



SISTEMA DE RECONOCIMIENTO DE PLACAS COLOMBIANAS POR MEDIO DE
REDES CONVOLUCIONALES PARA ACCESO A ÁREAS RESIDENCIALES

Colombian License Plate Recognition System Using Convolutional Neural Networks for
Residential Area Access Control

DANIEL ENRIQUE PINTO RESTREPO

Tesis

Asesor:

Tomás Olarte Hernandez

UNIVERSIDAD EAFIT

ESCUELA DE CIENCIAS APLICADAS E INGENIERÍA

MAESTRÍA EN CIENCIAS DE LOS DATOS Y ANALÍTICA

MEDELLÍN

2024

Sistema de reconocimiento de placas colombianas por medio de redes convolucionales para acceso a áreas residenciales

Daniel Enrique Pinto Restrepo

Junio 2024

Resumen

El presente estudio se centra en el desarrollo y evaluación de un sistema de reconocimiento de placas vehiculares en Colombia utilizando redes neuronales convolucionales, con el objetivo de mejorar la seguridad en el acceso a áreas residenciales. Para abordar este problema, se emplearon técnicas de transferencia de aprendizaje con modelos preentrenados en los conjuntos de datos COCO e ImageNet, y se utilizaron además conjuntos de datos específicos de Brasil e India para aumentar la disponibilidad de datos durante el reentrenamiento de los modelos. Se propusieron y compararon cinco modelos diferentes, variando las herramientas y los datos utilizados tanto para la detección de vehículos como para el reconocimiento de caracteres en las placas. Los resultados de las pruebas de campo mostraron un porcentaje de acierto del 60 %, destacando problemas específicos con la letra “Q” en las placas. A pesar de las mejoras en la precisión de las primeras fases del sistema, el principal desafío radica en la correcta identificación de caracteres específicos, lo que sugiere la necesidad de una mayor diversificación y representatividad de los datos de entrenamiento. Este trabajo proporciona una base sólida para futuros desarrollos y mejoras en sistemas de reconocimiento de placas vehiculares en contextos similares.

1. Introducción

Los sistemas de reconocimiento de placas tienen diversas aplicaciones en el mundo actual,

tales como la búsqueda de vehículos robados, la identificación de la falta de pago de seguros o impuestos, y la verificación de infracciones de tránsito. Además, estos sistemas facilitan el pago automático de peajes y el control de acceso a áreas privadas. En el contexto colombiano, cerca de 53 mil vehículos, incluyendo motos y carros, fueron hurtados en el año 2023, lo que representa un incremento del 7 %¹ en comparación con el año 2022. En el primer trimestre de 2024, la cifra de hurtos a vehículos particulares ha disminuido un 4 % en comparación con el mismo período de 2023.

Aunque no existen estadísticas precisas sobre los hurtos en unidades residenciales debido a la falta de control en el ingreso y salida de vehículos, sí hay estadísticas sobre el robo general a viviendas. Según la Policía Nacional, en 2023 se registraron 35 mil² denuncias por robo a viviendas, lo que representa un aumento de 3 mil casos en comparación con 2022. Si bien la solución presentada en este trabajo no está enfocada específicamente en el total de robos a viviendas, la automatización de los ingresos y salidas de vehículos puede contribuir significativamente a controlar este tipo de delitos.

En la actualidad, las empresas de vigilancia privada en Colombia utilizan diversas metodologías para registrar los movimientos de vehícu-

¹Cálculos propios basados en datos abiertos Colombia https://www.datos.gov.co/Seguridad-y-Defensa/HURTO-A-VEH-CULOS/csb4-y6v2/about_data

²<https://www.semana.com/nacion/articulo/los-robos-a-viviendas-estan-disparados-y-no-todas-las-empresas-de-seguridad-responden-asi-estan-actuando-los-ladrones/202426/>

los, incluyendo formatos manuales, sistemas de reconocimiento de placas, y sistemas basados en tarjetas o tags dentro de los vehículos. Por esta razón, el desarrollo de un sistema robusto de reconocimiento de placas resulta fundamental para mejorar el control de acceso a áreas privadas y, potencialmente, reducir la incidencia de robos en unidades residenciales.

En la literatura se han identificado distintos sistemas de reconocimiento de placas, que van desde metodologías clásicas de visión por computadora hasta algoritmos avanzados de aprendizaje profundo (Deep Learning), especialmente redes neuronales convolucionales (CNN). Dado el buen rendimiento de estas últimas, se propone utilizar redes convolucionales para el desarrollo del sistema. Sin embargo, una de las dificultades al elegir modelos de aprendizaje profundo es la necesidad de una gran cantidad de datos para entrenar los modelos desde cero [1]. Por ello, se opta por la transferencia de aprendizaje, una técnica de aprendizaje automático en la que un modelo ya entrenado en una tarea se utiliza como punto de partida para resolver una tarea diferente.

El desarrollo de este trabajo se estructura de la siguiente manera: primero, se presenta una revisión de los trabajos relacionados con el desarrollo de sistemas de reconocimiento de placas. En la siguiente sección, se proporciona una breve descripción de las placas colombianas. Luego, se describen los datos utilizados en el proyecto. A continuación, se detallan los modelos implementados para el desarrollo del proyecto. Posteriormente, se presentan los resultados obtenidos en las diferentes iteraciones realizadas. Después, se incluye una sección de experimentación y pruebas de campo de los algoritmos. Finalmente, se exponen las conclusiones del trabajo y se discuten las posibles direcciones para trabajos futuros.

2. Trabajos relacionados

Los sistemas basados en aprendizaje automático pueden diseñarse de dos maneras: un

sistema de principio a fin (end-to-end) que se encargue de encontrar relaciones complejas desde los insumos crudos hasta el resultado final, en este caso de una imagen o cuadro de un vídeo hasta detectar los caracteres de la placa. Sin embargo, se han encontrado limitantes en esta aproximación, en especial la cantidad de datos necesaria para que funcione [2]. La otra manera es un sistema dividido en problemas más pequeños a resolver, como lo sugiere Laroca [3], donde primero se localiza el vehículo y después solo esta porción se le envía al detector de placas, con el objetivo de reducir la complejidad del detector de placas y los falsos positivos.

En la etapa de localización del vehículo se evidencia el uso de redes neuronales convolucionales como YOLO [3] y FAST-YOLO [4]. Siguiendo esta línea, se evidencia el uso de diferentes aproximaciones de aprendizaje profundo para la ubicación de la placa como YOLO v2 [5], FAST YOLO [6], YOLO v3 [7] y MobileNet para la extracción de características y SSD como el framework de detección [8], entre otras. También, para la etapa de localización de la placa, se encuentran metodologías más clásicas como análisis de componentes conectados, análisis morfológico con estadísticas de ejes, análisis de puntos de eje y procesamiento de colores; todos los métodos mencionados son altamente dependientes de las condiciones naturales como la luz [5].

Siguiendo las etapas del sistema de reconocimiento de placas, en la segmentación de caracteres, se pueden evidenciar diferentes formas de llevar a cabo esta etapa. Se observan técnicas de aprendizaje profundo como YOLO v2, utilizado para la detección y extracción de caracteres de la imagen de la placa [9], así como otras metodologías que implementan CR-NET estableciendo 26 clases para las letras y 10 para los dígitos [3]. Otras aproximaciones se basan en un preprocesamiento de la imagen mediante el algoritmo de ecualización de histograma a nivel de grises, que permite incrementar el contraste de la región de la placa. Luego, se procede a convertir la imagen en una imagen binaria; finalmente, se segmentan los caracteres utilizando un análisis

de correlaciones canónicas [5].

Finalmente, para el reconocimiento de los caracteres segmentados, se encuentran metodologías como coincidencia de plantillas, el cual es un procedimiento de correlación cruzada entre el carácter segmentado y la plantilla de dicho carácter. También se pueden encontrar técnicas basadas en características como SVM [10]. Continuando con las metodologías utilizadas en esta etapa, también se evidencia el uso de técnicas de aprendizaje profundo como LPRNet [8], y metodologías donde se crea una arquitectura propia de una red neuronal convolucional para llevar a cabo la tarea del reconocimiento de los caracteres [9].

3. Placas Colombianas

En Colombia, el Ministerio de Transporte es la entidad encargada de diseñar y establecer las características de la placa única nacional, así como de la asignación de series, rangos y códigos, los cuales son autorizados y entregados a las autoridades de tránsito competentes. Las placas colombianas se clasifican en varias categorías: servicio oficial, particular, diplomático, consular y de misiones especiales, basándose en la razón del servicio del vehículo.

Según el Código de Tránsito, los vehículos automotores deben llevar dos placas iguales, una en la parte delantera y otra en la parte trasera del vehículo. En el caso de motocicletas, motonetas, mototriciclos y bicicletas, las placas se colocan en la parte trasera, manteniendo las mismas características y seriado que las placas de los demás vehículos.

En cuanto al contenido, las placas de vehículos de servicio público, particular y oficial tienen seis caracteres dispuestos horizontalmente en dos grupos separados: tres letras y tres dígitos. En la parte inferior de la placa se encuentra el nombre del municipio donde está registrada. Para motocicletas, maquinaria agrícola y similares, las placas tienen cinco caracteres, tres letras y dos dígitos en dos grupos, y en la parte inferior se incluye la palabra “COLOMBIA”. Las placas

de vehículos diplomáticos comienzan con las letras “CD” seguidas de cuatro dígitos, las placas de vehículos consulares inician con “CC” seguidas de cuatro dígitos, y las placas de vehículos de misiones técnicas comienzan con “YT” seguidas de cuatro dígitos, en la figura 1 se pueden ver ejemplos de estos contenidos.

Las dimensiones de las placas varían según el tipo de vehículo. Para vehículos automotores, las placas miden 330 milímetros (mm) de largo por 160 mm de ancho, mientras que para motocicletas las placas tienen 235 mm de largo por 105 mm de ancho.

Los colores de las placas también difieren según el tipo de servicio. Para el servicio particular y oficial, el fondo es amarillo (124 C) y los caracteres son negros (Black C). Las placas de servicio público tienen un fondo blanco reflectivo de al menos 50 milicandelas y caracteres negros (Black C). Las placas de motocicletas, maquinaria agrícola y similares también tienen fondo amarillo (124 C) y caracteres negros (Black C). Por último, las placas de vehículos diplomáticos, consulares y de misiones técnicas tienen un fondo azul y caracteres blancos.

Esta normativa asegura que las placas vehiculares en Colombia no solo cumplan con los estándares de identificación y regulación, sino que también sean fácilmente reconocibles y diferenciables según el tipo de servicio y vehículo.



Figura 1: Placas Colombianas

4. Datos del proyecto

4.1. Datos Colombia

Inicialmente se realizaron esfuerzos para capturar imágenes de vehículos y placas colombianas, pero se tuvieron conflictos para acceder a las cámaras de seguridad de unidades residenciales. Debido al esfuerzo y la magnitud de recoger información por medio de fotos y vídeos, se lograron capturar 43 imágenes cuyo objetivo inicial era entrenar el modelo; sin embargo, debido al reducido número de muestras, se decidieron utilizar para validación y testeo. En la figura 2 se pueden ver algunos ejemplos de los datos. Estas imágenes requirieron un proceso de etiquetado, que implicó la delimitación precisa de los contornos de los vehículos, así como la demarcación de la región que contiene la placa. Para llevar a cabo esta tarea, se utilizó un dispositivo iPhone Xs Max. Además, se empleó la herramienta de anotación de imágenes de código abierto conocida como **Labelimg** [11]. Esta aplicación resulta fundamental en la marcación y etiquetado de objetos en imágenes, lo que facilita la creación de conjuntos de datos de entrenamiento adecuados para modelos de aprendizaje profundo, tal como se requiere en este caso.



Figura 2: Datos Propios

4.2. Aumentando los datos

Para garantizar un proceso eficiente de entrenamiento y prueba del sistema, fue imperativo buscar fuentes adicionales de imágenes etiquetadas. A continuación, se detallan dos conjuntos de datos provenientes de Brasil e India, que resultaron fundamentales para el desarrollo y evaluación de la detección tanto de vehículos como de placas. Además, se describe un conjunto de datos especializado en la tarea específica de la lectura de placas.

4.2.1. Datos UFPR-ALPR

La primera fuente de aumento de datos proviene del trabajo de Laroca [3] y son los datos UFPR-ALPR el cual contiene 4500 imágenes tomadas a través de un vehículo siendo manejado por áreas urbanas. Estas imágenes fueron obtenidas de 150 vídeos con una duración de 1 segundo con una tasa de 30 cuadros por segundo (FPS). Por lo tanto, los datos se dividen en 150 vehículos cada uno con 30 imágenes y cada imagen con una sola placa visible. Las imágenes fueron obtenidas con las cámaras GoPro Hero 4 Silver, Huawei P9 Lite y Iphone 7 Plus, estas están disponibles en formato PNG con un tamaño de 1920×1080 píxeles. Estos datos vienen separados de la siguiente manera 40% para entrenamiento, 40% para testeo y 20% para validación. Cada imagen tiene disponible en los archivos de texto las siguientes anotaciones: la cámara con la cual se tomó la imagen, la posición y el tipo de vehículo (carro o motocicleta), empresa manufacturera del vehículo, modelo y año; adicional la posición de la placa. En la figura 3 podemos observar una muestra de los datos.

4.2.2. India

Además de los conjuntos de datos recopilados en Brasil, se incorporaron datos provenientes de la India, los cuales están disponibles de manera abierta en la plataforma Kaggle. Estos datos comprenden 1256 imágenes diseñadas para el desarrollo de soluciones de detección de pla-



Figura 3: UFPR-ALPR

cas adaptadas a las condiciones específicas de la India. Los datos se organizan en dos carpetas principales. La primera, denominada *State-wise_OLX*, contiene subcarpetas que representan cada estado en India, junto con las anotaciones de la ubicación de las placas en las imágenes. Estas imágenes se obtuvieron mediante el proceso de web scraping del sitio web OLX. La segunda carpeta contiene imágenes capturadas en las autopistas de Maharashtra utilizando un dispositivo móvil en diferentes condiciones de iluminación natural. Cada imagen está asociada con un archivo .xml que contiene las anotaciones correspondientes a las placas, utilizando el mismo nombre que el archivo de imagen para una fácil referencia, en la figura 4 se puede apreciar unos ejemplos de las imágenes de estas fuentes de datos junto con sus etiquetas en recuadro verde.

4.2.3. Roboflow

Los conjuntos de datos previamente mencionados se enfocan en las dos primeras etapas del proceso de reconocimiento de placas: la detección del vehículo y la localización de la placa. Para completar la última etapa, que involucra la lectura de la placa, se accedió a un conjunto de datos de código abierto disponible en la plataforma Roboflow [12]. Este conjunto comprende un total de 6076 imágenes, distribuidas en 4245



Figura 4: Datos India

para entrenamiento, 1221 para validación y 610 para pruebas, cubriendo un espectro de 36 categorías que incluyen números del 0 al 9 y letras de la A a la Z. La mayoría de las clases están bien representadas en los datos de entrenamiento, sin embargo, se observa una sub-representación de la letra Q, con solo 38 instancias presentes en las imágenes de entrenamiento. En la figura 5 se ilustra el número de instancias para cada clase dentro de los datos de entrenamiento.

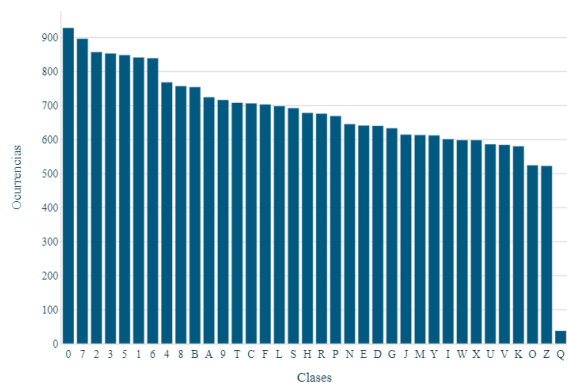


Figura 5: Ocurrencias clases

Es fundamental destacar que las imágenes asignadas para validación y pruebas no se emplearán en la evaluación del rendimiento del al-

goritmo; su inclusión en esta descripción es puramente informativa, brindando una comprensión completa sobre el origen y la distribución original de los datos en este conjunto.

Dado que el sistema de reconocimiento de placas está diseñado exclusivamente para el contexto colombiano, resulta imperativo contrastar los conjuntos de datos de Roboflow con los datos colombianos. Dada la amplitud del conjunto de Roboflow y su carencia de una descripción detallada, se llevó a cabo una inspección visual de las imágenes contenidas en el mismo. Se observó que la mayoría de estas imágenes correspondían a placas provenientes de Venezuela. No obstante, es importante destacar que el conjunto de datos también incluye una variedad de tipos de placas, así como imágenes que contienen únicamente un carácter. En la figura 6 se pueden observar algunos ejemplos de las imágenes encontradas en este conjunto de datos.

En cuanto a las placas de Venezuela, estas se caracterizan por una serialización compuesta por cuatro letras seguidas de tres números, formando un total de siete caracteres (por ejemplo, AB123CD). Dichas placas presentan una forma rectangular estándar de 300 mm x 150 mm, con bordes negros y un fondo blanco sobre el cual se superpone una impresión de la bandera nacional. El serial alfanumérico se sitúa en el centro de la placa y su secuencia de dígitos refleja el tipo de vehículo al que está asignada.

Si se comparan las placas de Venezuela con el estándar del Ministerio de Transporte colombiano, se encuentran diferencias notables, aunque también similitudes. A pesar de no ser idénticas, ambas comparten características como el tamaño y la ausencia de caracteres especiales, empleando exclusivamente letras mayúsculas.

4.3. Análisis descriptivo

Esta sección se centra en comparar las distribuciones de los diferentes conjuntos de datos utilizados para apoyar el desarrollo del proyecto. En la figura 7 se observan los dos mapas de calor para los datos de los vehículos de Bra-

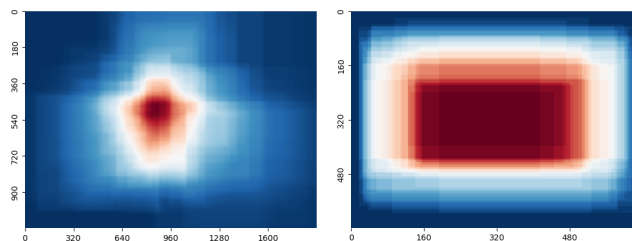


(a) Placa Venezuela (b) Otras Imágenes

Figura 6: Datos Roboflow

sil y Colombia. Es evidente que hay una mayor distribución en diferentes partes de los vehículos en Brasil, comparado con la centralidad en los datos de Colombia, debido a la diversidad y número de muestras que contiene el conjunto de datos de Brasil. Esto indica que el conjunto de datos de Brasil es versátil y útil como insumo para el entrenamiento de los modelos.

En cuanto a la distribución de las placas, el conjunto de datos de Brasil muestra una mayor diversidad como se puede ver en la figura 8, con placas ubicadas en distintas partes del mapa de calor. Por otro lado, los datos de India y Colombia presentan una centralidad en la ubicación de sus placas, aunque con algunas observaciones fuera de esta centralidad. En resumen, las distribuciones de los conjuntos de datos de Brasil e India pueden ser útiles para aplicaciones en distribuciones similares a las de Colombia, tanto en términos de localización de placas como de vehículos



(a) Brasil (b) Colombia

Figura 7: Mapas de calor vehículos

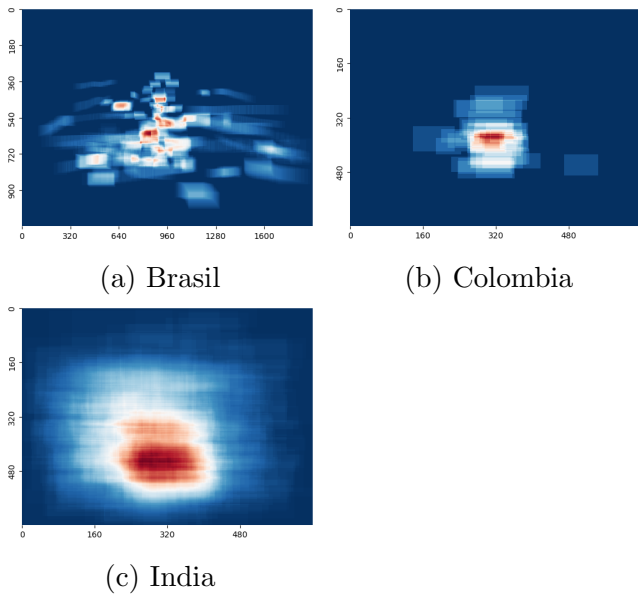


Figura 8: Mapa de calor placas

4.4. Pre-Procesamiento

En esta sección, se detallan los procesos llevados a cabo previo al modelado. El primer paso consistió en unificar las diversas fuentes de información disponibles. Para ello, se realizó un proceso de extracción y transformación de cada una de las fuentes de imágenes mencionadas anteriormente, junto con sus respectivas etiquetas, a datos tabulares. Esta transformación permitió un manejo más eficiente de los datos.

Es importante destacar que los datos de India incluían etiquetas para las placas, así como el número de las mismas. Sin embargo, no se disponía de etiquetas para los vehículos. Dado el volumen y el tipo de imágenes que requerían etiquetado, se optó por crear etiquetas ficticias para los vehículos, generando un margen del 80% de cada una de las imágenes. Estas etiquetas adicionales de los vehículos de la India permiten abordar el problema de la detección parcial del vehículo, ya que dichas imágenes presentan un acercamiento significativo a la parte trasera del vehículo, como se puede ver en algunos ejemplos de la figura 4. Este enfoque permite que el algoritmo identifique correctamente la etiqueta “auto”, incluso cuando solo una parte del vehículo es visible, y proceda con la segunda etapa de detección de la placa. Asimismo, como parte del

proceso de preparación de los conjuntos de datos, se llevó a cabo un reescalado de las etiquetas de las placas. Esta acción se realizó después de recortar las imágenes según las etiquetas de los contornos del vehículo. Posteriormente, se demostrará el impacto de no haber realizado este reescalado durante el entrenamiento de los dos modelos.

Una vez completada la unificación de los datos, se procedió a la división de estos en conjuntos de entrenamiento, validación y prueba tanto para el modelo de detección del vehículo como para el de detección de placas. En el caso de los datos de Brasil, estos ya estaban separados por defecto en estas tres categorías, por lo que se mantuvieron así. En cuanto a los datos de India, se dividió el conjunto en un 50% para entrenamiento, reservando el restante 50% para pruebas, con una subseparación del 80% para pruebas y el 20% para validación. Respecto a los datos de Colombia, las 43 imágenes se dividieron en un 80% para pruebas y un 20% para validación. Esta distribución se realizó con el propósito de maximizar el número de imágenes en el conjunto de prueba, asegurando así una evaluación más robusta del modelo.

Por último, se encontraron etiquetas con caracteres especiales, especialmente en placas de Brasil, donde se utiliza un guión para separar letras y números. En este contexto, se decidió eliminar los caracteres especiales de las etiquetas, con la finalidad de simplificar el proceso y obtener resultados consistentes en el reconocimiento de placas, limitando la salida a letras y números únicamente.

Para mejorar la robustez y la generalización del modelo, se aplicaron dos técnicas de aumento de datos durante el preprocesamiento. En primer lugar, se implementó un desenfoque con un radio máximo de 2.5 píxeles, con el objetivo de simular condiciones de captura menos nítidas y más realistas. Además, se introdujo ruido en las imágenes, alterando hasta un 1.84% de los píxeles, lo que contribuye a fortalecer la capacidad del modelo para manejar la variabilidad inherente en los datos de entrada. En la figura 9 se muestra una imagen en su estado original jun-

to con las dos fuentes de aumento de datos; en 9b se observa el desenfoque aplicado a la imagen, mientras que en 9c se aprecia la distorsión en los píxeles. Estas técnicas de aumento de datos se emplearon para mejorar la capacidad de generalización del modelo y su resistencia a condiciones adversas durante la inferencia.

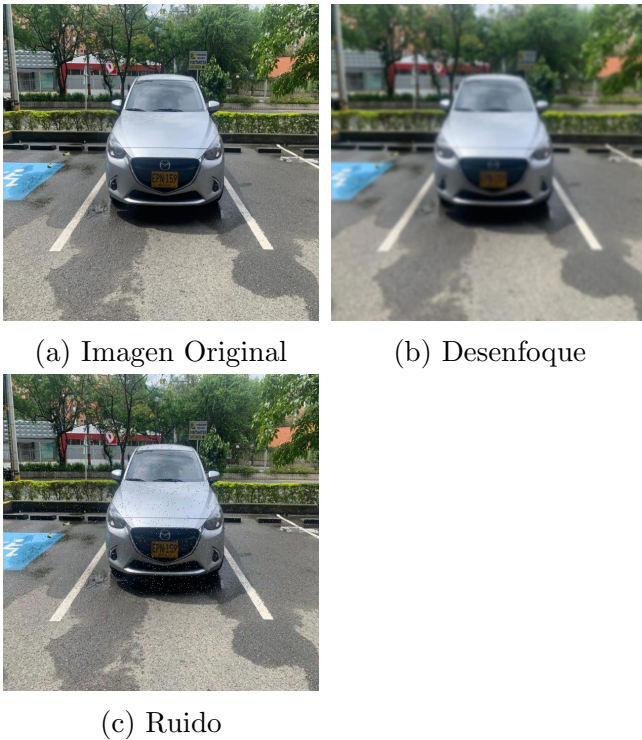


Figura 9: Comparación aumento de datos

5. Metodología propuesta para el sistema de reconocimiento de placas

La siguiente sección describe la aproximación propuesta para solucionar el sistema de reconocimiento de placas y se encuentra dividida en tres sub secciones, una por cada etapa del sistema de reconocimiento de placas.

Dentro de cada etapa se utilizaron redes neuronales convolucionales (CNN's), de esta forma se pueden entrenar los parámetros con el objetivo de mejorar el desempeño en cada una de las etapas. Los modelos utilizados son: SSD Mobilenet, YOLOv8 y Pytesseract.

5.1. Detección del vehículo

En esta primera etapa se re-entrenó SSD Mobilenet[13] utilizando los pesos pre-entrenados de COCO[14] para la tarea específica de detectar un vehículo. De igual forma se empleó SSD Mobilenet entrenado en COCO y YOLOv8[15] entrenado en ImageNet[16] para hacer una comparación entre un modelo base ya entrenado. Es crucial destacar que los modelos pre-entrenados en conjuntos de datos como COCO e ImageNet están diseñados con una gran especificidad en las categorías que pueden detectar. Por ejemplo, estos conjuntos dividen los vehículos en categorías como carros, camiones y autobuses, mientras que en este proyecto se optó por agrupar todos estos vehículos bajo la etiqueta simplificada de carro. Es por esto que al utilizar los modelos ya entrenados con esto, se mantuvieron solo estas clases, desechando las predicciones de categorías diferentes a estas. Además, se mantuvieron las categorías de personas, motocicletas y bicicletas debido a la frecuente ocurrencia de predicciones donde la persona era la parte más prominente, como cuando una persona está montada en una motocicleta. En estos casos, la predicción del modelo se basó en el cuadro delimitador (bbox) con el área más grande en la imagen.

Cómo se puede observar en la figura 10, la primera parte del sistema de reconocimiento de placas recibe la imagen normal, en esta parte se evidencia en verde la detección del vehículo y ese recorte del cuadro delimitador es el insumo que se le pasa a la siguiente etapa.

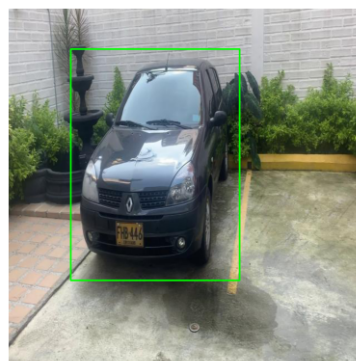


Figura 10: Detección Vehículo

5.2. Detección de la placa

Para la segunda etapa, se empleó SSD Mobilenet re-entrenado con los pesos de COCO para la localización de placas. No se encontraron otras metodologías comparables específicamente para esta tarea en este contexto. En esta sección, luego de recibir el insumo de la detección del vehículo, se detecta la placa tal y cómo se muestra en la figura 11, el recuadro verde delimita nuevamente el insumo que recibirá la parte final del sistema.



Figura 11: Detección placa

5.3. Detección de caracteres

Una vez detectada la placa, se utilizó Pytesseract como herramienta de reconocimiento óptico de caracteres (OCR) basada en el motor Tesseract OCR de Google. Pytesseract permite extraer texto de imágenes y documentos escaneados. Se ajustaron hiperparámetros para optimizar su rendimiento, determinando que tratar la imagen como una sola palabra ofrecía los mejores resultados para este proyecto.

Sin embargo, los resultados iniciales no alcanzaron el nivel de precisión deseado para todo el sistema de reconocimiento de placas. Por lo tanto, se implementó un OCR personalizado basado en YOLOv8, re-entrenado con un conjunto de datos específico de Roboflow para mejorar la precisión en la lectura de placas.

Para esta última etapa, se puede evidenciar en la figura 12 el insumo de la sección anterior, de acuerdo a esto el modelo entrenado detecta

y categoriza cada uno de los caracteres evidenciados en la imagen para de esta forma entregar la predicción final del sistema.



Figura 12: Detección caracteres

5.4. Listado Modelos

En la sección de resultados, se compararán los siguientes sistemas: el Modelo 1 integró SSD Mobilenet entrenado en COCO para la detección de vehículos, SSD Mobilenet entrenado con datos de Brasil para la detección de placas, y Pytesseract como herramienta de lectura de placas. El Modelo 2 varió únicamente en la herramienta utilizada para detectar vehículos, empleando SSD Mobilenet entrenado exclusivamente con datos de Brasil. Por otro lado, el Modelo 3 modificó el modelo de detección de vehículos utilizando YOLOv8 entrenado en ImageNet.

En busca de una metodología más robusta tras no alcanzar los resultados deseados, se procedió a reentrenar ambos modelos de detección de vehículos y placas, incorporando los datos mencionados de India. Así, el Modelo 4 utilizó SSD Mobilenet entrenado con datos combinados de Brasil e India para la detección de vehículos y placas, conservando Pytesseract para la lectura de placas.

Aunque las métricas mejoraron para las dos primeras partes del sistema, la parte final, que es crucial, no cumplió con los objetivos esperados. Por esta razón, el Modelo 5 mantuvo las dos primeras partes del Modelo 4, pero varió la herramienta utilizada para la lectura de placas,

empleando el OCR de YOLOv8 entrenado con datos de Roboflow.

6. Capacidad Computacional

Debido a la necesidad de entrenar las diversas arquitecturas seleccionadas en una GPU, y considerando el tiempo significativo requerido para el entrenamiento de cada una de las etapas (vehículo, placa y caracteres), se decidió utilizar los recursos de Google Colab, que ofrece acceso a GPUs. Este entorno proporcionó la capacidad computacional necesaria para manejar las demandas del entrenamiento intensivo.

Para llevar a cabo el proceso de entrenamiento, se realizó una suscripción a Google Colab Pro, lo que permitió aumentar los tiempos de ejecución. Entre las opciones de GPU disponibles se encontraban la T4, A100 y L4. Finalmente, se optó por utilizar la T4, ya que no consumía todos los recursos de la sesión de Google Colab y cumplía con el objetivo de entrenar los modelos de manera eficiente. Esta elección permitió reducir considerablemente los tiempos de entrenamiento y actualizar los pesos de cada una de las arquitecturas de manera efectiva.

Una vez completado el entrenamiento y obtenidos los pesos actualizados para cada una de las etapas del proceso de reconocimiento de placas, estos pesos fueron transferidos y utilizados en un procesador AMD Ryzen 7 PRO 4750U con gráficos Radeon a 1.70 GHz para probar los diferentes modelos. Aunque el enfoque principal del trabajo no es el despliegue del sistema, garantizar que este pueda funcionar en una CPU es de suma importancia para futuras instancias del proyecto. Este enfoque híbrido, combinando el entrenamiento en GPU y la prueba en CPU, asegura la viabilidad del sistema en diferentes entornos computacionales y proporciona una evaluación integral de los modelos desarrollados.

7. Resultados

Antes de adentrarnos en el análisis de los resultados obtenidos por el sistema de reconocimiento de placas, es crucial comprender las métricas de calidad utilizadas para evaluar el rendimiento global del algoritmo. La métrica principal se centra en la precisión del sistema en el reconocimiento de placas, representando la tasa de aciertos que el sistema logra al identificar correctamente las placas. En cuanto a las dos etapas previas del sistema, aunque el resultado final es fundamental, su función principal radica en identificar posibles deficiencias en el sistema antes de la etapa crítica del reconocimiento de placas.

Con el propósito de ofrecer una evaluación más detallada, se emplea la métrica de intersección sobre la unión (IoU) entre la predicción del sistema y el valor real de la etiqueta. Se considera un acierto cuando el IoU es igual o superior a 0.8; de lo contrario, se registra como un fallo. Al final, todas estas métricas se presentan como porcentajes de acierto en relación con el total de datos evaluados, proporcionando una visión integral del desempeño del sistema.

Finalmente, es importante mencionar el número de observaciones utilizadas en cada iteración para calcular las tasas de acierto. Para Brasil, se contaba con un total de 1800 imágenes de prueba. En el caso de India, se utilizaron 1004 imágenes, mientras que Colombia, con la muestra más pequeña, contó con un total de 36 imágenes.

7.1. Iteración 1

En esta primera iteración, los resultados de los tres modelos evaluados muestran un desempeño similar en general para la métrica final, como lo muestra el cuadro 1. Se observa un mejor rendimiento en la lectura de placas de Colombia tanto para el Modelo 1 como para el Modelo 3. Por otro lado, los resultados son notablemente bajos para las placas de India, siendo el desempeño del Modelo 2 el más deficiente en este contexto. En contraste, el Modelo 2 demuestra

el mejor rendimiento en los datos de Brasil, lo que sugiere que el entrenamiento específico con datos brasileños proporciona una ligera ventaja en la lectura de placas en ese país.

Al analizar las métricas del pipeline con mayor detalle, se encontró que el Modelo 3 presenta un bajo acierto en la detección de vehículos en Colombia, aunque su desempeño en la lectura de placas es comparable al de los otros modelos. Específicamente, el rango de IoU para la detección de vehículos en Colombia con este modelo varía de 0.93 a 0.50. Al revisar una predicción particular ilustrada en la figura 13 y compararla con su verdad absoluta (representada en rojo) y la predicción del Modelo 3 (representada en verde), se observa que, aunque la métrica no la considera un acierto, la predicción del modelo proporciona suficiente información para que el pipeline pueda generar una lectura de placas precisa.

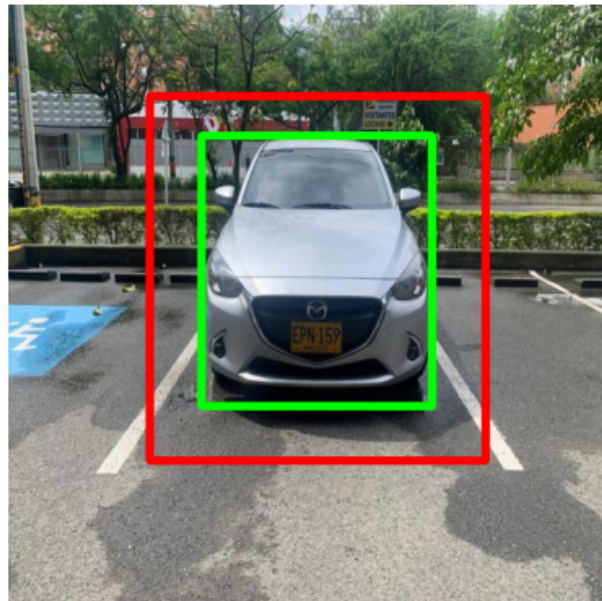


Figura 13: Error Modelo 3

Modelo	País	Acierto Carro	Acierto Placa	Acierto Lectura
1	Brasil	53 %	55 %	9 %
	India	12 %	51 %	10 %
	Colombia	14 %	47 %	28 %
2	Brasil	86 %	90 %	16 %
	India	24 %	30 %	5 %
	Colombia	67 %	47 %	22 %
3	Brasil	74 %	81 %	12 %
	India	11 %	52 %	10 %
	Colombia	8 %	50 %	28 %

Cuadro 1. Resultados Iteración 1

7.2. Iteración 2: Re-entrenando el Modelo

Los resultados de esta segunda iteración muestran una mejora considerable en la localización de vehículos y placas en comparación con el modelo anterior, esto evidenciado en el cuadro 2. Sin embargo, los resultados en cuanto a la lectura de las placas siguen siendo insatisfactorios, sin mejoras significativas en ninguno de los tres países evaluados.

Modelo	País	Acierto Carro	Acierto Placa	Acierto Lectura
4	Brasil	100 %	100 %	17 %
	India	99 %	95 %	11 %
	Colombia	71 %	76 %	18 %

Cuadro 2. Resultados Iteración 2

7.3. Iteración 3: Propio OCR

En comparación con los resultados de las iteraciones anteriores, en el cuadro 3 se observa una mejora significativa en la precisión de la lectura de placas, con un aumento del 20 % en los aciertos para Brasil e India. En cuanto a Colombia, que es el foco principal de este trabajo, la tasa de acierto aumentó al 88 %. Es importante destacar que la muestra para Colombia es mucho más reducida.

En el análisis de error para Colombia, se encontró que los errores en la lectura de placas se concentraban principalmente en vehículos que incluían la letra “Q” en sus placas. El modelo tenía dificultades para diferenciar entre la “Q”, la “O” y el número “0”, lo que resultaba en predicciones incorrectas. Esta deficiencia se atribuye a la subrepresentación de la letra “Q” en el

conjunto de entrenamiento utilizado para la lectura de placas, lo que afectó negativamente la capacidad del modelo para reconocerla correctamente. En la figura 14 se puede evidenciar como el modelo confundió en el caso de la primera placa la letra “Q” por el número “0”, mientras que para la otra placa tuvo una doble predicción, ya que arrojó tanto la letra “O” y el número “0”.

Modelo	País	Acierto Carro	Acierto Placa	Acierto Lectura
5	Brasil	100 %	100 %	34 %
	India	99 %	95 %	30 %
	Colombia	71 %	76 %	88 %

Cuadro 3. Resultados Iteración 3



(a) LR0311



(b) LO0M060

Figura 14: Errores Modelo 5

8. Experimentos

8.1. Experimento 1: Leyendo placas perfectamente cortadas

Una experimento importante consistió en verificar cómo sería la lectura de las placas si todo el pipeline previo funcionara al 100 % de precisión. Para realizar esta evaluación, se utilizaron las etiquetas verdaderas de las placas y se aplicó el lector de placas entrenado con YOLOv8. Los resultados obtenidos de esta prueba están plasmados en el cuadro 4

Estos resultados muestran un desempeño muy similar al del pipeline entrenado en la iteración 3, lo que indica que la etapa final de lectura de placas sigue siendo un desafío considerable.

País	Acierto Lectura
Brasil	34 %
India	31 %
Colombia	88 %

Cuadro 4. Resultados Experimento Placas Cortada

8.2. Experimento 2: ChatGPT

En un segundo experimento, se probó ChatGPT en la misma tarea que se había planteado para el pipeline anterior. Para ello, se utilizó la API de OpenAI con el modelo **gpt-4-turbo**, que tiene capacidad para analizar imágenes. La prueba en estas imágenes tuvo un costo de 23.85 USD, los resultados arrojados del experimento se pueden ver en el cuadro 5

País	Acierto Lectura
Brasil	55 %
India	69 %
Colombia	82 %

Cuadro 5. Resultados Experimento gpt-4-turbo

Comparando esta iteración con las anteriores, este modelo demostró ser el más potente en la lectura de placas de todos los países evaluados. Sin embargo, para Colombia, el desempeño siguió siendo similar al de la iteración anterior. Esto evidencia la fortaleza de estos modelos en la realización de tareas de visión por computadora sin necesidad de un entrenamiento exhaustivo.

En la figura 15 podemos observar la evolución que se tuvo en el desempeño final del sistema de reconocimiento de placas en todo el proceso realizado, vemos como en unas etapas iniciales no se lograban resultados relevante, finalmente con chatGPT y con el OCR propio se logran resultados prometedores específicamente para Colombia.

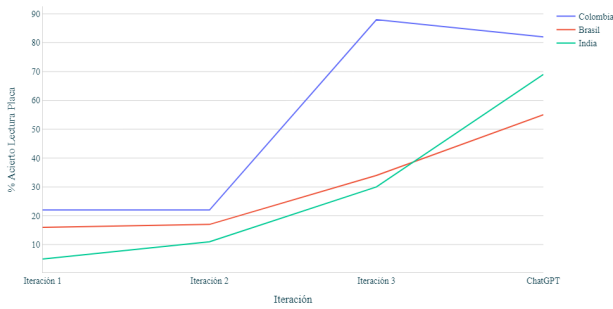


Figura 15: Evolución acierto placas

9. Prueba en campo

En esta sección, se llevó a cabo un experimento práctico para evaluar el desempeño del sistema de reconocimiento de placas en condiciones reales. Se grabaron vídeos de vehículos ingresando a unidades residenciales y centros comerciales, simulando la posición típica de una cámara de seguridad en un entorno real, como se muestra en la figura 16. La captura de los vídeos se realizó a diferentes horas del día: a las 4 de la tarde en la unidad residencial y a las 2 de la tarde en el centro comercial.

La metodología del experimento consistió en iniciar la grabación de un vídeo al observar un vehículo en movimiento, con la duración del vídeo siendo aleatoria y dependiendo de la velocidad a la que cruzaba el vehículo. En total, se utilizaron diez vídeos, cada uno capturando un vehículo particular. Las grabaciones se realizaron con un iPhone 14 Pro, y la duración de los vídeos varió entre ellos, reflejando diferentes escenarios de ingreso. Cada vídeo fue guardado en una carpeta independiente, anotando la placa del vehículo correspondiente.

Debido a restricciones de tiempo y a la concurrencia limitada de vehículos en los lugares seleccionados para la captura de información, no fue posible obtener más vídeos. Sin embargo, este experimento proporciona una evaluación inicial del sistema en condiciones reales y ofrece información valiosa para futuros ajustes y mejoras.

A partir de estos vídeos, se extrajeron un to-

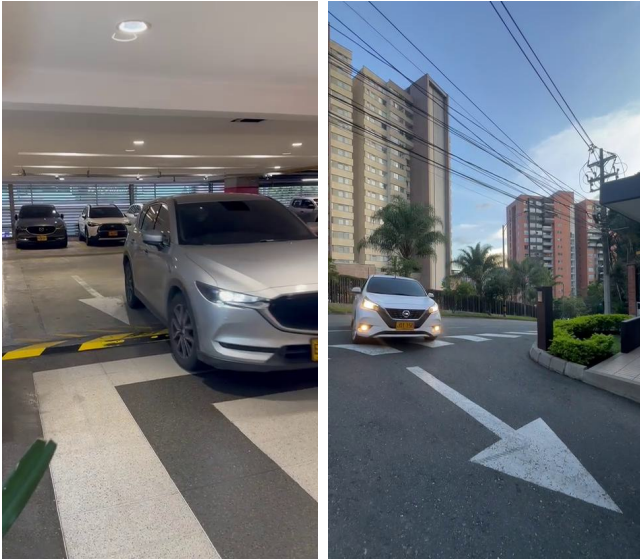
tal de 2,231 imágenes individuales. Cada imagen fue procesada mediante el sistema de reconocimiento de placas propuesto y el primer modelo para realizar una comparación entre ellos. Dado el volumen de imágenes por vídeo, se adoptó un enfoque estadístico para determinar la predicción final de cada vídeo. Específicamente, se tomó la moda de las predicciones obtenidas de todas las imágenes de un vídeo específico, asegurando así una mayor robustez en los resultados.

El análisis reveló un porcentaje de acierto del 60 % para el sistema propuesto, a comparación con el primer modelo que para la moda de los videos no logró ningún acierto, aunque en algunos imágenes puntuales si logró la predicción correcta de la placa. Dentro de las placas en las cuales no hubo acierto, se identifican temas de inclusión de la letra ‘Q’, así como letras ‘O’ que fueron confundidas con el número ‘‘0’’, también a tener en cuenta es el tema de la letra ‘X’ que tampoco fue acertada. Este hallazgo sugiere que el sistema enfrenta dificultades particulares con la distinción de estos caracteres, lo cual puede estar relacionado con la subrepresentación de la letra ‘Q’ en el conjunto de datos de entrenamiento, lo cual refleja una similitud visual con el ‘‘0’’ y ‘‘O’’ y por lo tanto una confusión del sistema.

En la figura 17, se puede observar el porcentaje de acierto del modelo 5 para los datos de testeo en Colombia como para los datos de la prueba en campo, la gráfica revela un desacierto total en la letra ‘Q’ y ‘X’, al igual que un acierto bajo para la letra ‘O’, reflejando lo anteriormente mencionado sobre la confusión entre este carácter y el ‘‘0’’.

10. Conclusiones

En este estudio, se desarrolló y evaluó un sistema de reconocimiento de placas vehiculares enfocado en el contexto colombiano, empleando redes neuronales convolucionales y técnicas de transferencia de aprendizaje. A través de múltiples iteraciones y ajustes en los modelos, se implementó un sistema que combina la detección



(a) Centro Comercial (b) Unidad Residencial

Figura 16: Escenarios Prueba en campo

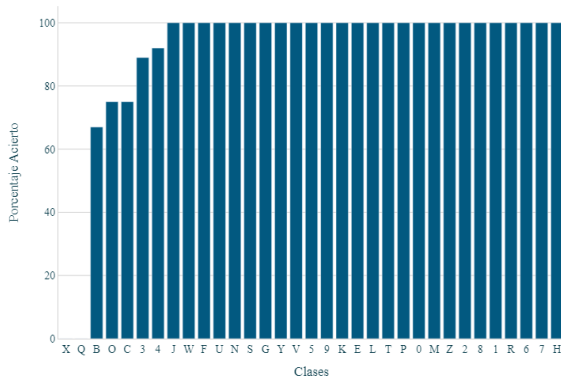


Figura 17: Acierto caracteres

de vehículos y placas con el reconocimiento óptico de caracteres (OCR). Los resultados de las pruebas en campo mostraron un porcentaje de acierto del 60% para el sistema de reconocimiento de placas, mientras que el Modelo 2 no logró ningún acierto en esta prueba. Esto evidencia un avance significativo en la detección y lectura de placas, aunque persisten limitaciones en la identificación de ciertos caracteres.

Se aprovecharon modelos preentrenados para obtener ventajas al momento de entrenar los modelos en la tarea específica de este estudio. Esto permitió un desarrollo más eficiente y preciso, aprovechando la especificidad y precisión

de estos modelos en tareas de detección y clasificación.

Un aspecto crucial que se destacó es el desafío de obtener conjuntos de datos etiquetados para tareas específicas. La recopilación y etiquetado de datos es una tarea sumamente importante y, a la vez, costosa ya que requiere un esfuerzo humano, que requiere un esfuerzo significativo para garantizar la robustez del proyecto. Recurrir a trabajos previos que ya han llevado a cabo esta labor facilita el proceso de entrenamiento y acelera el desarrollo. No obstante, para abordar problemáticas puntuales, la tarea de etiquetado sigue siendo fundamental para asegurar que el sistema pueda adaptarse y funcionar eficazmente en contextos específicos.

En general, los resultados indican que, aunque se han alcanzado mejoras considerables, aún existen desafíos importantes, particularmente en la diversidad y representatividad de los datos de entrenamiento. La precisión en la lectura de caracteres sigue siendo un área crítica que necesita atención para lograr un sistema de reconocimiento de placas completamente robusto y fiable.

11. Trabajo futuro

Para futuras investigaciones y desarrollos, se proponen las siguientes direcciones:

1. Ampliación y Diversificación del Conjunto de Datos: Es fundamental aumentar la cantidad y diversidad de datos de entrenamiento, especialmente aquellos que incluyen caracteres menos representados como la letra "Q". La inclusión de datos de diferentes regiones y condiciones ambientales también puede mejorar la robustez del sistema.
2. Evaluación en Diferentes Contextos: Probar y evaluar el sistema en una variedad de contextos y escenarios reales, como diferentes tipos de vehículos y condiciones de iluminación, para garantizar su eficacia y adaptabilidad.

Este estudio sienta las bases para el desarrollo continuo de sistemas de reconocimiento de placas vehiculares, proporcionando una hoja de ruta clara para futuras mejoras y adaptaciones que pueden llevar a un sistema más preciso y fiable en el reconocimiento de placas en diversos contextos.

12. Anexos

12.1. Comparación Pre-Procesamiento

Como se mencionó anteriormente, en la etapa de pre-procesamiento de los datos, para la etapa de localización de las placas no se usaron las imágenes originales, sino que se utilizó la imagen recortada por el cuadro de delimitación del vehículo. Esto ayudó enormemente al desempeño del modelo, haciendo mucho más estable la localización de las placas, dicho esto, se utilizó un mismo vídeo probando las dos iteraciones, la primera sin recortar las imágenes y la segunda recortando las imágenes, luego, separamos cuadro por cuadro el vídeo y realizamos la comparación. En la figura 18 se puede observar una localización mucho más estable al realizar el pre-procesamiento mencionado.

12.2. Repositorio Github

El trabajo realizado y todo el código generado para llevar a cabo este proyecto se encuentran en el repositorio de GitHub <https://github.com/dpr3619/colombian-anpr/tree/dpr3619-patch-1>

Referencias

- [1] Asmaul Hosna, Ethel Merry, Jigmei Gyalmo, Zulfikar Alom, Zeyar Aung, and Mohammad Abdul Azim. Transfer learning: a friendly introduction. *Journal of Big Data*, 9, 12 2022.

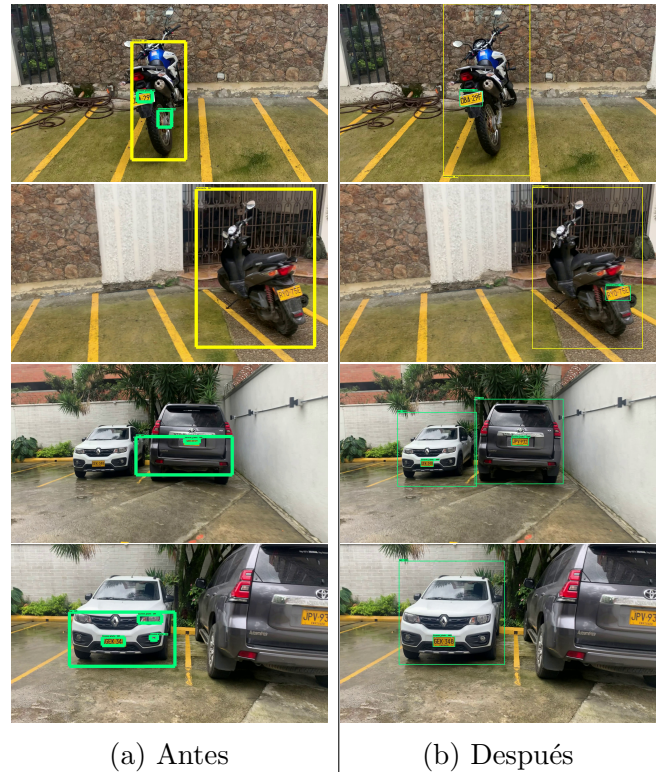


Figura 18: Comparación pre-procesamiento

- [2] Tobias Glasmachers. Limits of end-to-end learning. *CoRR*, abs/1704.08305, 2017.
- [3] Rayson Laroca, Evair Severo, Luiz A. Zanolensi, Luiz S. Oliveira, Gabriel Resende Gonçalves, William Robson Schwartz, and David Menotti. A robust real-time automatic license plate recognition based on the yolo detector. 2 2018.
- [4] Sergio Montazzolli Silva and Claudio Rosito Jung. Real-time brazilian license plate detection and recognition using deep convolutional neural networks. pages 55–62. Institute of Electrical and Electronics Engineers Inc., 11 2017.
- [5] Kh Tohidul Islam, Ram Gopal Raj, Syed Mohammed Shamsul Islam, Sudanthi Wijewickrema, Md Sazzad Hossain, Tayla Razmovski, and Stephen O’leary. A vision-based machine learning method for barrier access control using vehicle license plate authentication. *Sensors (Switzerland)*, 20:1–18, 6 2020.

- [6] Rayson Laroça, Luiz A. Zanlorensi, Gabriel R. Gonçalves, Eduardo Todt, William Robson Schwartz, and David Menotti. An efficient and layout-independent automatic license plate recognition system based on the yolo detector. *IET Intelligent Transport Systems*, 15:483–503, 4 2021.
- [7] Narayana Darapaneni, Kiran Mogeraya, Shubham Mandal, Ashwin Narayanan, Prasanna Siva, Anwesh Reddy Paduri, Faisal Khan, and Praful Mohan Agadi. Computer vision based license plate detection for automated vehicle parking management system. pages 0800–0805. Institute of Electrical and Electronics Engineers Inc., 10 2020.
- [8] Yusef Alborzi, Talayeh Sarraf Mehraban, Javad Khoramdel, and Ali Najafi Ardekany. Robust real time lightweight automatic license plate recognition system for iranian license plates. pages 352–356. Institute of Electrical and Electronics Engineers Inc., 11 2019.
- [9] Bhavin Dhedhi, Prathamesh Datar, Anuj Chiplunkar, Kashish Jain, Amrith Ranganathan, and Jayshree Kundargi. Automatic license plate recognition using deep learning. 2019.
- [10] Lubna, Naveed Mufti, and Syed Afaq Ali Shah. Automatic number plate recognition: a detailed survey of relevant algorithms, 5 2021.
- [11] Darrenl Tzutalin. Labelimg.gitcode. <https://github.com/tzutalin/labelImg>, 2015.
- [12] W251OCR. Ocr dataset. <https://universe.roboflow.com/w251ocr/ocr-oy9a7>, mar 2023. visited on 2024-04-29.
- [13] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. Ssd: Single shot multibox detector. 12 2015.
- [14] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. Microsoft coco: Common objects in context. 5 2014.
- [15] Ultralytics. Yolov8: Real-time object detection and image segmentation, 2024. Available at <https://github.com/ultralytics/yolov8>.
- [16] Jia Deng, Wei Dong, Richard Socher, Li Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009.