REIDENTIFICACIÓN DE VEHÍCULOS MEDIANTE TÉCNICAS DE DEEP LEARNING

Álvaro Ramajo Ballester[1]
aramajo@pa.uc3m.es

Jacobo González Cepeda[1]
100307736@alumnos.uc3m.es

José María Armingol Moreno[1]
armingol@ing.uc3m.es

Arturo de la Escalera Hueso[1]
escalera@ing.uc3m.es

[1] Laboratorio de Sistemas Inteligentes, Universidad Carlos III de Madrid
Avda de la Universidad, 30, 28911, Leganés, Madrid

https://doi.org/10.17979/spudc.9788497498418.1031

Resumen

El nivel de precisión de las redes neuronales profundas en tareas de percepción visual permite captar información crucial del entorno para futuros proyectos, como los vehículos autónomos y las ciudades inteligentes. Una de las posibilidades que permitiría este tipo de sistemas es el control y seguimiento de determinados vehículos sospechosos. Teniendo en cuenta el uso de esta tecnología por parte de la policía, se facilitaría el seguimiento de determinados coches bajo investigación. Con esta visión, el objetivo de este trabajo es el estudio del estado del arte actual de los métodos y el desarrollo de un sistema que resuelva dos tareas de forma eficiente: la caracterización visual y reidentificación de vehículos y la segmentación de matrículas y reconocimiento de caracteres. Esta doble identificación puede adaptarse a las condiciones ambientales, a la distancia del objetivo y a las capacidades y resolución de las cámaras. Para probar y validar este sistema, se ha creado un conjunto de datos personalizado para minimizar la diferencia entre el laboratorio y el entorno real.

Palabras clave: Reidentificación de vehículos, deep learning, smart cities.

1 INTRODUCCIÓN

Actualmente, el campo de la visión por ordenador está experimentando un auge imparable tanto en términos de desarrollo como de implementación de sistemas en la vida real. Según el Global AI in Computer Vision Market [26] el volumen de mercado de la IA en este campo se valoró en 7.040 millones de dólares en 2020 y se espera que alcance los 144.460 millones de dólares en 2028. Este mercado está siendo impulsado por la creciente necesidad de inspección de calidad y automatización, así como por la creciente demanda de sistemas de visión por ordenador.

Una de las razones de este incremento ha sido la aparición de técnicas de aprendizaje profundo. Gracias a ello, y en concreto a las redes neuronales convolucionales [8] y a los mecanismos de atención [25], la precisión de estos sistemas de visión es órdenes de magnitud superior a la de los algoritmos clásicos. Junto a este mejor rendimiento, la velocidad de respuesta de este tipo de visión artificial es considerablemente más rápida, con el uso de unidades de procesamiento gráfico (GPU). Esto permite implementar nuevos sistemas, como la percepción e interpretación del entorno en vehículos autónomos, el control de calidad en procesos industriales, la identificación de objetos y personas, el diagnóstico médico por imagen, la videovigilancia y las infraestructuras inteligentes en ciudades inteligentes, entre otros. Estos dos últimos campos constituyen el marco de este trabajo ya que las herramientas desarrolladas permiten tanto el procesamiento automático de imágenes en tareas de videovigilancia como el análisis de escenas de tráfico en infraestructuras inteligentes.

En cuanto a esta primera tarea, el uso cada vez más extendido de las cámaras de seguridad es una de las mejores herramientas disponibles en la actualidad para prevenir y combatir la delincuencia. Sin embargo, el aumento de la cantidad de datos disponibles conlleva un incremento lineal de las horas dedicadas a su análisis. Esto ha sido así tradicionalmente hasta la introducción de los sistemas de procesamiento automático de imágenes. Uno de los puntos de inflexión que aceleró la implantación de estos fue la investigación del atentado terrorista de Londres en 2017.

En cuanto a las aplicaciones de las ciudades inteligentes, el objetivo de la reidentificación de vehículos es identificar el mismo vehículo a través de múltiples cámaras, que pueden obtener imágenes desde diferentes perspectivas del vehículo. A través de una red de vigilancia ubicua, un sistema de reidentificación puede obtener rápidamente la ubicación y la hora del vehículo objetivo en el espacio visual que cubren. De este modo, el vehículo puede ser detectado, localizado y rastreado automáticamente a través de múltiples cámaras, lo que ahorra trabajo y dinero. Además, estos sistemas tienen muchas posibles aplicaciones prácticas, como la asistencia al aparcamiento, el seguimiento de vehículos sospechosos, la monitorización en directo o el seguimiento de vehículos con múltiples cámaras para la vigilancia urbana, lo que hace que este tipo de desarrollo sea crucial para la construcción de un sistema de transporte inteligente.

En este contexto, a pesar de los grandes desarrollos en el campo de la inteligencia artificial y la visión, existe una necesidad que aún no está totalmente cubierta. Esta carencia se refiere a un sistema que permita la caracterización y reidentificación (o búsqueda) de un vehículo que permita utilizar toda la información visual disponible. Para ello, se pretende procesar las imágenes tanto de las cámaras de tráfico de gran altura (caracterización visual) como de las cámaras adicionales más cercanas al tráfico rodado (identificación de matrículas), de forma que el análisis se adapte tanto a la resolución como a la posición y orientación de las cámaras. Este proyecto pretende facilitar el trabajo de futuras investigaciones policiales que requieran el análisis y seguimiento de sujetos en vehículos, ya que gracias a las cámaras ya instaladas en carreteras y autopistas, es posible reconocer la presencia de un vehículo de interés con gran precisión,

 https://doi.org/10.17979/spudc.9788497498418.1031

lo que de otro modo requeriría la visualización manual por parte de un agente.

2 ANTECEDENTES

La clasificación y detección de objetos con deep learning es una tarea que ha evolucionado y mejorado claramente en los últimos años. Gracias a estas técnicas, se puede alcanzar un nivel de precisión relativamente alto. Otra ventaja es su versatilidad, ya que los modelos pueden adaptarse a diferentes tareas modificando ligeramente la estructura de la red y los datos de entrenamiento. En el caso de la reidentificación de vehículos, debido a la naturaleza bimodal del sistema propuesto, se han examinado dos aspectos a la hora de confirmar la identidad del vehículo. El primero es el reconocimiento automático de matrículas (ALPR), mientras que el segundo modo trata de extraer las características visuales del coche completo y realizar una comparación de similitud entre el objetivo y la imagen del vehículo procesada.

Dentro del ALPR, existen dos enfoques principales en el estado del arte: los sistemas de reconocimiento multietapa o los de una sola etapa. Los sistemas multietapa realizan una extracción de la región de la imagen donde se encuentra la matrícula para segmentar los caracteres dentro de esa región y aplicar un OCR (Optical Character Recognition) sobre ellos. Para esta primera tarea, existen métodos clásicos basados en formas [29] y colores [5]. Una vez extraídas las regiones, los caracteres se segmentan mediante conectividad [17] o redes neuronales convolucionales [7]. Por último, los caracteres se reconocen mediante técnicas de emparejamiento [18] o redes probabilísticas [1], entre otras. En cambio, los métodos de una sola etapa suelen basarse en técnicas de aprendizaje profundo, como [11] y [27], utilizando redes VGG16. Este tipo de métodos pueden ser más rápidos y eficientes, ya que existe una correlación entre la detección y el reconocimiento, por lo que los modelos pueden compartir parámetros y reducir su tamaño y tiempo de inferencia [21].

En cuanto al segundo modo de reidentificación, es necesario realizar una extracción mucho más fina y rigurosa de las características, ya que las diferencias entre los distintos tipos de coches son mucho menores que entre los objetos comunes (diferentes colores, formas, etc.). Hay ejemplos de caracterización e identificación de vehículos en [16] y [9] con buenos resultados. Sin embargo, utilizan conjuntos de datos con imágenes muy similares en términos de posición e iluminación de los vehículos, como se muestra en la Figura 1. Esto sugiere que podría ser menos exportable a un escenario real en el que haya que analizar los coches que vienen de todas las direcciones.

Figura 1. Ejemplo de imágenes en [9]

Del mismo modo, también hay trabajos en los que se utilizan conjuntos de datos de gran variedad y cantidad de imágenes, como Stanford-Cars [6], VeRi-776 [14] y VeRi-Wild [15]. En la Figura 2 se muestra un ejemplo de las imágenes recogidas en este último.

Figura 2. Ejemplo de imágenes en [15]

Para estos casos, se han presentado algunas soluciones que mejoran generalmente los resultados en la reidentificación de vehículos [4] como con objetos genéricos [3]. Además, este último también proporciona una *toolbox* donde se pueden encontrar los modelos preentrenados para su uso.

3 ARQUITECTURA DEL MODELO

El sistema global se compone de cuatro modelos diferentes y se muestra en la Figura 3:

- YOLOv4 [2]: la imagen de entrada es procesada por el primer modelo, un detector de objetos, que proporciona los cuadros delimitadores de todos los vehículos presentes en la imagen.
- WPOD-Net + OCR-Net [22]. Una vez extraídas las regiones de los vehículos, la primera rama paralela realiza el reconocimiento de las matrículas. Si la confianza de la región de la matrícula está por encima del umbral, se aplica el reconocimiento de caracteres con OCR-Net, que produce la secuencia final de caracteres.
- FastReid [3]: la segunda rama paralela codifica la región de la imagen del vehículo según sus características visuales en un vector de 4096 características. La similitud entre diferentes

https://doi.org/10.17979/spudc.9788497498418.1031

vehículos se medirá en la distancia euclídea entre sus correspondientes vectores.

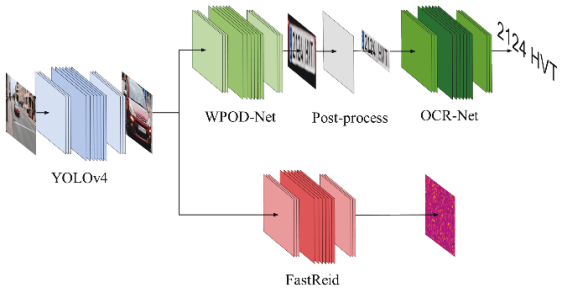

Figura 3. Arquitectura multi-red neuronal

3.1 DETECCIÓN DE VEHÍCULOS

La detección de las regiones de interés (ROI) tiene una doble vertiente en este trabajo. En primer lugar, se trata de indicar los fragmentos rectangulares dentro de la imagen que corresponden a un vehículo y, en segundo lugar, de buscar el recorte del vehículo que corresponde a su matrícula. En este primer caso, se ha abordado con la red YOLO [2] en su cuarta versión. Uno de los requisitos generales del sistema es la capacidad de ser implementado en tiempo real. Con esta premisa, se hace prácticamente imprescindible reducir al máximo el tiempo de inferencia en cada una de las redes tanto de detección como de reconocimiento. Esta es la principal razón para elegir YOLOv4 frente a otros modelos como YOLOv3 [20], EfficientDet [24], ATSS [28], ASFF [12] o CenterMask [10], como puede verse en la Figura 4.

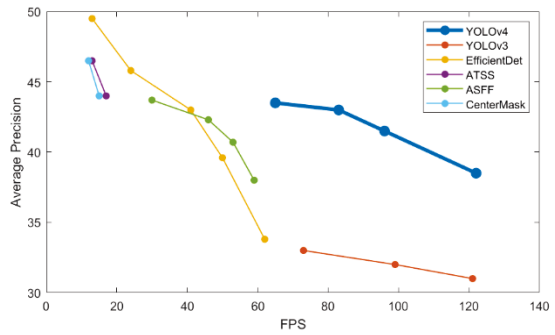

Figura 4. Velocidad de inferencia y *average precisión* en los modelos de detección

3.2 RECONOCIMIENTO DE MATRÍCULA

Las matrículas suelen ser objetos rectangulares y planos, que se colocan en los vehículos para su identificación. Para aprovechar su forma, se utiliza una red neuronal convolucional llamada Warped Planar Object Detection Network (WPOD-Net) [22]. El objetivo es que la red pueda aprender a detectar placas con diferentes perspectivas infiriendo los coeficientes de una transformada afín que "recompone" la imagen en una perspectiva rectangular como una vista frontal. La WPOD-NET se desarrolló a partir de los conocimientos de YOLO, SSD [13] y las redes de transformadas espaciales (STN). YOLO y SSD realizan una rápida detección y reconocimiento de múltiples objetos a la vez, pero no tienen en cuenta las transformaciones espaciales, generando sólo bordes delimitadores rectangulares para cada detección. Por el contrario, los STN pueden utilizarse para detectar regiones no rectangulares, pero no manejan múltiples transformaciones al mismo tiempo, realizando sólo una única transformación espacial en toda la entrada.

El proceso de detección mediante WPOD-NET se muestra en la Figura 5. Inicialmente, la red es alimentada por la salida redimensionada del módulo de detección de vehículos. Después de pasar por la red, se genera un mapa de características de 8 canales, que codifica las probabilidades y los parámetros de la transformación afín. A continuación, se crea un cuadrado imaginario de tamaño fijo alrededor del centro de una celda (m, n). Si la probabilidad del objeto para esta celda está por encima de un determinado umbral de detección, algunos de los parámetros inferidos se utilizan para construir una matriz afín que transforma el cuadrado ficticio en una región alrededor de la placa. De este modo, la región puede reconvertirse fácilmente en un objeto alineado horizontal y verticalmente.

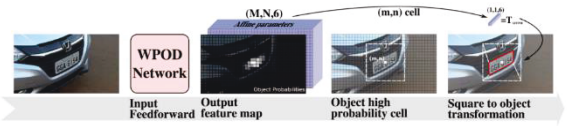

Figura 5. Diagrama de operación de WPOD-Net [22]

La Figura 6 muestra un ejemplo del procesamiento del recorte de la región de la matrícula.

Figura 6. Detección y rectificación de la matrícula

La segmentación y el reconocimiento de caracteres en la región rectificada de la matrícula (Figura 7) se realiza mediante OCR-Net [22], una red YOLO modificada [19], similar a lo comentado en la detección del contorno de la placa, pero con caracteres.

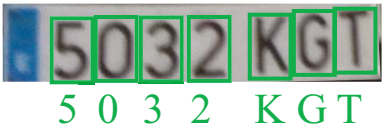

Figura 7. Reconocimiento de caracteres

https://doi.org/10.17979/spudc.9788497498418.1031

3.3 REIDENTIFICACIÓN VISUAL

Una vez completada y validada la identificación del vehículo con matrícula, se aborda el segundo modo de reidentificación. Esta reidentificación se realiza calculando la distancia euclídea entre los 4096 vectores de características del modelo de reconocimiento visual. Cada una de las imágenes procesadas tendrá su correspondiente vector de características que se compara con el vector de características del vehículo objetivo para calcular dicha distancia.

4 RECONOCIMIENTO VISUAL: ENTRENAMIENTO

Para abordar esta tarea, se ha probado una comparación entre algunos modelos del estado del arte y varios backbone entrenados. Para la reidentificación de vehículos, la librería FastReid [3] ofrece arquitecturas ya preentrenadas y optimizadas.
Con el objetivo de cumplir estos estándares, se han llevado a cabo diferentes estrategias de entrenamiento con la familia de redes neuronales EfficientNet [23]. La característica distintiva de este tipo de arquitectura que permite mejorar su rendimiento es la precisa escalabilidad de las dimensiones de la red. Como se muestra en la Figura 8, las redes convolucionales estándar intentan mejorar su rendimiento aumentando las dimensiones del mapa de características, es decir, su anchura (b); otras intentan aumentar el número de capas intermedias haciendo una red más profunda (c) o con imágenes de mayor resolución (d). Este tipo de modelos se ha utilizado como backbone durante el proceso de entrenamiento con capas de max pooling y convolucionales añadidas a su salida para el ajuste fino.

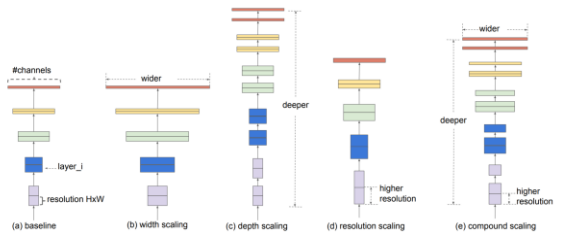

Figura 8. Escalabilidad de EfficientNet [23]

El entrenamiento inicial se ha realizado con el conjunto de datos Stanford-Cars [6], que se utiliza con frecuencia en el estado del arte actual, como se ha comentado en la sección anterior. Se ha realizado un primer entrenamiento de prueba con una versión reducida (aproximadamente el 10% del conjunto de datos), con el fin de ajustar el *dropout* y *learning rate*. Este primer valor se refiere a la proporción de redes neuronales en determinadas capas que se "apagan" aleatoriamente durante el entrenamiento. De este modo, la extracción de características se realiza mediante varias rutas (las neuronas "encendidas") y así el modelo se generaliza mejor. El *learning rate* se refiere a la velocidad de actualización de los pesos. Un valor reducido permite añadir muchos más pesos, pero a costa de un mayor tiempo de entrenamiento, por lo que es aconsejable ajustarlo de forma óptima. Todas estas pruebas se muestran en la Figura 9.

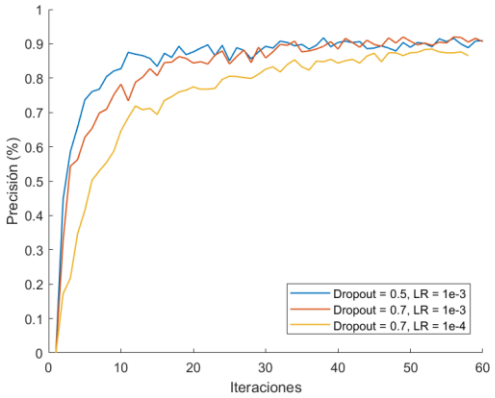

Figura 9. Efecto del *dropout* y el *learning rate*

En este gráfico se observan dos resultados notables. El primero es que las redes con un dropout de 0,7 generalizan ligeramente mejor que las de 0,5. A pesar de tardar algo más en las primeras epochs, la tendencia es a favor de las primeras, ya que con 0,5 se alcanza un máximo significativamente menor. Por otro lado, la tasa de aprendizaje más adecuada es 1e^(-3), ya que maximiza la precisión más rápidamente. Con esto, y varias pruebas más que evitamos incluir para no extender demasiado la demostración de entrenamiento, probamos el rendimiento de 3 versiones de la red EfficientNet: B0, B3 y B7. La salida se ha configurado utilizando un pooling global máximo para cada uno de los filtros de salida y una capa de clasificación densa con el dropout previamente ajustado. Los resultados se muestran en la Figura 10.

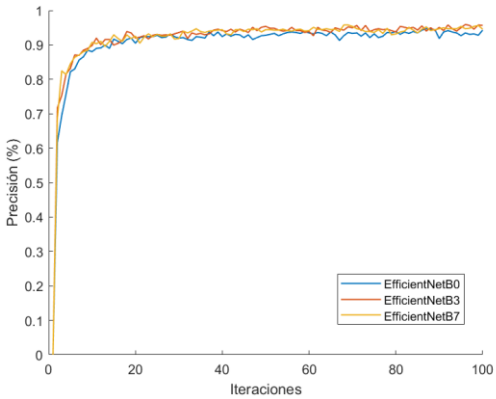

Figura 10. Comparativa de entrenamiento con Stanford-Cars

Como se muestra, el rendimiento es muy similar en los tres casos, por lo que al ser los modelos más grandes más pesados y lentos, tiene sentido adoptar el modelo B0 como red de caracterización, ya que se busca un sistema que funcione en tiempo real, si es posible.
Además, se ha realizado otro entrenamiento con las mismas redes, pero en este caso con el conjunto de datos VeRi-776 [14]. Para facilitar el proceso de entrenamiento, se han modificado ligeramente las clases de salida, ya que la red programada está concebida con el propósito de clasificación, para posteriormente eliminar la última capa softmax y codificar las imágenes con la salida de la penúltima capa. Por tanto, se han utilizado las mismas clases en el entrenamiento que en la validación, a diferencia de lo que se propone en el conjunto de datos original. Los resultados se muestran en la Figura 11. Esta modificación hace que los resultados sean numéricamente más favorables, aunque la evaluación comparativa se realizará en el siguiente apartado.

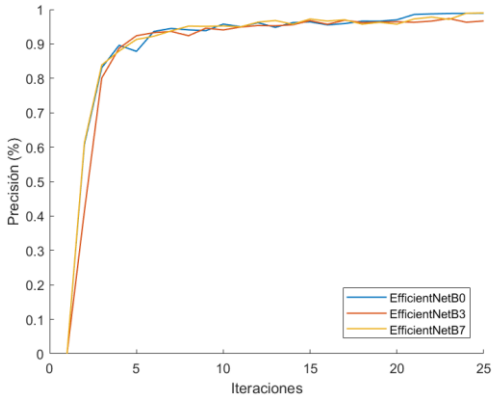

Figura 11. Comparativa de entrenamiento con VeRi-776

Como puede verse, la precisión es similar entre los tres modelos, por lo que se elige la opción EfficientNetB0 por las mismas razones que en el caso anterior. El mismo procedimiento se ha probado con un nuevo conjunto de datos etiquetados manualmente, y los resultados se mostrarán en la sección de resultados.

5 ADQUISICIÓN DE DATOS DE EVALUACIÓN

Para reducir al máximo la diferencia entre las condiciones de evaluación de este sistema y el entorno real de producción, se han creado una serie de conjuntos de datos propios para evaluar el rendimiento del sistema de reidentificación visual. En total hay dos tipos de datasets con características diferentes, en función de la similitud, la perspectiva y la iluminación de los vehículos, así como del número de cámaras de entrada.

5.1 DATASET AUTOVÍA

Este primer conjunto de datos contiene imágenes tomadas desde un poste de la autopista A-7 desde una posición elevada, oblicua y trasera. Este conjunto se caracteriza por una gran similitud entre tomas de la misma clase, con idéntica perspectiva, misma iluminación y sin oclusiones. Sirve como punto de partida en el proceso de evaluación, ya que es más sencillo y se pueden esperar resultados más favorables. Incluye un total de 458 imágenes correspondientes a 200 modelos de vehículos. En la Figura 12 se ofrece una muestra de estas.


Figura 12. Muestra del dataset autovía

5.2 DATASET INTERSECCIÓN

Este segundo grupo de imágenes incluye escenas de tráfico correspondientes a intersecciones en El Toyo, Almería. Este conjunto de datos está dividido en 2 lugares de grabación diferentes (v1 y v2) con multitud de perspectivas y oclusiones diferentes entre vehículos y con la vegetación. Cada una de las escenas ha sido capturada simultáneamente por dos cámaras (c1 y c2) y representa el mayor grado de dificultad, ya que representa el entorno normal de funcionamiento. Además, al tener dos fuentes de entrada, permite buscar vehículos anotados desde una cámara en la otra, con diferente perspectiva, que es el principal objetivo que se persigue en este trabajo. Agrupa un total de 1255 imágenes de 69 clases con criterios de anotación ligeramente diferentes. En la v1 se han incluido todas las apariciones de vehículos, incluso con vistas muy lejanas y parciales, mientras que en la v2 sólo se anotan los vehículos completos con un tamaño mínimo reconocible (Figura 13).


Figura 13. Muestra del dataset intersección

6 RESULTADOS

Como conclusión de este trabajo, las siguientes tablas muestran los resultados obtenidos en la fase de fine-

https://doi.org/10.17979/spudc.9788497498418.1031

tuning del sistema completo. En primer lugar, la tabla 1 muestra la precisión de los dos modelos entrenados (EfficientNetB0) frente a los modelos FastReid preentrenados en los conjuntos de datos públicos. Esta evaluación corresponde a la precisión de una prueba de par positivo-negativo. Cada par positivo-negativo se ha creado con cada imagen del conjunto de evaluación, una imagen de su clase (positiva) y una imagen aleatoria del resto de clases (negativa).

De estos resultados se puede extraer que el modelo FastReid preentrenado con VeRi-Wild, que es un conjunto de datos más grande y con menos restricciones que VeRi-776, es un mejor candidato.

Tabla 1. Precisión en datasets públicos

Modelo	Precisión		
	Stanford-Cars	VeRi-776	VeRi-Wild
EfficientNetB0 (Stanford-Cars)	**83.5 %**	62.5 %	72.7 %
EfficientNetB0 (VeRi-776)	59.1 %	77.2 %	79.7 %
FastReid (VeRi-776)	60.6 %	**96.8 %**	90.8 %
FastReid (VeRi-Wild)	67.6 %	90.5 %	**99.5 %**

Sin embargo, una vez realizada la evaluación con los conjuntos de datos propios, mucho más cercanos al entorno de producción real, las métricas favorecen al modelo FastReid preentrenado con VeRi-776, consiguiendo los mejores resultados, como se muestra en la tabla 2.

Tabla 2. Precisión en datasets propios

Model	Road	Precisión			
		Intersec. v1c1	Intersec. v2c1	Intersec. v1[1]	Intersec. v2[1]
EfficientNetB0 (Stanford-Cars)	85.1 %	73.6 %	84.1 %	62.5 %	61.9 %
EfficientNetB0 (VeRi-776)	96.6 %	88.2 %	91.6 %	71.5 %	78.1 %
FastReid (VeRi-776)	**97.9 %**	**94.0 %**	**96.6 %**	**87.8 %**	**91.5 %**
FastReid (VeRi-Wild)	96.7 %	90.9 %	89.7 %	78.8 %	82.5 %

[1] Dataset con dos cámaras de entrada

Además, se ha calculado la métrica rank@n. Estos valores se refieren a la presencia del modelo de búsqueda en las n posiciones más probables según las predicciones del modelo. Es decir, el rank@1 se refiere a la probabilidad de que el primer resultado del modelo corresponda al vehículo buscado, mientras que el rank@10 indica la probabilidad de que el vehículo se encuentre en los 10 primeros resultados.

La tabla 3 muestra, una vez más, que el modelo FastReid VeRi-776 alcanza los mejores valores para la métrica.

Tabla 3. Métrica Rank@1 y Rank@10 en datasets propios

Model	Rank@1		Rank@10	
	Intersec. v1	Intersec. v2	Intersec. v1	Intersec. v2
EfficientNetB0 (Stanford-Cars)	42.6 %	43.1 %	73.6 %	70.6 %
EfficientNetB0 (VeRi-776)	61.2 %	60.1 %	84.3 %	87.1 %
FastReid (VeRi-776)	**75.4 %**	**90.6 %**	**94.0 %**	**99.4 %**
FastReid (VeRi-Wild)	75.3 %	41.3 %	87.6 %	65.2 %

En cuanto al reconocimiento de matrículas, la red utilizada alcanza una precisión del 89,33%, lo que permite predecir la matrícula de los vehículos con un alto grado de exactitud. Estos valores se han extraído de [21] y se muestran en la tabla 4.

Tabla 4. Precisión en el reconocimiento de matrícula

Model	OpenALPR		SSIG	AOLP	
	EU	BR	Test	RP	CD-Hard
WPOD-Net+ OCR-Net	93.52 %	91.23 %	88.56 %	98.36 %	75.00 %

Por otro lado, en lo que respecta a la velocidad de inferencia, los modelos FastReid están muy optimizados, consiguiendo un tiempo de inferencia hasta 10 veces menor (tabla 5), lo que es ideal en el caso de la ejecución en tiempo real.

Tabla 5. Tiempo de inferencia en reidentificación visual

Model	Time (ms)
EfficientNetB0	28.76
EfficientNetB3	33.79
EfficientNetB7	46.63
FastReid (VeRi-776)	4.04
FastReid (VeRi-Wild)	3.28

Como se ha comentado en la sección introductoria, el reconocimiento de la matrícula sólo es posible en determinadas condiciones favorables. Sin embargo, siendo conscientes de que no todas las imágenes pueden reunir tales características, el segundo sistema de reconocimiento visual ofrece más flexibilidad en cuanto a las restricciones de funcionamiento y su versatilidad proporciona un rendimiento notable en situaciones más adversas. Al basarse en la extracción de las características visuales (forma y color) de todo el vehículo, es menos sensible a la distancia. Por lo tanto, cuando el reconocimiento de la matrícula es posible, amplía el rango de reconocimiento válido, por ejemplo, de 15 a 40 metros en la vista de la cámara de gran ángulo de cuatro carriles. También permite el seguimiento del vehículo para los siguientes

fotogramas de un vídeo en función de la similitud con los anteriores. Esto constituye un punto distintivo respecto a los enfoques clásicos, proporcionando una solución más robusta.

7 CONCLUSIÓN

Una vez validado, este sistema no sólo muestra un gran rendimiento en la identificación de vehículos objetivo, sino que además ofrece una mayor flexibilidad gracias a su módulo dual (visual y OCR), lo que le permite operar bajo diferentes características de entorno y resoluciones de cámara.

Una de las posibles líneas de investigación futuras derivadas de este trabajo es el desarrollo de este sistema como una herramienta totalmente funcional para su uso por parte de las Fuerzas de Seguridad del Estado. Como se ha mencionado en el apartado inicial, la automatización de estos procesos de visionado de cámaras liberará una gran cantidad de horas de los agentes dedicados a esta tarea. Asimismo, permitirá abarcar un número mucho mayor de fuentes de entrada, en este caso imágenes, para ampliar la búsqueda y asegurar una mayor probabilidad de éxito.

Agradecimientos

Subvención PID2019-104793RB-C31 y PDC2021-121517-C31 financiados por MCIN/AEI/ 10.13039/501100011033 y por la Unión Europea "NextGenerationEU/PRTR" y la Comunidad de Madrid a través de SEGVAUTO-4.0-CM (P2018/EMT-4362). Nuevo paradigma para la gestión de los servicios de transporte de emergencia: AMBULATE-CM. Este artículo forma parte del convenio entre la Comunidad de Madrid (Consejería de Educación, Universidades, Ciencia y Portavocía) y la UC3M para la concesión directa de ayudas para la financiación de proyectos de investigación sobre la enfermedad COVID-19 financiados con los recursos REACT-UE del Fondo Europeo de Desarrollo Regional A Way for Europe.

English summary

VEHICLE RE-IDENTIFICATION IN ROAD ENVIRONMENTS USING DEEP LEARNING TECHNIQUES

Abstract

The level of precision of deep neural networks in visual perception tasks allows to capture crucial information from the environment for future projects, such as autonomous vehicles and smart cities. One possibility that this type of system would allow is the control and tracking of certain suspicious vehicles. Considering the use of this technology by police, it would facilitate the tracking of certain cars under investigation. With this vision, the objective of this work is the study of the current state-of-the-art of the methods and the development of a system that solves two tasks efficiently: the visual characterization and re-identification of vehicles and the license plates segmentation and character recognition. This dual identification can adapt to the environmental conditions, target distance and cameras capabilities and resolution. To test and validate this system, a custom dataset has been created to minimize the difference between lab and real environment.

Keywords: Vehicle re-identification, deep learning, smart cities.

Referencias

[1] Anagnostopoulos, C.N.E., Anagnostopoulos, I.E., Loumos, V., Kayafas, E. (2006) A license plate-recognition algorithm for intelligent transportation system applications. IEEE Transactions on Intelligent transportation systems 7, pp. 377–392

[2] Bochkovskiy, A., Wang, C.-Y., Liao, H.-Y.M. (2020) Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:200410934

[3] He, L., Liao, X., Liu, W., Liu, X., Cheng, P., Mei, T. (2020) FastReID: A Pytorch Toolbox for General Instance Re-identification. arXiv preprint arXiv:200602631

[4] Huynh, S. v (2021) A Strong Baseline for Vehicle Re-Identification. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition pp. 4147–4154

[5] Jia, W., Zhang, H., He, X., Wu, Q. (2006) Gaussian weighted histogram intersection for license plate classification. In: Proceedings - International Conference on Pattern Recognition. pp 574–577

[6] Krause, J., Stark, M., Deng, J., Fei-Fei, L. (2013) 3d object representations for fine-grained categorization. In: Proceedings of the IEEE international conference on computer vision workshops. pp 554–561

[7] Laroca, R., Zanlorensi, L.A., Gonçalves, G.R., Todt, E., Schwartz, W.R., Menotti, D. (2019) An efficient and layout-independent automatic license plate recognition system based on the YOLO detector. arXiv preprint arXiv:190901754

 https://doi.org/10.17979/spudc.9788497498418.1031

[8] LeCun, Y., Haffner, P., Bottou, L., Bengio, Y. (1999) Object recognition with gradient-based learning. In: Shape, contour and grouping in computer vision. Springer, pp 319–345

[9] Lee, H.J., Ullah, I., Wan, W., Gao, Y., Fang, Z. (2019) Real-time vehicle make and model recognition with the residual SqueezeNet architecture. Sensors 19, pp. 982

[10] Lee, Y., Park, J. (2020) Centermask: Real-time anchor-free instance segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp 13906–13915

[11] Li, H., Wang, P., Shen, C. (2018) Toward end-to-end car license plate detection and recognition with deep neural networks. IEEE Transactions on Intelligent Transportation Systems 20, pp. 1126–1136

[12] Liu, S., Huang, D., Wang, Y. (2019) Learning spatial fusion for single-shot object detection. arXiv preprint arXiv:191109516

[13] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C. (2016) Ssd: Single shot multibox detector. In: European conference on computer vision. pp 21–37

[14] Liu, X., Liu, W., Ma, H., Fu, H. (2016) Large-scale vehicle re-identification in urban surveillance videos. In: 2016 IEEE International Conference on Multimedia and Expo (ICME). pp 1–6

[15] Lou, Y., Bai, Y., Liu, J., Wang, S., Duan, L.-Y. (2019) VERI-Wild: A Large Dataset and a New Method for Vehicle Re-Identification in the Wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp 3235–3243

[16] Naseer, S., Shah, S., Aziz, S., Khan, M.U., Iqtidar, K. (2020) Vehicle Make and Model Recognition using Deep Transfer Learning and Support Vector Machines. In: 2020 IEEE 23rd International Multitopic Conference (INMIC). pp 1–6

[17] Nukano, T., Fukumi, M., Khalid, M. (2004) Vehicle license plate character recognition by neural networks. In: Proceedings of the International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS). pp 771–775

[18] Rahman, C.A., Badawy, W., Radmanesh, A. (2003) A real time vehicle's license plate recognition system. In: Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance, 2003. pp 163–166

[19] Redmon, J., Farhadi, A. (2017) YOLO9000: better, faster, stronger. Proceedings of the IEEE conference on computer vision and pattern recognition pp. 7263–7271

[20] Redmon, J., Farhadi, A. (2018) Yolov3: An incremental improvement. arXiv preprint arXiv:180402767

[21] Shashirangana, J., Padmasiri, H., Meedeniya, D., Perera, C. (2020) Automated license plate recognition: a survey on methods and techniques. IEEE Access 9, pp. 11203–11225

[22] Silva, S.M., Jung, C.R. (2018) License plate detection and recognition in unconstrained scenarios. In: Proceedings of the European conference on computer vision (ECCV). pp 580–596

[23] Tan, M., Le, Q. (2019) Efficientnet: Rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning. pp 6105–6114

[24] Tan, M., Pang, R., Le, Q. v (2020) Efficientdet: Scalable and efficient object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp 10781–10790

[25] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I. (2017) Attention is all you need. In: Advances in neural information processing systems. pp 5998–6008

[26] Verified Market Research (2021) AI in Computer Vision Market Size And Forecast. https://www.verifiedmarketresearch.com/product/ai-in-computer-vision-market/

[27] Xu, Z., Yang, W., Meng, A., Lu, N., Huang, H., Ying, C., Huang, L. (2018) Towards end-to-end license plate detection and recognition: A large dataset and baseline. In: Proceedings of the European conference on computer vision (ECCV). pp 255–271

[28] Zhang, S., Chi, C., Yao, Y., Lei, Z., Li, S.Z. (2020) Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp 9759–9768

[29] Zheng, D., Zhao, Y., Wang, J. (2005) An efficient method of license plate location. Pattern Recognit Lett 26, pp. 2431–2438