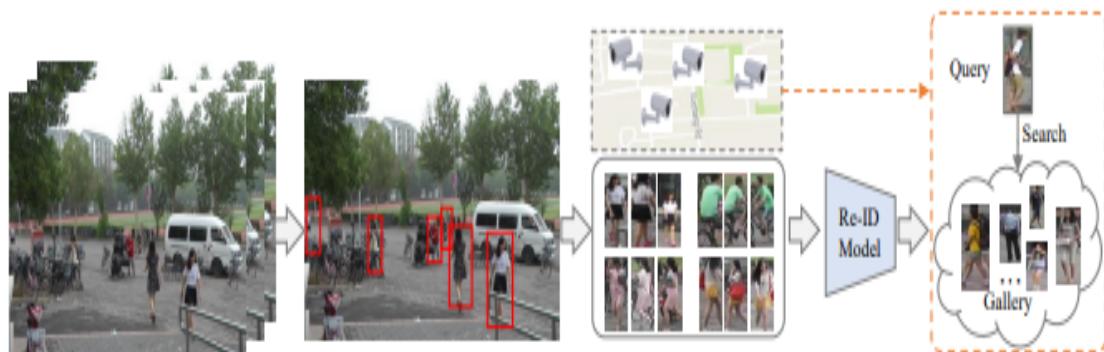


Figura 2.1: Proceso de un sistema de identificación, mostrando los pasos anteriores (fuente: [1]).

Como se mencionó previamente, existen dos entornos en la ReID, el entorno cerrado, donde las identidades que aparecen en las pruebas coinciden con los que se encuentran en la galería, y el entorno abierto, donde puede ser que en la galería no se encuentren las mismas identidades que en el conjunto de prueba o viceversa, e incluso que el número de imágenes a tratar sea infinito, aumentando muchísimo su dificultad.

De aquí se puede identificar varios puntos que diferencian ambos entornos a la hora de trabajar con ellos:

- **Datos homogéneos vs datos heterogéneos:** en los datos sin procesar, las identidades pueden estar representadas por imágenes o vídeos

capturados por una sola modalidad en un entorno cerrado. Sin embargo, en un experimento en entorno abierto, se podría necesitar procesar datos heterogéneos, como imágenes infrarrojas, bocetos, descripción de texto, etc.

- **Cuadros delimitadores vs imágenes y vídeos sin procesar:** en la generación de las cajas delimitadoras, en entorno cerrado el entrenamiento y las pruebas se realizan con respecto a las susodichas cajas delimitadoras, debido que es en este punto donde se encuentra la información de la apariencia de la persona. Mientras tanto, en entorno abierto la búsqueda de los individuos se debe realizar sin procesar las imágenes o vídeos.
- **Datos anotados vs datos no etiquetados:** para la anotación de los datos de entrenamiento, en entorno cerrado se considera que se tiene suficientes anotaciones de entrenamiento para desarrollar un modelo de ReID. No obstante, en entorno abierto es posible no tener suficientes datos anotados o incluso ninguna información.
- **Anotación correcta vs anotaciones ruidosa:** en el entrenamiento del modelo, en el entorno cerrado, se considera que todas las anotaciones son correctas. Sin embargo, en un entorno abierto, se producen errores de anotación o hay resultados imperfectos en la detección de individuos.
- **Entidad existente en la galería vs conjunto abierto:** en entorno cerrado, el individuo de interés aparece en la galería. Pero en entorno abierto, puede no aparecer en la galería, o se necesita la verificación en lugar de la recuperación.

2.2. Aprendizaje profundo

El aprendizaje profundo (también conocido como “deep learning”) es una rama del aprendizaje automático que se enfoca en entrenar redes neuronales artificiales para que aprendan y realicen tareas complejas, como reconocimiento de patrones, clasificación de imágenes y procesamiento del lenguaje natural. Estas redes neuronales artificiales están diseñadas para imitar el comportamiento de las neuronas en el cerebro humano, lo que les permite aprender de manera autónoma a partir de un gran conjunto de datos y mejorar su rendimiento a medida que se les proporciona más información.

El aprendizaje profundo ha tenido un gran impacto en áreas como la Visión por Computador, el reconocimiento de voz, la traducción automática y el procesamiento del lenguaje natural. Esto se debe a que las redes neuronales profundas pueden aprender y detectar patrones en los datos de entrada de manera más efectiva que los algoritmos de aprendizaje automático tradicionales, lo que permite una mayor precisión en las predicciones y clasificaciones.

El aprendizaje profundo se basa en modelos de redes neuronales profundas que constan de múltiples capas interconectadas de neuronas artificiales. Cada capa de neuronas procesa la información de entrada de manera diferente, lo que permite la detección de patrones en diferentes niveles de abstracción. A medida que la información fluye a través de la red, cada capa refina y procesa la información de entrada para generar una salida final que representa una predicción o clasificación.

2.2.1. Redes neuronales

Las redes neuronales son sistemas de computación que imitan el cerebro humano. Están compuestas por nodos llamados neuronas artificiales que trabajan juntos para aprender y reconocer patrones.

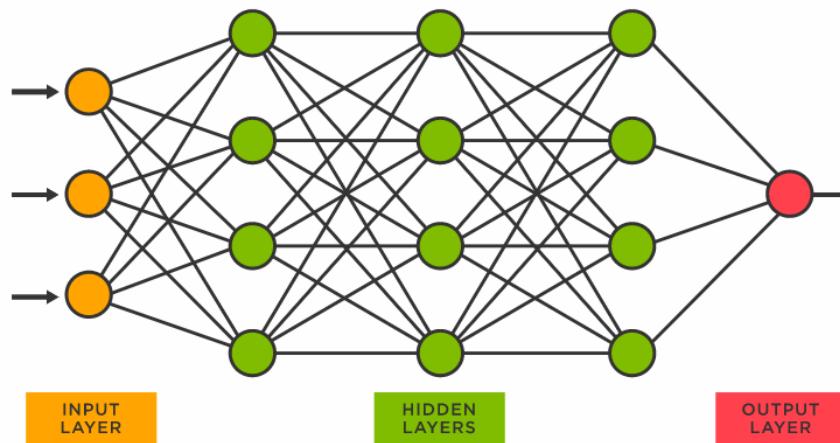


Figura 2.2: Ejemplo red neuronal (fuente: [2])

Cada neurona artificial en una red neuronal procesa información utilizando una combinación de entradas ponderadas y una función de activación. Como podemos ver en la imagen 2.2, estas neuronas se organizan en capas,

y la información fluye de la capa de entrada a través de capas ocultas hasta la capa de salida.

Para entrenar una red neuronal, se ajustan los pesos de las conexiones entre las neuronas para reducir la diferencia entre las salidas deseadas y las salidas reales. Este proceso de aprendizaje se basa en la retroalimentación y la optimización iterativa.

Las redes neuronales pueden aprender de ejemplos de entrenamiento y generalizar ese conocimiento para reconocer patrones complejos, hacer predicciones y tomar decisiones en diversas aplicaciones, como reconocimiento de imágenes, procesamiento del lenguaje natural, sistemas de recomendación, conducción autónoma, y muchos otros campos.

2.2.2. Redes convolutivas

Las redes convolutivas, también conocidas como “Convolutional Neural Networks” (CNN), son un tipo de modelo de aprendizaje profundo que se utiliza en el procesamiento de imágenes y en otras aplicaciones que trabajan con datos de tipo matriz. Las redes convolutivas han demostrado ser altamente efectivas en la tarea de reconocimiento de patrones visuales, como la clasificación de imágenes y el análisis de objetos en tiempo real.

Una red convolutiva se compone de varias capas de convolución, activación y agrupamiento. En la capa de convolución, la red aprende a identificar características relevantes en la imagen utilizando un conjunto de filtros que se aplican a la imagen de entrada. Estos filtros son matrices de números que se ajustan durante el entrenamiento de la red para encontrar patrones específicos en la imagen.

La capa de activación utiliza una función no lineal para introducir no linealidades en la red, lo que permite que la red aprenda patrones más complejos en la imagen. La capa de agrupamiento reduce la dimensionalidad de la salida de la capa de convolución, lo que reduce el costo computacional y evita el sobreajuste.

En general, las redes convolutivas son altamente efectivas en el procesamiento de imágenes debido a su capacidad para aprender automáticamente características relevantes y reducir la dimensionalidad de los datos. Esto hace que sean una herramienta importante en aplicaciones como el reconocimiento facial, la detección de objetos en imágenes y la segmentación de imágenes médicas, etc.

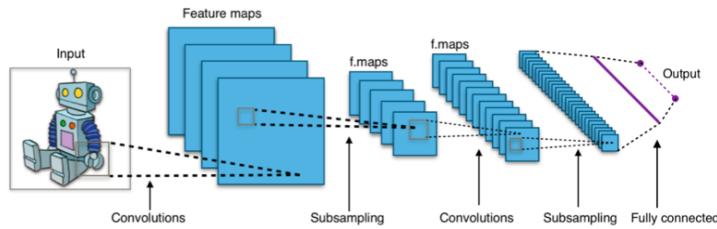


Figura 2.3: Esquema red convolutiva (fuente: [3])

Se puede también mencionar la existencia de la “Fully Connected Network“, también conocida como capa de red neuronal densa, siendo una capa en una red neuronal artificial en la que cada neurona está conectada a todas las neuronas de la capa anterior. En otras palabras, todas las entradas a una neurona están conectadas a todas las salidas de las neuronas de la capa anterior. Esto significa que cada entrada a la capa está ponderada por un peso y una función de activación se aplica a la suma ponderada de las entradas para producir la salida de la neurona.

Las capas totalmente conectadas son comúnmente utilizadas en la fase final de una red neuronal, donde se toman las características aprendidas por las capas convolutivas y se utilizan para clasificar o hacer predicciones sobre los datos de entrada. Por ejemplo, en una red neuronal convolutiva utilizada para clasificar imágenes, las capas convolutivas extraen características útiles de las imágenes, y luego estas características se pasan a través de capas totalmente conectadas para producir una predicción final sobre la clase o las clases perteneces en la imagen.

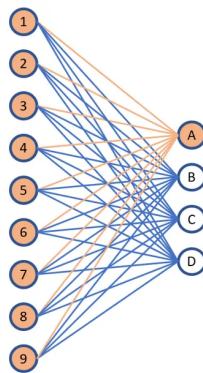


Figura 2.4: “Fully connected“ (fuente: [4])

2.2.3. Aprendizaje de características

Hay diferentes estrategias para el aprendizaje de características en el ámbito de la ReID [1].

- **Aprendizaje de características globales:** consiste en extraer un vector de características globales para cada imagen de persona, siendo una elección común integrarlas en técnicas avanzadas de aprendizaje profundo. Se suelen utilizar imágenes individuales o una combinación de múltiples imágenes en forma de tripleta.
- **Aprendizaje de características locales:** consiste en aprender características agregadas en región. Las partes del cuerpo son generadas automáticamente por estimación de pose o división más o menos horizontal.
- **Aprendizaje de características auxiliares:** este aprendizaje solicita información adicional anotada o muestras de entrenamiento generadas para mejorar el aprendizaje.
- **Aprendizaje de características en vídeo:** tratándose de un aprendizaje que ha ganado un interés creciente en la comunidad ReID, donde los individuos están representados por una secuencia de vídeo. El principal desafío es capturar con precisión la información temporal. Para ello, se han creado diferentes arquitecturas de red neuronal para este tipo de diseño, donde se optimiza conjuntamente la capa recurrente final para la propagación de la información temporal y la capa de agrupación temporal. Otro gran desafío es el seguimiento de fotogramas dentro de los videos, donde se ha empleado una regularización de diversidad [133] para extraer datos discriminativos en las diversas secuencia de vídeo. También es un desafío manejar las diferentes longitudes de secuencias de vídeo, donde se divide un vídeo largo en secuencias de múltiples fragmentos cortos, agregando los fragmentos mejor clasificados para aprender una característica concreta.

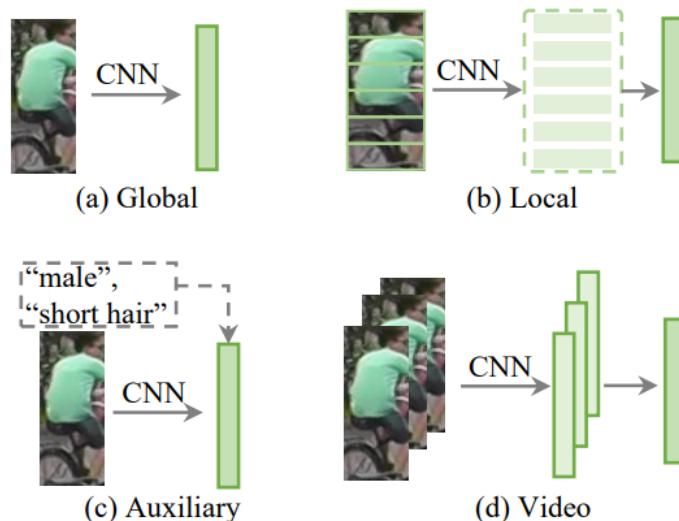


Figura 2.5: Diferentes estrategias para el aprendizaje de características en la ReID de personas (fuente: [1])

2.3. Transfer learning

El ”transfer learning” [5] es una técnica en el aprendizaje automático que consiste en utilizar un modelo pre-entrenado en una tarea específica y adaptarlo para otra tarea relacionada. En lugar de entrenar un modelo desde cero, utiliza un modelo ya entrenado en un conjunto de datos grande y complejo, para luego ajustar y re-entrenar el modelo en un conjunto de datos más pequeño y específico.

La idea detrás del ”transfer learning” es que, en muchos casos, los modelos pre-entrenados ya han aprendido características relevantes de los datos que se pueden aplicar a otras tareas relacionadas. Al utilizar un modelo pre-entrenado, se puede ahorrar tiempo y recursos de entrenamiento, ya que el modelo ya ha aprendido una representación de características útiles y solo se necesita ajustar para la tarea específica.

En términos generales, podemos distinguir 3 tipos de ”Transfer Learning“:

- Aprendizaje por transferencia inductiva: Donde se usan los mismos datos, pero las tareas de fuente y objetivo son diferentes. La idea consiste entonces en usar los modelos existentes para reducir el campo de aplicación de los modelos posibles (sesgo del modelo).

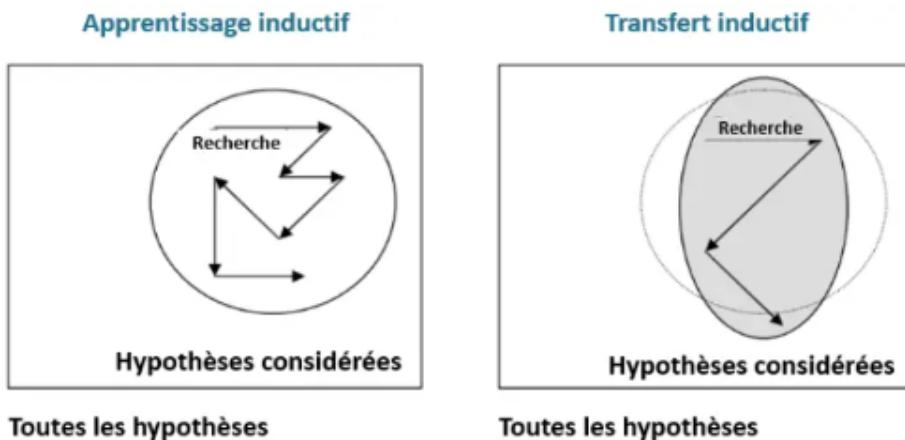


Figura 2.6: Aprendizaje por transferencia inductiva(Fuente [5])

- Aprendizaje por transferencia no supervisada: Debido a las grandes cantidades de datos no etiquetados que pueden haber, usar el aprendizaje no supervisado combinado con el “Transfer Learning” ha generado mucho interés. Un ejemplo es el “Self-taught clustering”, el cual es un método que permite realizar el “clustering” de pequeñas conjuntos de datos objetivo no etiquetados, con la ayuda de una gran cantidad de datos no etiquetados.
- Aprendizaje por transferencia transductiva: Donde las tareas fuente y objetivo son similares, pero sus campos correspondientes son diferentes en términos de datos o de distribuciones de probabilidad. Pueden ser adaptados a datos extraídos de redes sociales, cuyo contenido es diferente pero se asemeja al de los periódicos.

También se pueden determinar 2 tipos de estrategias al usar con respecto al “Deep Learning”:

- Utilización de modelos pre-entrenados como extractores de características: La arquitectura de estos modelos se presenta como un compilado de capas de neuronas. Estas capas adquieren diferentes características en función del nivel en el que se sitúan. La última capa (generalmente una “fully connected”) es utilizada para obtener el resultado final. La idea es reutilizar una red pre-entrenada sin capa final, donde la nueva red funciona como un extracto de características para realizar otras tareas.

- Ajustes de modelos pre-entrenados: En este método, no solo se ajusta la última capa, sino que también otras capas son re-entrenadas de manera selectiva. La idea es congelar, es decir fijar el peso de ciertas capas durante el entrenamiento y afinar el resto para responder al problema, permitiendo reutilizar los conocimientos en términos de arquitectura global de la red y explotar sus estados como punto de partida para el entrenamiento, permitiendo obtener mejores resultados en un tiempo de entrenamiento más corto.

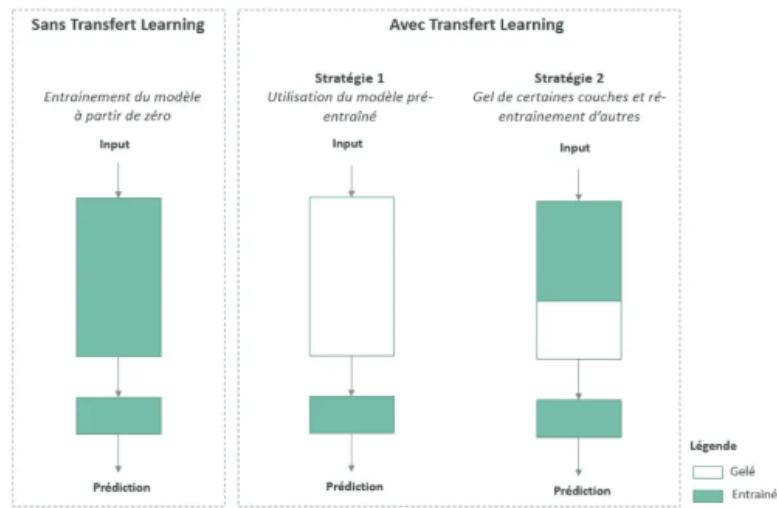


Figura 2.7: Estrategias de “Transfer Learning“ en “Deep Learning“ (Fuente [5])

2.4. Zero Shot Learning

El “Zero-Shot Learning“ (ZSL) [18] es una técnica de aprendizaje automático que permite a un modelo clasificar objetos de clases que nunca ha visto antes, sin necesidad de un entrenamiento específico para esas clases. Se suele utilizar para sistemas autónomos que deben ser capaces de identificar y categorizar nuevos objetos por sí mismos.

Como se puede apreciar en la imagen 2.8, el modelo se entrena previamente utilizando un conjunto de clases conocidas (clases vistas) y luego se le solicita que realice clasificaciones en un conjunto diferente de clases (clases no vistas), sin recibir entrenamiento adicional. El objetivo de ZSL es transferir el conocimiento ya contenido en el modelo a la tarea de clasificar las clases

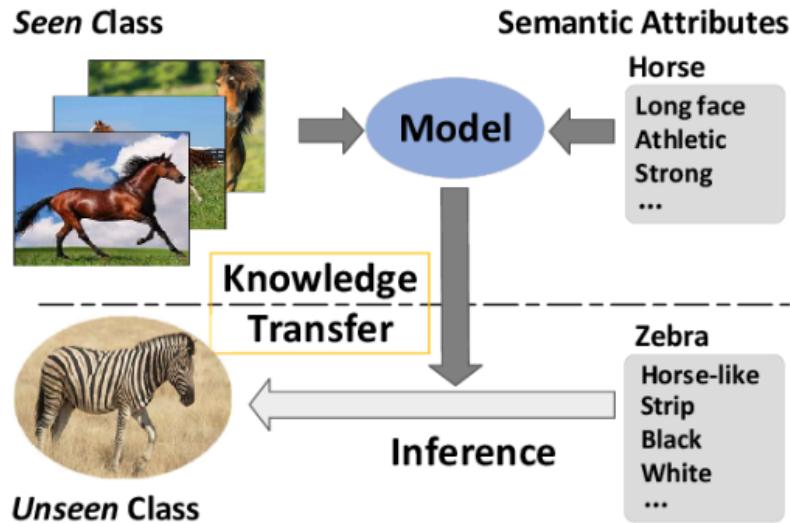


Figura 2.8: “Zero-Shot Learning”. (fuente: [6])

no vistas. ZSL es un subcampo del “transfer learning” y resulta extremadamente útil porque ayuda a superar los desafíos y costos asociados con la etiquetación de datos. Los dos enfoques más comunes utilizados para resolver los problemas de reconocimiento de tiro cero son:

- Métodos basados en clasificadores: implican el uso de un enfoque de uno-contra-todos, donde se entrena un clasificador binario separado para cada clase no vista.
- Métodos basados en instancias: implican en primero obtener ejemplos etiquetados de las clases no vistas y luego usar estos ejemplos para entrenar un clasificador.

Para evaluar el rendimiento de un clasificador ZSL, se calcula la precisión del top-k promedio por categoría. Esta métrica mide la proporción de muestras de prueba cuya clase real se encuentra entre las k clases con las probabilidades más altas pronosticadas por el clasificador.

La precisión de k superior promedio por categoría se calcula tomando el promedio de la precisión de k superior para cada clase, donde C indica el número de clases no vistas.

$$Accuracy = \frac{1}{C} \sum_{i=0}^{C-1} \frac{\# \text{ predicciones correctas en } i}{\# \text{ muestras en } i} \quad (2.1)$$

2.5. ReID de personas en competiciones deportivas

La ReID en el ámbito deportivo presenta varios desafíos únicos. En primer lugar, el entorno deportivo puede ser caótico y dinámico, con múltiples atletas en movimiento simultáneo, lo que dificulta la identificación precisa. Además, las condiciones de iluminación y los ángulos de cámara pueden variar, lo que afecta a la calidad de las imágenes y la visibilidad de los rasgos faciales u otras características distintivas. Además, los atletas pueden cambiar de vestimenta o usar accesorios como gorras, gafas o cascos, lo que dificulta aún más la tarea de ReID.

En las competiciones deportivas, la ReID de personas juega un papel crucial en diversas aplicaciones. Por ejemplo, puede utilizarse para el seguimiento de jugadores, permitiendo un análisis detallado del rendimiento individual y colectivo. También se utiliza para rastrear y registrar el rendimiento de los competidores en diferentes pruebas. Además, la ReID de personas puede ser útil en la seguridad de eventos deportivos, facilitando la identificación de individuos sospechosos o la detección de comportamientos anómalos.

En competiciones deportivas, las técnicas de ReID de personas se han beneficiado de los avances en el aprendizaje automático y la Visión por Computador. El uso de redes neuronales convolutivas, el aprendizaje profundo y el uso de conjuntos de datos etiquetados específicos han mejorado significativamente la precisión y eficacia de los sistemas de ReID. Además, la disponibilidad de hardware potente y la capacidad de procesamiento en tiempo real han permitido implementaciones prácticas de estas técnicas en entornos deportivos.

Para garantizar un seguimiento preciso de los corredores en competiciones deportivas, es fundamental tener conocimiento de la ubicación de los participantes, teniendo en cuenta que pueden pertenecer a diversas categorías. Esta complejidad añade un desafío adicional al problema. El objetivo principal es evitar cualquier intento de hacer trampas y garantizar la seguridad de los propios participantes. En la actualidad, se utiliza un sistema de seguimiento que implica el uso de balizas que los participantes deben llevar consigo. Estas balizas son detectadas cuando los corredores pasan por los puntos de control establecidos a lo largo del recorrido. Uno de los mayores desafíos consiste en asegurar la confiabilidad de estos sistemas, de manera que sean capaces de identificar a los participantes de manera precisa y consistente en todos los puntos de control.

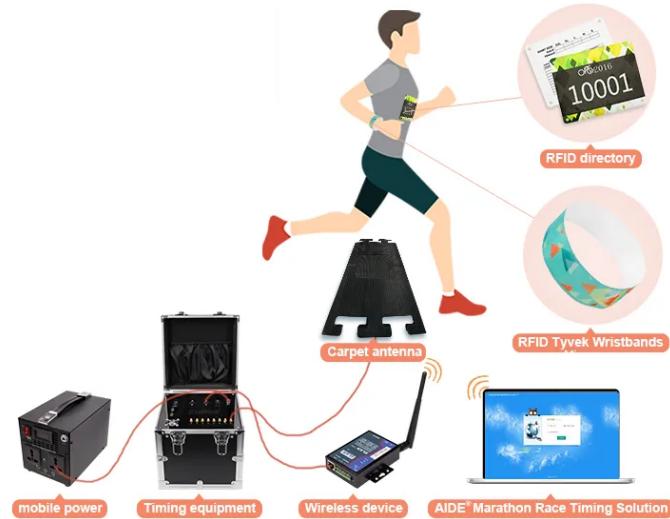


Figura 2.9: Sistema de control de tiempos en carreras deportivas de AIDE. (fuente: [7])

En la imagen 2.9 se puede apreciar cómo los corredores llevan una baliza colocada en el pecho, la cual se activa al pasar por el punto de control. Esta baliza establece una comunicación con los sensores ubicados en el punto de control, los cuales se encargan de identificar a la persona correspondiente. La información recopilada es transmitida al sistema de gestión, el cual cuenta con un sistema de respaldo, generalmente en formato papel, para garantizar la consistencia y precisión del sistema.

El mayor desafío de esta tecnología radica en la posibilidad de extravío o fallas en los sensores, lo que podría afectar negativamente a los corredores e incluso llevar a su descalificación de la carrera. Además, dependiendo de la velocidad de los competidores y la calidad de los sensores, existe el riesgo de que el seguimiento del corredor no se realice de manera precisa. Además, este tipo de sistemas solo permite verificar que la baliza ha pasado por los puntos de control, sin asegurar que sea el mismo participante que la llevaba inicialmente, lo que podría abrir la puerta a posibles fraudes.

Por otro lado, también se puede realizar un seguimiento manual de los corredores, donde se requiere la intervención de personal humano para observar los dorsales de los corredores que pasan por un punto determinado y registrar manualmente en papel o en un dispositivo la posición y el tiempo actual. Sin embargo, este enfoque no resulta muy eficiente, especialmente cuando hay una gran cantidad de participantes, lo que puede dar lugar a inconsistencias en los registros. Además, este tipo de sistemas conlleva un costo humano

considerable, lo que dificulta su implementación en competiciones con una gran afluencia de participantes y/o distancias largas.

Considerando todas estas necesidades, se ha reconocido la necesidad de implementar sistemas automatizados, en este caso, sistemas basados en aprendizaje automático centrados en la ReID de personas, con el fin de automatizar el registro de tiempo y atletas en eventos deportivos. De esta manera, se busca eliminar la dependencia de métodos manuales y propensos a errores, a la vez que mejora la integridad y eficiencia de las competiciones y proporcionar resultados más precisos para los atletas y los organizadores del evento.

2.6. AlignedReID++

El estado del arte, en la re-identificación, se encuentra en constante evolución, con numerosos enfoques y algoritmos desarrollados para mejorar la precisión y la eficacia de los sistemas. Entre las diversas opciones disponibles, se ha elegido utilizar AlignedReID++ [8] como base para llevar a cabo las pruebas y experimentos en este proyecto, debido a que ha demostrado ser una elección prometedora debido a su rendimiento sobresaliente en comparación con otros métodos existentes. Además, AlignedReID++ ha sido diseñado específicamente para abordar los desafíos de la re-identificación de personas, como la variación de pose, iluminación y fondo.

Se trata de una evolución de la arquitectura AlignedReID [19]. Se compone de un método denominado “Dynamically Matching Local Information” (DMLI), que alinea las imágenes en franjas horizontales sin necesidad de supervisión externa, permitiendo solucionar el problema de la variación de las poses de los individuos objetivos, el cuadro delimitador, etc. El método es aplicable a la mayoría de redes convolucionales basadas en ReID.

Como se puede apreciar en la imagen 2.10 de la arquitectura de AlignedReID++, se basa en el diseño de la arquitectura Resnet50 para la extracción del mapa de características ($C \times H \times W$, donde ‘ C ’ es el número de canales y ‘ $H \times W$ ’ el tamaño espacial) siendo la salida de la última capa de convolución. Además, se aplica una media global “pooling” para convertir el mapa de características en un vector de características globales ($C \times 1$), suponiendo que f_A y f_B son características globales de las imágenes A y B, permitiendo describir la distancia global como:

$$d_g(A, B) = \|f_A - f_B\|_2 \quad (2.2)$$

Esto se debe a que mediante la distancia global se puede medir, de una forma más sencilla, la similitud entre 2 imágenes, pero a costa de ignorar la información local. Para remediarlo se utiliza el DMLI, haciendo que la rama

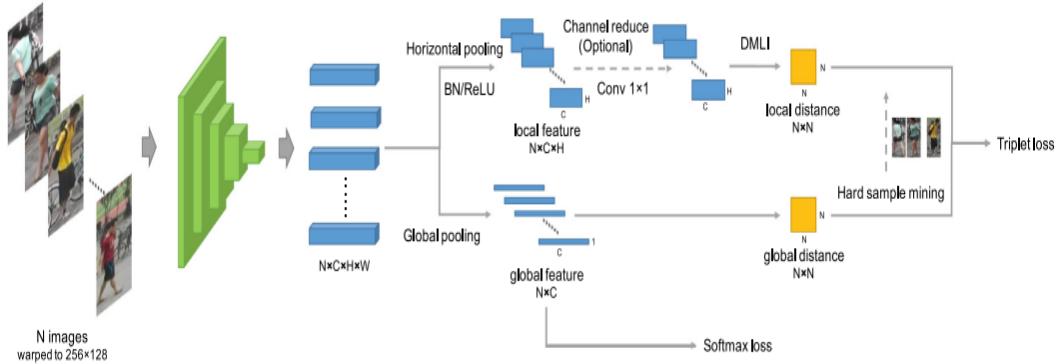


Figura 2.10: Estructura alignedreid++ (fuente: [8])

local utilice una agrupación horizontal máxima (o promedio) para convertir el mapa de características ($C \times H \times W$) a un mapa de características locales ($C \times H \times 1$) que a su vez será redimensionado al tamaño $H \times C$. Permitiendo que se defina l_A y l_B como características locales de las respectivas imágenes. Además se realiza una normalización de distancias entre 0 y 1 mediante transformación de elementos, donde $d_{i,j}$ se tratará de distancia entre la i -ésima parte vertical de la primera imagen y la j -ésima parte vertical de la segunda.

$$d_{i,j} = \frac{e^{\|l_A^i - l_B^j\|_2} - 1}{e^{\|l_A^i - l_B^j\|_2} + 1} \quad i, j \in 1, 2, 3, \dots, H \quad (2.3)$$

Con estas distancias $d_{i,j}$ se puede generar la matriz de distancias D. Es decir, la distancia local entre las imágenes se define como la distancia total de la ruta más corta desde (1,1) hasta (H,H) en la matriz D. Estas distancias pueden calcularse a través de la ecuación:

$$S_{i,j} = \begin{cases} d_{i,j} & i = 1, j = 1 \\ S_{i-1,j} + d_{i,j} & i \neq 1, j = 1 \\ S_{i,j-1} + d_{i,j} & i = 1, j \neq 1 \\ \min(S_{i-1,j}, S_{i,j-1}) + d_{i,j} & i \neq 1, j \neq 1 \end{cases} \quad (2.4)$$

Donde $S_{i,j}$ es la distancia total más corta desde (1,1) a (i, j) en la matriz D, y $S_{H,H}$ es la distancia total más corta al final, en este caso, la distancia local, entre las dos imágenes: $d_l(A, B) = S_{H,H}$. En la etapa de inferencia, la distancia total de dos imágenes es:

$$d(A, B) = d_g(A, B) + \lambda d_l(A, B) \quad (2.5)$$

Donde λ es el peso para equilibrar la distancia global y local.

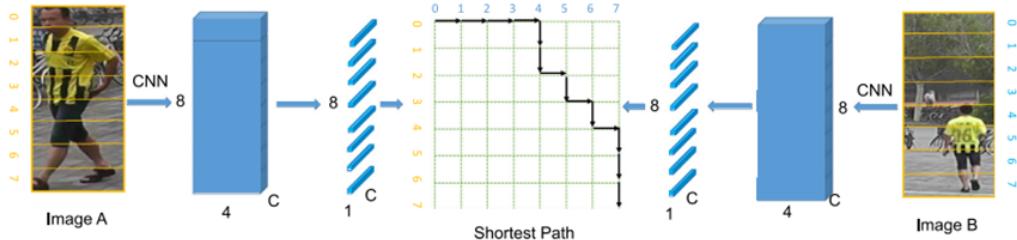


Figura 2.11: Funcionamiento de DMLI (fuente: [8])

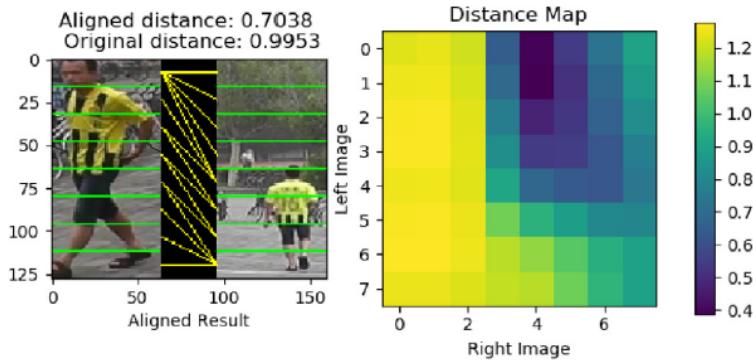


Figura 2.12: Resultados de alineado y mapa de distancia. (fuente: [8])

Sin embargo, las características locales pueden detectar algunas diferencias. Por ello, en la red AlignedReID++ las características globales se aprenden junto con las locales basadas en DMLI.

La red troncal y la rama global, son las mismas que con la mayoría de los trabajos anteriores de ReID que se entrena con pérdida “softmax” (L_{ID}), combinadas con “TriHard loss” ($L_{ID} + L_T^g$), este último consiste en que en el entrenamiento, agrega una capa de normalización de lotes y “ReLU” antes de agruparlos horizontalmente. Además de una capa convolutiva 1x1, para reducir el número de canales del mapa de características, con el fin de adelantar el entrenamiento.

Utilizando distancias globales para extraer las muestras debido a que el cálculo es mucho más rápido y que la optimización de las ramas globales y

locales con diferentes tripletas no proporciona ninguna ayuda a la red. Considerando L_T^l como la “triplet loss” [9] de la rama local, siendo el “triplet loss” una pérdida que está basada en el contexto de la clasificación del vecino más cercano. Pretendiendo asegurar que una imagen x_i^a (ancla) de un individuo concreto sea la más próxima de todas las imágenes x_i^p (positivos) de la misma persona que de cualquier otra imagen x_i^n (negativos) de cualquier otra persona, como se puede ver en la siguiente imagen:

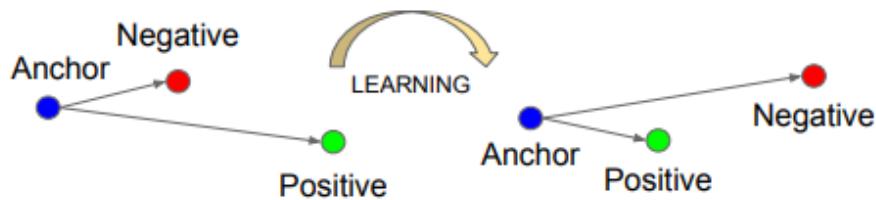


Figura 2.13: Triplet Loss (fuente: [9])

De esta forma se llega a las ecuaciones:

$$\|f(x_i^a) - f(x_i^p)\|_2^2 + \alpha < \|f(x_i^a) - f(x_i^n)\|_2^2, \quad (2.6)$$

$$\forall (f(x_i^a), f(x_i^p), f(x_i^n)) \in T \quad (2.7)$$

Donde α es un margen que se aplica entre pares positivo y negativos. T es el conjunto de todas las tripletas posibles en el conjunto de entrenamiento y tiene cardinalidad N .

Como se pretenden minimizar la perdida se llega a la ecuación:

$$L = \sum_i^N [\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha] \quad (2.8)$$

Esto permite que todas las tripletas posibles den como resultado muchas tripletas que cumplen las ecuaciones anteriores. Estos no aporta al entrenamiento ni el resultado en una convergencia más lenta, pero todavía actúan a través de la red.

Teniendo en cuenta lo anterior, se determina que la perdida total de AlignedReID++ es:

$$L = L_{ID} + L_T^g + L_T^l \quad (2.9)$$

La propagación hacia atrás de AlignedReID++ se calcula fácilmente por la diferenciación automática del entorno del aprendizaje profundo. Utilizando

las ecuaciones 1 y 2 y denotando $x = ||l_A^i - l_B^j||_2$. Considerando que d_{ij} cumple 2 apartados. El primero, $d_{ij}\alpha x$, no provoca cambios de x, en un rango entre 0 y 1, permitiendo aplicar pequeños caminos para alinear las características locales, dejando el modelo más estable. El segundo, $\frac{\partial d_{ij}}{\partial x} \frac{1}{\alpha}x$ donde el cálculo sería:

$$\frac{\partial d_{ij}}{\partial x} = \frac{2}{e^x + \frac{1}{e^x} + 2} \quad e^x \in [1, e^{\max(x)}] \quad (2.10)$$

Donde $\frac{\partial d_{ij}}{\partial x} \frac{1}{\alpha}x$ desde que e^x no es menor a 1. La alineación no correspondiente tiene una gran distancia L2, y su gradiente es próximo a 0 en la ecuación anterior. Por ello, la contribución de tales alineaciones en el camino más corto es pequeña. Por lo que la distancia local entre las imágenes está determinada principalmente por las alineaciones correspondientes.

Capítulo 3

Conjuntos de datos

Los conjuntos de datos de ReID son colecciones de imágenes y metadatos asociados que se utilizan para entrenar y evaluar algoritmos de reidentificación de personas. Estos conjuntos de datos son fundamentales en la investigación y desarrollo de técnicas de ReID, ya que proporcionan una base para el entrenamiento y la evaluación de modelos.

Los conjuntos de datos de ReID varían en tamaño y complejidad y Aligned-ReID++ ha sido evaluado para trabajar con algunos de los conjuntos de datos más conocidos, los cuales son MSMT17 [20], CUHK03 [21], DukeMTMCReID [22] y Market1501 [10].

A continuación, nos centraremos en el conjunto de datos Market1501, esto debido a que su estructura de datos, donde hay diferentes identidades en distintas cámaras, resulta similar al que se usa en el conjunto TGC20ReID [11], el cual es el que se plantea para realizar los futuros experimentos.

3.1. Market1501

Market1501 [10] es un conjunto de datos que contiene imágenes capturadas por seis cámaras situadas frente a un supermercado. Estas cámaras incluyen cinco cámaras HD de alta resolución (1280×1080) y una cámara SD de resolución estándar (720×576), donde hay superposición entre estas cámaras. El conjunto de datos incluye 32668 cajas delimitadoras (“bboxes”) correspondientes a 1501 identidades. Debido al entorno abierto, cada identidad en el conjunto de datos es capturada por un máximo de seis cámaras para permitir la búsqueda entre cámaras, por lo que se aseguró que cada identidad anotada es capturada por al menos dos cámaras.

Este conjunto de datos presenta las siguientes características:

- A diferencia de otros conjuntos de datos que utilizan “bboxes” recorta-

dos manualmente, Market-1501 emplea el Modelo de Partes Deformables (DPM) [23] para la detección de “bboxes“.

- Dentro de este conjunto de datos, se aplica una clasificación a las “bboxes“ detectadas utilizando la relación de área entre ellas. Si la relación de área supera el 50 %, se considera que la bbox detectada es “buena“. Por otro lado, si la relación de área es menor al 20 %, se etiqueta como “distractor“. Cualquier bbox que no cumpla con estas condiciones se clasifica como “basura“, lo que indica que la imagen no tiene ninguna influencia significativa en la precisión de la ReID. Además, se identifican y marcan como “distractores“ algunas “bboxes“ que son falsas alarmas evidentes en el conjunto de datos.
- Cada identidad puede tener varias imágenes en cada cámara, lo que permite tener múltiples consultas y análisis para cada identidad durante la búsqueda entre cámaras. Esto refleja mejor la situación práctica donde se pueden aprovechar múltiples consultas para obtener información más discriminatoria sobre la persona de interés.

Conjunto	Market-1501
# identidades	1501
# “bboxes“	32668
# distractores	2,793 + 500K
# cámaras	6

Tabla 3.1: Datos Market-1501

3.2. TGC20ReID

Como se dijo al principio de esta sección, el conjunto de datos a utilizar se trata del TGC20ReID [11], formado por datos adquiridos en un escenario deportivo, que contiene imágenes de los corredores de la Transgrancanaria (TGC) 2020 “Classic“ donde los participantes fueron grabados en cinco puntos de registro diferentes a lo largo de la carrera.

Lo interesante de utilizar este conjunto de datos es que el mismo corredor puede aparecer en condiciones muy diferentes debido a la duración de la carrera y las características de los puntos de registros, llegando a notarse cambios de iluminación, e incluso la apariencia del corredor no se mantiene durante toda la carrera, debido a cambios de accesorios. Haciendo que



Figura 3.1: Imágenes de consulta de muestra de Market1501. Las “bboxes” son dibujadas manualmente. Cada persona o identidad en el conjunto de datos tiene un máximo de 6 consultas asociadas, una para cada cámara presente en el escenario. (fuente: [10])

usar este conjunto de datos permita un desafío más complejo para probar diferentes algoritmos de ReID.

En la tabla 3.2 se muestra una comparación entre este conjunto de datos con otros de tipo deportivo y de ReID.

Como se mencionó anteriormente, el conjunto de datos TGC20ReID contiene imágenes de corredores de la Transgrancanaria (TGC) 2020 Classic, una carrera de 128KM que recorre la isla de Gran Canaria de norte a sur. Para la carrera de 2020, 435 de los 677 terminaron la carrera a tiempo. Un conjunto importante de ellos fueron registrados en los cinco puntos de registros (RP1, RP2, RP3, RP4, RP5) a lo largo de la carrera. Dos registros (RP1 y RP2) ocurren durante la noche y el resto durante el día. Debido a la duración de 30 horas de la carrera, solo se registraron a 109 corredores en todos los puntos.

A diferencia de otros conjuntos de datos deportivos, en el TGC20ReID un mismo corredor puede aparecer en condiciones muy diferentes por la duración de la carrera y las características de los RP (Fig 3). Otra diferencia notable entre los conjuntos de datos existentes y los nuestros es la variación de la iluminación desde la noche hasta el día, e incluso cambios de iluminación en los RP debido a sombras relacionadas con edificios, montañas, etc. Además la apariencia del corredor no se mantiene del todo a lo largo de la carrera

Dataset	#Sports	Annotation	Media	#Sequences	Conditions	#Identities	Quality
APIDIS [24]	Basketball	Trajectories & Actions	Video Same locations	1	Indoor	12	1600x1200 22fps
CVBase 06 [25]	3 types	Trajectories & Actions	Video Same locations	4	Indoor	22	368x288 384x28825fps
Leeds Sports Pose [26]	8 types	Joints locations	Imagenes	2000	Indoor & Daylight	N.A.	Varying
Sports videos in the wild [27]	30 types	Actions locations	Video	4200	Indoor & Daylight	N.A.	480x270 480x36030fps
UIUC Sports Event [28]	8 types	Actions	Imagenes	1579	Indoor & Daylight	N.A.	Varying
Marathon RBN [29]	Marathon	Bib numbers	Imagenes	9706	Daylight	N.A.	N.A.
Market-1501 [10]	No	Identities	Imagenes	-	Daylight	1501	N.A.
MARS [30]	No	Identities	Video	20k	Daylight	1261 2-6 cameras	1080x1920 640x480
TGC20ReId	Trail race	Identities	Imagenes	4373	Day & Nightlight	416 2 5 locations	1920x1080

Tabla 3.2: Características conjunto de datos (fuente: [11])

debido a los cambios de ropa (Fig 4). Todo esto genera que el conjunto de datos sea una recopilación de datos muy desafiante para probar diferentes algoritmos de ReID.



Figura 3.2: Puntos de registros (fuente: [11])

3.3. Preparación conjunto de datos

El conjunto de datos Market1501 posee una nomenclatura específica que consta de varias partes separadas por guiones bajos. Estas partes son:

- 'idad': Un número de 4 dígitos que representa el ID de la persona en la imagen.
- 'c': Indica la cámara que capturó la imagen.



Figura 3.3: Variaciones de un corredor (fuente: [11])

- 's': Un número que indica la secuencia de imágenes, en este caso siempre es 1.
- 'fotograma': el cual indica el número de fotograma de la imagen en la secuencia.
- 'recuadro delimitado': Un número de 2 dígitos que indica el ID del recuadro delimitador de la persona en la imagen, en este caso siempre es 1.

Para adaptar el conjunto deportivo TGC20Reid a esta nomenclatura, se procederá a cambiar los tamaños y renombrar las imágenes del conjunto de datos siguiendo la susodicha estructura de nombres. Por ejemplo, la imagen “1_Arucas_frame_01_06_58_000.jpg” pasará a llamarse como “0001_c1s1_010658000_01.jpg”.

TGC	Market
Arucas	c1s1
Teror	c2s1
Preso de Hornos	c3s1
Ayagauras	c4s1
Parque sur	c5s1

Tabla 3.3: Transformación nombres conjunto de datos



Figura 3.4: Imagen tras procesado

Una vez modificado los nombres, se procede a replicar el formato de carpetas de Market1501, es decir, se repartirán las imágenes en “bounding_box_train”, “bounding_box_test” y “query”, donde “bounding_box_test” tendrán las imágenes de Ayagaures (c4s1) de los corredores que aparecen en todos los puntos de registros, “query” contendrá las imágenes de Parque sur (c5s1) de los mismos corredores, y el resto de imágenes del conjunto de datos se usará en el “bounding_box_train” para el entrenamiento. Una vez que hemos preprocesado el conjunto de datos, se procederá a realizar tres tipos de pruebas distintos para analizar los resultados que se obtendrán.

Capítulo 4

Herramientas

A continuación se comentarán las herramientas utilizadas en este proyecto. El cual se ha usado como lenguaje de programación Python a través de Jupyter Notebook. Además del entorno de Virtual Box para trabajar con librerías específicas de Linux como Bob.

En este proyecto se han utilizado diversas herramientas que han permitido llevar a cabo la tarea de manera eficiente y efectiva.

4.1. Python

Python [31] es un lenguaje de programación de alto nivel, interpretado y de propósito general, que se ha convertido en uno de los lenguajes de programación más populares en la actualidad. Es utilizado para una variedad de aplicaciones, incluyendo desarrollo web, análisis de datos, inteligencia artificial y aprendizaje automático, entre otros.

Una de las principales características de Python es su sintaxis simple y legible, lo que lo convierte en un lenguaje fácil de aprender para los principiantes. Además, es un lenguaje interpretado, lo que significa que no es necesario compilar el código antes de ejecutarlo, lo que ahorra tiempo y permite una mayor flexibilidad en el desarrollo.

Python también cuenta con una gran cantidad de bibliotecas y módulos, lo que lo hace altamente extensible y personalizable. Algunas de las bibliotecas más populares son NumPy, Panda, Matplotlib, SciPy, TensorFlow y PyTorch, que son ampliamente utilizados en el análisis de datos y el aprendizaje automático.

Python es compatible con múltiples sistemas operativos, incluyendo Windows, Linux y Mac OS, lo que lo convierte en una opción popular para el desarrollo multiplataforma. Además, es un lenguaje de código abierto, lo que

significa que el código fuente está disponible para su uso y modificación libre.



Figura 4.1: Logo Python

4.1.1. Jupyter Notebook

Jupyter Notebook [32] es un entorno de programación interactivo que permite la creación y el uso de documentos interactivos llamados cuadernos Jupyter. Es una herramienta popular entre los científicos de datos y los investigadores porque permite la creación de informes reproducibles y la exploración interactiva de datos.

El cuaderno Jupyter permite integrar en un mismo documento tanto código como texto, gráficos, imágenes y otros recursos. Esto permite que los usuarios puedan escribir código, ejecutarlo y ver los resultados en tiempo real en el mismo documento, lo que facilita la exploración de datos y el análisis.

Jupyter Notebook soporta varios lenguajes de programación, incluyendo Python, R, Julia y MATLAB. Cada cuaderno está organizado en celdas, que pueden contener código, texto o gráficos. Los usuarios pueden ejecutar cada celda de forma independiente, lo que permite iterar y depurar el código de manera interactiva.

Jupyter Notebook es una herramienta muy útil para enseñar programación y análisis de datos, ya que permite crear ejercicios interactivos y ejemplos que los estudiantes pueden ejecutar y modificar en tiempo real. También es una herramienta popular en la industria, donde se utiliza para la exploración de datos y la creación de informes interactivos.



Figura 4.2: Logo jupyter

4.1.2. VirtualBox

VirtualBox [33] es un software de virtualización gratuito y de código abierto que permite a los usuarios crear y ejecutar máquinas virtuales en su computadora. Admite la virtualización de sistemas operativos invitados como Windows, Linux, macOS, Solaris y otros sistemas operativos.

Una de las principales ventajas de VirtualBox es que es multiplataforma, lo que significa que puede ejecutarse en diferentes sistemas operativos, como Windows, Linux y macOS. Además, es muy fácil de instalar y configurar, lo que lo hace ideal para usuarios que buscan una solución de virtualización rápida y fácil.

VirtualBox utiliza un enfoque de virtualización de tipo 2, lo que significa que se ejecuta en un sistema operativo anfitrión y crea máquinas virtuales que se ejecutan como procesos en ese sistema operativo anfitrión. Esto es diferente de la virtualización de tipo 1, que se ejecuta directamente en el hardware y es más adecuada para entornos empresariales de alto rendimiento.

Con VirtualBox, los usuarios pueden crear máquinas virtuales desde cero o importar máquinas virtuales existentes de otros software de virtualización. Además, es posible configurar la cantidad de recursos que se asignan a cada máquina virtual, como RAM, espacio en disco y procesadores.

Otra característica interesante de VirtualBox es la capacidad de crear y administrar instantáneas de máquinas virtuales, lo que permite a los usuarios guardar el estado de una máquina virtual en un momento dado y volver a él más tarde. Esto es útil para pruebas de software y experimentación, ya que permite a los usuarios probar diferentes configuraciones y revertir fácilmente a un estado anterior si algo sale mal.



Figura 4.3: Logo jupyter

4.1.3. Librería Bob

Junto a VirtualBox, se ha utilizado la librería Bob [34] para la elaboración de las gráficas CMC. Bob es una librería de código abierto que proporciona una colección de herramientas para el procesamiento de señales, imágenes y reconocimiento de patrones. Ofrece una amplia gama de algoritmos para la extracción de características, el análisis de datos y la clasificación, lo que la convierte en una herramienta muy útil para el análisis de imágenes en este proyecto.



Figura 4.4: Librería bob

4.1.4. Microsoft Teams

Microsoft Teams [35] es una plataforma de colaboración y comunicación en línea desarrollada por Microsoft. La plataforma está diseñada para facilitar la colaboración en equipo, la comunicación y la organización de proyectos. Con Microsoft Teams, los usuarios pueden comunicarse y trabajar juntos en tiempo real, desde cualquier lugar y en cualquier momento.

Entre las principales características de Microsoft Teams se incluyen:

- Chat en tiempo real: los usuarios pueden comunicarse entre sí mediante mensajes de chat en tiempo real.

- Videoconferencias: Microsoft Teams ofrece la posibilidad de realizar videoconferencias con otros usuarios.
- Compartir pantalla: los usuarios pueden compartir su pantalla con otros miembros del equipo durante las videoconferencias.
- Colaboración en documentos: los usuarios pueden trabajar juntos en documentos en tiempo real, lo que facilita la colaboración y la edición conjunta de documentos.
- Integración con otras herramientas de Microsoft: Microsoft Teams se integra con otras herramientas de Microsoft, como Office 365 y SharePoint, lo que permite a los usuarios acceder a estas herramientas directamente desde la plataforma.
- Personalización: Microsoft Teams ofrece opciones de personalización para adaptarse a las necesidades y preferencias de cada usuario.

Microsoft Teams es una herramienta muy útil para equipos de trabajo, ya que permite una comunicación fluida y una colaboración efectiva en proyectos y tareas. Además, ofrece una interfaz fácil de usar y una amplia gama de características para mejorar la eficiencia y la productividad de los equipos.



Figura 4.5: Logo jupyter

Capítulo 5

Experimentos

En este apartado se contará acerca de los experimentos realizados a lo largo de este proyecto y de los resultados acerca de estos. Como punto inicial se puede indicar que se realizarán tres experimentos, en los cuales se evaluarán métodos diferentes, el “zero shot learning“, el entrenamiento desde cero y el “transfer learning“, los cuales ya se han tratado anteriormente. Se basará en el uso del conjunto TGC20ReID para su evaluación al cual se le han realizado en la sección 3.3, donde se ha elaborado las diferentes carpetas siguiendo el orden del que esta diseñada el conjunto de datos de Market1501, “query“ con 329 imágenes, “bounding box test“ con 208 imágenes y “bounding box train“ donde esta última tendrá el mayor número de individuos con 817 imágenes .

Para cada uno de los experimento se obtendrán sus respectivos pesos óptimos para cada uno de los métodos de entrenamiento, se procede a calcular los “embedding“ de las imágenes de los corredores que aparecen en todos los puntos, esto consiste en representar las susodichas imágenes con vectores de números, lo que ha permitido mejorar significativamente las tareas de descubrimiento de conocimiento y de recomendación de contenido. Con estos vectores, podemos determinar la matriz de proximidad de las imágenes y calcular una curva de coincidencia acumulada (CMC) la cual es una métrica de evaluación de uso común en el campo de la Visión por Computador, particularmente para tareas de ReID de personas. Consiste en la representación de la probabilidad de que un modelo coincida correctamente con una persona objetivo en un conjunto de imágenes recuperadas clasificadas, en función del rango (rank), proporciona una evaluación completa del rendimiento del modelo. En concreto, se mirarán los rank 1, 5, 10 y 20:

El rank 1 consiste porcentaje de predicciones en las que la predicción coincide con la etiqueta correcta. Se trata de un sistema de precisión que consiste en tomar la cantidad total de predicciones correctas y divídala por

la cantidad de puntos de datos en el conjunto de datos. Mientras que el resto de rank son una extensión de la precisión del rank 1, donde en lugar de preocuparnos solo por la predicción de un clasificador, tendremos en cuenta las de 5, 10 y 20 **predicciones principales de la red respectivamente.**

Para esto, se utilizan las matrices de distancia que se obtuvieron anteriormente, en este caso se revisarán las matrices que relacionan Ayagaures y Parque Sur, ya que son las imágenes en las que se pueden ver con mas claridad a los corredores.

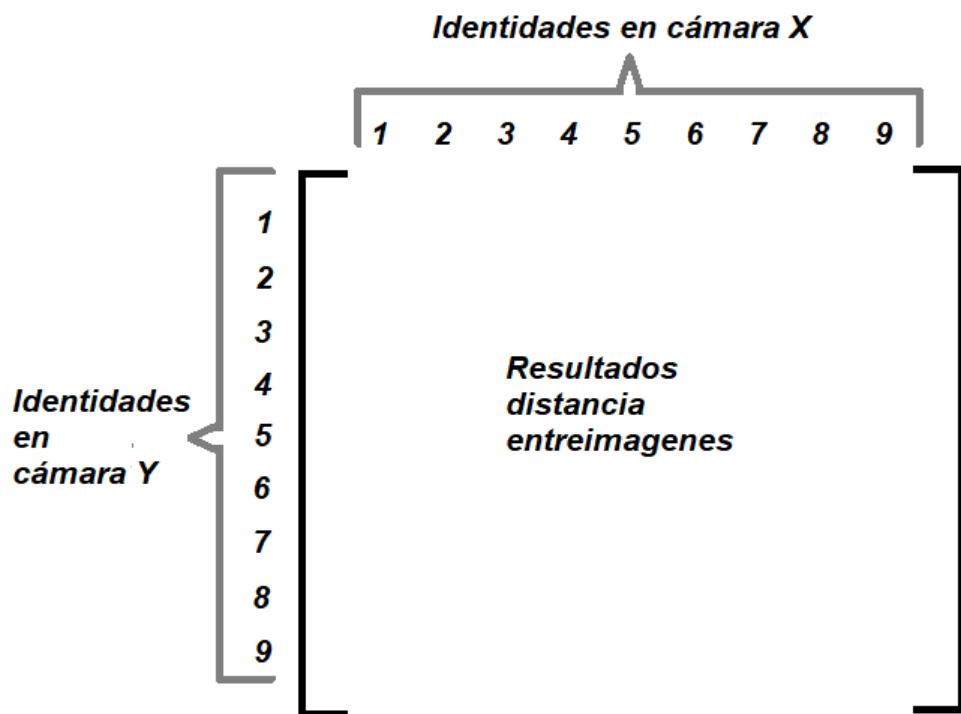


Figura 5.1: Esquema matriz de distancias

El análisis se hará a través de 3 técnicas diferenciadas entre si:

5.1. Zero shot learning

Como se ha hablado anteriormente, el ZSL [18] es una técnica de aprendizaje automático que permite que un modelo pueda clasificar objetos de clases nunca antes vistas, sin haber recibido entrenamiento específico para esas clases en particular debido a la capacidad del modelo para entender y utilizar información descriptiva sobre las clases, como atributos o características asociadas a ellas.

Su objetivo es ampliar la capacidad de clasificación de los modelos, permitiéndoles reconocer y clasificar objetos nuevos y desconocidos sin tener que ser entrenados específicamente para cada clase.

El primer método consistió en utilizar esta técnica con el conjunto TGC20ReId, empleando los pesos preentrenados de Market1501.

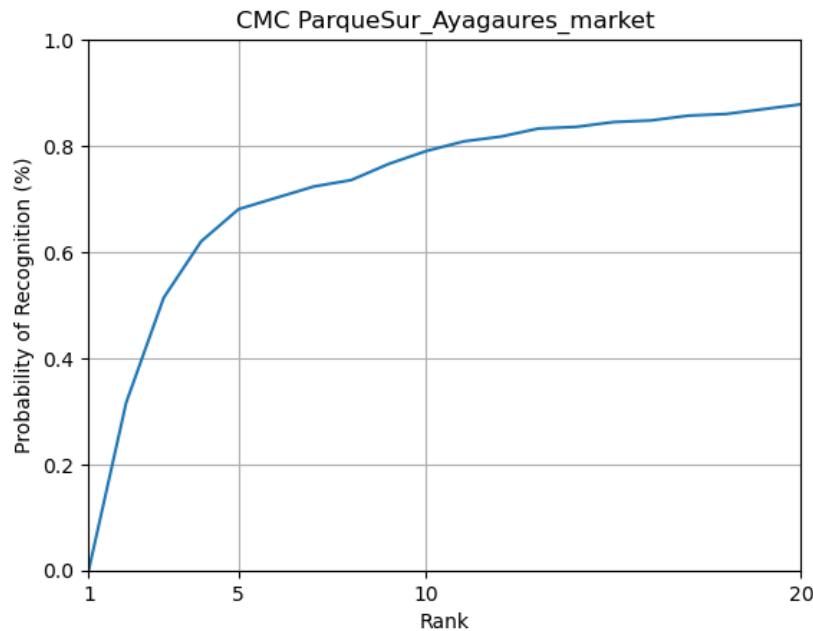


Figura 5.2: CMC zero shot learning

5.2. Entrenamiento desde cero

El segundo método implica entrenar AlignedReId++ únicamente con el conjunto TGC20ReId, y dado que el conjunto de datos solo contiene una serie de corredores que aparecen en todos los puntos de captura de la imagen, y es importante asegurar que la red pueda identificar correctamente a cualquier corredor nuevo que se añada en el futuro, se colocaran todos los corredores que aparecen en todos los puntos de captura en la carpeta de test, mientras que el resto de las imágenes se utilizarán para entrenar la red. Cabe mencionar que estas imágenes de entrenamiento aparecen en varios puntos de captura, pero no en todos. A continuación se evaluará los resultados utilizando la técnica de AlignedReid++ desde el inicio. Esto nos permitirá determinar los resultados trabajando solo con nuestro conjunto de datos.

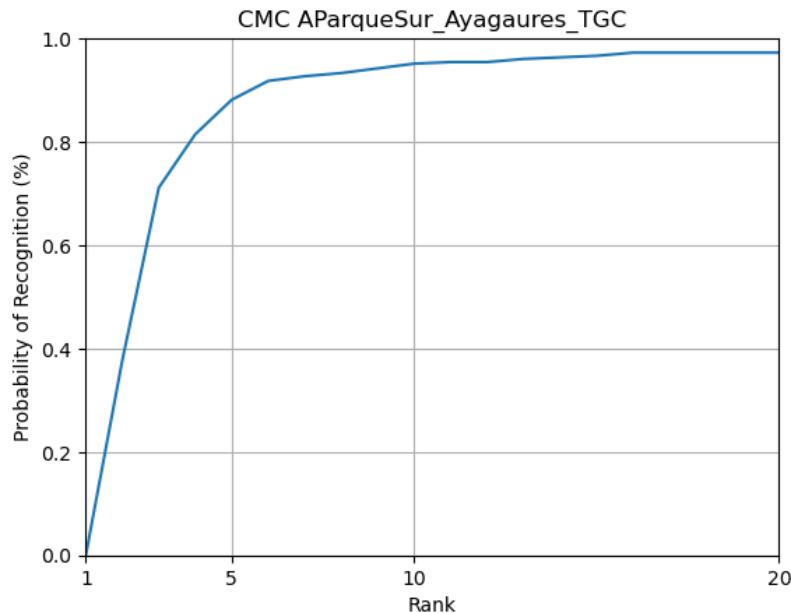


Figura 5.3: CMC entrenado con TGCReId

5.3. Transfer learning

En este enfoque, se empleará la técnica de “transfer learning” anteriormente mencionado la cual, a modo de resumen, se trata de una técnica que consiste en utilizar un modelo pre-entrenado en una tarea específica y adaptarlo para otra tarea relacionada en vez de entrenar un modelo desde cero. Utilizando un modelo ya entrenado en un conjunto de datos grande y complejo, para luego ajustar y re-entrenar el modelo en un conjunto de datos más pequeño y específico.

En este experimento, aprovechando los pesos previamente entrenados en el conjunto de datos de Market1501. Mediante esta estrategia, se espera lograr una mejora significativa en los resultados al utilizar las características previamente aprendidas por Market1501 como punto de partida en el entrenamiento del modelo con el conjunto de datos TGCReId. Al hacerlo, se espera que el modelo sea capaz de capturar patrones más complejos y representativos, mejorando así su capacidad de reconocimiento y clasificación de personas de manera más precisa y eficiente.

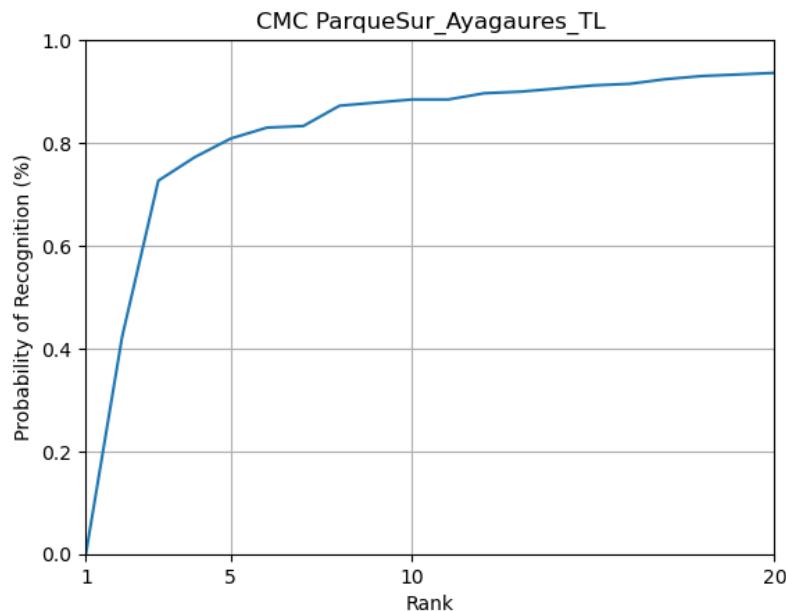


Figura 5.4: CMC entrenado con TGC

5.4. Resultados de los experimentos

A continuación se presenta una tabla donde se pueden contemplar los resultados obtenidos para los experimentos previos, “zero shot learning”, entrenamiento desde cero y “transfer learning”, para los distintos Ranks probados, además de una gráfica CMC comparativa donde se pueden apreciar mejor las variaciones entre los experimentos.

Experimentos	mAP	Rank-1	Rank-5	Rank-10	Rank-20
Zero Shot Learning	45.2 %	41.9 %	66.0 %	79.0 %	90.0 %
Entrenamiento desde cero	79.3 %	77.8 %	93.9 %	97.3 %	98.2 %
Transfer learning	76.0 %	76.6 %	90.3 %	93.6 %	97.3 %

Tabla 5.1: Experimentos

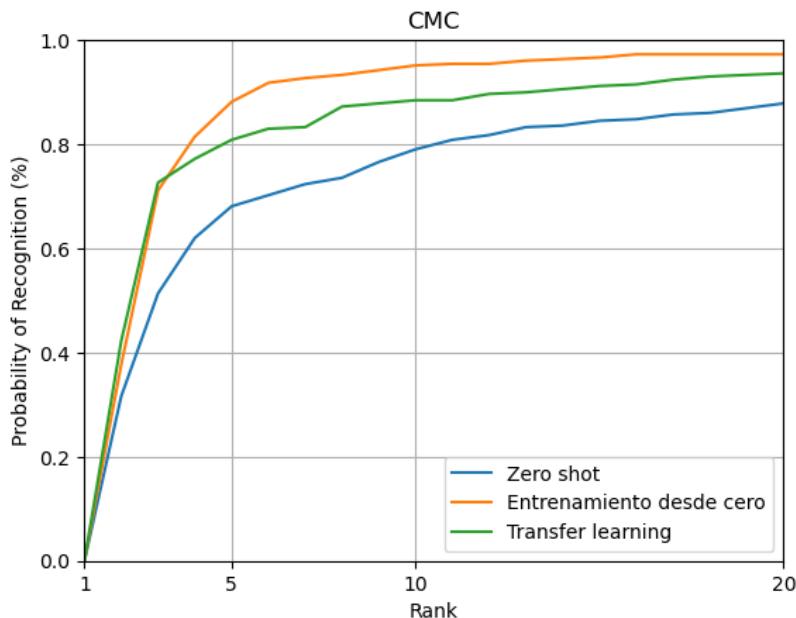


Figura 5.5: CMC comparativa

Como podemos apreciar en la imagen 5.5 se puede apreciar que los resultados obtenidos desde el entrenamiento del conjunto de datos TGCReId desde cero han dado unos resultados muchísimos mejores que el resto de experimentos realizados con respecto al mismo conjunto de datos. Como nota

informativa, en el artículo original de AlignedReID++ se puede ver los resultados obtenidos desde un entrenamiento desde cero con distintos conjuntos de datos [5.6](#), en los que se puede apreciar que se obtuvieron mejores resultados para Rank-1 que los obtenidos en las pruebas anteriores, tanto en el conjunto DukeMTMC como en el de Market1501, siendo este último el cual se han llegado a utilizar los pesos obtenidos para generar con los experimentos.

Method		Market1501		DukeMTMC		CUHK03		MSMT17	
		mAP	r = 1	mAP	r = 1	mAP	r=1	mAP	r=1
SVD [47]	ICCV17	62.1	82.3	56.8	76.7	37.3	41.5	-	-
IRNN [48]	PR2018	65.3	86.2	-	-	-	-	-	-
PDC [6]	ICCV17	63.4	84.1	-	-	-	-	29.7	58.0
PCE&ECN [4]	CVPR18	69.0	87.0	62.0	79.8	27.3	30.2	-	-
GLAD [5]	ACMMM17	73.9	89.9	-	-	-	-	34.0	61.4
MGCAM [49]	CVPR18	74.3	83.8	-	-	46.9	46.7	-	-
AML [50]	PR2018	74.1	89.5	-	-	-	-	-	-
AWTL [13]	CVPR18	75.7	89.5	63.4	79.8	-	-	-	-
MLFN [51]	CVPR18	74.3	90.0	62.8	81.0	47.8	52.8	-	-
HA-CNN [52]	CVPR18	75.7	91.2	63.8	80.5	38.3	41.7	-	-
Baseline		70.9	86.4	61.0	76.9	56.2	57.3	37.3	61.5
AlignedReID++		77.6	91.0	68.0	80.7	59.7	60.9	40.6	66.3
AlignedReID++(LS)		79.1	91.8	69.7	82.1	59.6	61.5	43.7	69.8
CamStyle(RK) [53]	CVPR18	71.5	89.5	57.6	78.3	-	-	-	-
DaRe(RK) [54]	CVPR18	82.0	88.3	74.5	80.4	63.6	62.8	-	-
GSRW [55]	CVPR18	82.5	92.7	66.4	80.7	63.6	62.8	-	-
AlignedReID++(RK)		88.5	92.0	81.2	85.2	70.7	67.6	-	-
AlignedReID++(LS,RK)		89.4	92.8	82.8	86.2	70.7	67.9	-	-

Figura 5.6: Resultados comparando distintos métodos (fuente: [\[8\]](#))

Podemos considerar que una de las razones de esto puede deberse a una mayor complejidad en nuestro conjunto, esto se puede apreciar en el experimento del “zero shot learning“, el cual se ha utilizado la evaluación directamente con los pesos ya mencionado, han dado los peores resultados, probablemente a que no se encontraban todas las características necesaria para obtener un mayor rendimiento, situación que no se ha visto afecta en el entrenamiento desde cero y el “transfer learning“ donde se han detectado características más discriminativas permitiendo aumentar la correcta identificación de los corredores

Capítulo 6

Conclusión y trabajos futuros

En este Trabajo de Fin de Máster, hemos logrado alcanzar los objetivos propuestos inicialmente, centrándonos en evaluar la eficiencia de la reidentificación de personas utilizando AlignedReID++, basándonos en distintos tipos de métodos, como el “zero shot learning”, el “transfer learning” y entrenado directamente con el conjunto de datos a examinar . A lo largo de este capítulo, compartiremos nuestras impresiones sobre el proyecto y analizaremos la importancia de los resultados obtenidos, así como los métodos explorados durante el desarrollo de esta investigación.

6.1. Valoración personal

Después de analizar los resultados, se ha llegado a que el “zero shot learning” ha dado los peores resultados con un 41,9 % en Rank-1, en comparación con los resultados obtenidos con el entrenamiento utilizando directamente el conjunto de datos a explorar y el método de “transfer learning” con un 77.8 % y un 76.6 % respectivamente. Estos resultados indican cómo la aplicación de estas técnicas ha permitido detectar patrones previamente no identificados en el conjunto original de datos de Market1501, lo cual sugiere que aún existen oportunidades para perfeccionar y desarrollar estas técnicas antes de su implementación completa.

Es importante destacar que, si bien hemos logrado una mejora significativa mediante el uso del método de “transfer learning”, también hemos observado mejoras notables con el entrenamiento utilizando directamente el conjunto de datos a explorar. Esto podría deberse al hecho de que el conjunto de datos utilizado en nuestras evaluaciones no era lo suficientemente amplio como para brindar una representación completa de los patrones y características subyacentes. Por lo tanto, es crucial continuar investigando y

ampliando el conjunto de datos para seguir mejorando nuestros resultados.

En conclusión, este estudio ha demostrado el potencial y la efectividad del “transfer learning” y el entrenamiento utilizando directamente el conjunto de datos a explorar en la reidentificación de personas. Sin embargo, aún hay margen para perfeccionar estas técnicas, especialmente a través de la ampliación del conjunto de datos y la exploración de enfoques adicionales. Estos esfuerzos continuarán impulsando avances significativos en la mejora de la precisión y robustez de los sistemas de reidentificación de personas en diversos escenarios y aplicaciones.

6.2. Mejoras futuras

En este sentido, se recomienda continuar con la exploración de diversas técnicas y algoritmos de aprendizaje profundo, así como la expansión del conjunto de datos, para poder mejorar la precisión y eficacia del modelo. Además, se sugiere llevar a cabo pruebas en situaciones más complejas y variadas, para poder evaluar el rendimiento del modelo en diferentes entornos y situaciones. En definitiva, el uso de técnicas de “transfer learning” y entrenamiento utilizando directamente el conjunto de datos a explorar demuestra ser una estrategia prometedora en el campo de la ReID de personas, y su potencial solo puede seguir creciendo con el avance de la tecnología y la investigación en esta área.

Es posible que el conjunto de datos utilizado para las evaluaciones haya sido limitado en su diversidad y tamaño, lo que podría no haber capturado completamente la amplia gama de patrones y características subyacentes presentes en los datos del mundo real. Por lo tanto, es importante reconocer la necesidad de seguir investigando y realizar proyectos futuros.

6.2.1. Ampliar conjunto de datos

Una posible mejora para el futuro consistiría en ampliar el conjunto de datos con el fin de obtener resultados más sólidos y generalizables. Además, es recomendable llevar a cabo pruebas en situaciones más complejas y variadas para evaluar el rendimiento del modelo en diferentes entornos y situaciones. El uso de técnicas de “transfer learning” y entrenamiento utilizando directamente el conjunto de datos a explorar demuestra ser una estrategia prometedora en el campo de la ReID de personas, y su potencial solo puede seguir creciendo con el avance de la tecnología y la investigación en esta área.

Una estrategia recomendada es continuar explorando diversas técnicas y algoritmos de aprendizaje profundo que permitan extraer y modelar de ma-

nera efectiva los patrones intrínsecos en los datos. Esto implica aprovechar la capacidad de los modelos preentrenados y aplicar “transfer learning”, utilizando arquitecturas y pesos de modelos previamente entrenados en conjuntos de datos masivos y diversos.

Además, es crucial ampliar y diversificar el conjunto de datos utilizado en el entrenamiento y la evaluación del modelo. Esto puede lograrse mediante la recopilación de más datos de diferentes fuentes, asegurándose de abarcar una amplia variedad de escenarios y condiciones. También es recomendable considerar técnicas de aumento de datos, como la transformación y combinación de imágenes existentes para aumentar la variabilidad y la representatividad de los datos de entrenamiento.

Al ampliar y mejorar el conjunto de datos, se espera obtener una representación más completa y precisa de las características y patrones presentes en los datos del mundo real. Esto, a su vez, puede conducir a modelos más robustos, con mejor capacidad de generalización y mayor eficacia en diversas aplicaciones y escenarios.

6.2.2. Aprendizaje no supervisado

Otras opción interesante que se puede considerar para mejorar aún más el modelo puede ser la exploración de técnicas de aprendizaje no supervisado para la extracción de características. Estas técnicas permiten que el modelo aprenda patrones y características relevantes de los datos sin la necesidad de etiquetas explícitas.

El aprendizaje no supervisado ofrece enfoques innovadores para descubrir patrones ocultos, estructuras y relaciones dentro de los datos. Al aplicar técnicas de aprendizaje no supervisado, el modelo puede aprender de forma automática y encontrar representaciones latentes o características significativas que pueden ser útiles para clasificar, agrupar o visualizar los datos.

Al explorar y aplicar técnicas de aprendizaje no supervisado en el contexto del modelo, se pueden obtener beneficios significativos, como una mejor comprensión de los datos, una representación más compacta y expresiva de las características y una mayor capacidad para manejar datos no etiquetados o escasamente etiquetados.

6.2.3. Reducción de ruido

Otro investigación a futuro puede ser la integración de técnicas de eliminación de ruido para mejorar la calidad de los datos utilizados en el entrenamiento del modelo. La presencia de ruido en los datos puede afectar

negativamente el rendimiento y la precisión del modelo, ya que introduce información incorrecta o irrelevante que puede conducir a decisiones erróneas.

La integración de técnicas de eliminación de ruido en el proceso de entrenamiento del modelo puede ayudar a mejorar la calidad de los datos utilizados, lo que a su vez puede llevar a una mejor capacidad de generalización y a la obtención de resultados más precisos. Estas técnicas pueden ayudar a eliminar datos erróneos, datos faltantes o datos inconsistentes, lo que resulta en un conjunto de datos más limpio y confiable para el entrenamiento del modelo.

En conclusión, se han logrado avances significativos en la ReID de personas mediante la implementación de técnicas de “transfer learning” y entrenamiento utilizando directamente el conjunto de datos a explorar. No obstante, aún existen oportunidades para seguir investigando y mejorando las técnicas utilizadas en este ámbito.

Es importante destacar que el campo de la ReID de personas sigue evolucionando y ofreciendo nuevas oportunidades de investigación. La exploración de enfoques innovadores, como el aprendizaje por refuerzo y la interpretación de modelos, puede brindar una mayor comprensión de cómo el modelo toma decisiones y ayudar a identificar patrones subyacentes en los datos.

Al abordar estos aspectos ya mencionados se podrá alcanzar un mejor rendimiento del modelo y una mayor capacidad para enfrentar situaciones más complejas y variadas en la ReID de personas.

Bibliografía

- [1] Mang Ye, Jianbing Shen, Gaojie Lin, Tao Xiang, Ling Shao, and Steven C. H. Hoi. Deep learning for person re-identification: A survey and outlook, 2021.
- [2] DataScientest. <https://www.tibco.com/reference-center/what-is-a-neural-network>, 19 de Enero de 2023.
- [3] Wikipedia. https://en.wikipedia.org/wiki/File:Typical_cnn.png, 19 de Enero de 2023.
- [4] Diego Unzueta. <https://towardsdatascience.com/convolutional-layers-vs-fully-connected-layers-364f05ab460b>, 19 de Enero de 2023.
- [5] DataScientest. <https://datascientest.com/es/que-es-el-transfer-learning>, 19 de Enero de 2023.
- [6] Analytics Vidhya. <https://www.analyticsvidhya.com/blog/2022/12/know-about-zero-shot-one-shot-and-few-shot-learning/>, 2023.
- [7] DataScientest. <https://www.aidemed.cn/new/new-19-113.html>, 19 de Enero de 2023.
- [8] H. Luo, W. Jiang, X. Zhang, X. Fan, J. Qian, and C. Zhang. Alignedreid++: Dynamically matching local information for person re-identification. *Pattern Recognition*, 94:53–61, 2019. Cited By :123.
- [9] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 815–823, 2015.
- [10] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *Proceedings*

- of the IEEE International Conference on Computer Vision (ICCV), December 2015.
- [11] A. Penate-Sánchez, D. Freire-Obregón, A. Lorenzo-Melián, J. Lorenzo-Navarro, and M. Castrillón-Santana. Tgc20reid: A dataset for sport event re-identification in the wild. *Pattern Recognition Letters*, 138:355–361, 2020. Cited By :4.
 - [12] Bici Home. <https://bicihome.com/historia-de-la-vuelta-a-espana-desde-sus-inicios-hasta-hoy/>, 19 de Enero de 2023.
 - [13] eitb.eus. <https://www.eitb.eus/es/deportes/ciclismo/detalle/8934218/listado-de-ciclistas-participantes-de-vuelta-a-espana-2022-nombre-de-corredores-y-sus-dorsales/>, 19 de Enero de 2023.
 - [14] Ciclo21. <https://www.ciclo21.com/vuelta-espana-2022/>, 19 de Enero de 2023.
 - [15] Las Provincias. <https://www.lasprovincias.es/deportes/ciclismo/vuelta-espana/premios-vuelta-ciclista-espana-20220816173706-nt.html>, 19 de Enero de 2023.
 - [16] Nationale nederlanden. <https://www.sansilvestrevallcana.com/historia.php>, 19 de Enero de 2023.
 - [17] Corriendo voy. <https://corriendovoy.com/descalificados-por-transferir-su-dorsal-en-el-trail-valle-de-tena/>, 19 de Enero de 2023.
 - [18] Petru Potrimba. <https://blog.roboflow.com/zero-shot-learning-computer-vision/>, 2023.
 - [19] Xuan Zhang, Hao Luo, Xing Fan, Weilai Xiang, Yixiao Sun, Qiqi Xiao, Wei Jiang, Chi Zhang, and Jian Sun. Alignedreid: Surpassing human-level performance in person re-identification, 2017.
 - [20] Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. Person transfer gan to bridge domain gap for person re-identification, 2018.
 - [21] Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *CVPR*, 2014.
 - [22] Ergys Ristani, Francesco Solera, Roger S. Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set for multi-target, multi-camera tracking, 2016.

- [23] Pedro F. Felzenszwalb, Ross B. Girshick, David McAllester, and Deva Ramanan. Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1627–1645, 2010.
- [24] Fan Chen, Damien Delannay, and Christophe Vleeschouwer. An autonomous framework to produce and distribute personalized team-sport video summaries: A basketball case study. *Multimedia, IEEE Transactions on*, 13:1381 – 1394, 01 2012.
- [25] J. Pers, M. Bon, G. Vuckovic, Cvbase 06 dataset. <https://vision.fe.uni-lj.si/cvbase06/downloads.html>, 2006.
- [26] Sam Johnson and Mark Everingham. Clustered pose and nonlinear appearance models for human pose estimation. In *Proceedings of the British Machine Vision Conference*, pages 12.1–12.11. BMVA Press, 2010. doi:10.5244/C.24.12.
- [27] Seyed Morteza Safdarnejad, Xiaoming Liu, Lalita Udma, Brooks Andrus, John Wood, and Dean Craven. Sports videos in the wild (svw): A video dataset for sports analysis. In *Proc. International Conference on Automatic Face and Gesture Recognition*, Ljubljana, Slovenia, May 2015.
- [28] Li-Jia Li and Li Fei-Fei. What, where and who? classifying events by scene and object recognition. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8, 2007.
- [29] Kamlesh, Pei Xu, Yang Yang, and Yongchao Xu. Person re-identification with end-to-end scene text recognition. In Jinfeng Yang, Qinghua Hu, Ming-Ming Cheng, Liang Wang, Qingshan Liu, Xiang Bai, and Deyu Meng, editors, *Computer Vision*, pages 363–374, Singapore, 2017. Springer Singapore.
- [30] Liang Zheng, Zhi Bie, Yifan Sun, Jingdong Wang, Chi Su, Shengjin Wang, and Qi Tian. Mars: A video benchmark for large-scale person re-identification. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, pages 868–884, Cham, 2016. Springer International Publishing.
- [31] python. <https://www.python.org>, 19 de Enero de 2023.
- [32] Jupiter Notebook. <https://jupyter.org>, 19 de Enero de 2023.
- [33] virtualbox. <https://www.virtualbox.org>, 19 de Enero de 2023.

- [34] libreria bob. <https://www.idiap.ch/software/bob/docs/bob/bob.measure/v2.4.1/index.html>, 19 de Enero de 2023.
- [35] Microsoft Teams. <https://www.microsoft.com/es-es/microsoft-teams/group-chat-software>, 19 de Enero de 2023.