ISSN 1870-4069

Detección y clasificación de señales de tráfico mexicanas mediante aprendizaje profundo

Rubén Castruita Rodríguez, Carlos Mendoza Carlos, Osslan Osiris VergaraVillegas, Vianey Guadalupe Cruz Sánchez

Universidad Autónoma de Ciudad Juárez,
Instituto de Ingeniería y Tecnología,
México

{al176515, al176524}@alumnos.uacj.mx,
{overgara, vianey.cruz}@uacj.mx

Resumen. La detección y clasificación automática de señales de tráfico es una tarea para apoyar la seguridad de un conductor e incluso para asistir en la navegación de un automóvil autónomo. El objetivo del artículo es presentar una metodología para la detección y clasificación de señales de tráfico mexicanas mediante aprendizaje profundo. La metodología se divide en cinco etapas: 1) se realiza la recolección de 1284 imágenes de señales de tráfico en un ambiente no controlado, 2) se propone un proceso manual para la detección de señales de tráfico en las imágenes, 3) se entrena una red neuronal convolucional con el conjunto CIFAR-10 para obtener conocimiento amplio de características de diferentes objetos, 4) se utiliza una red neuronal convolucional basada en regiones para detectar las señales, y 5) se realiza un proceso de transferencia de conocimiento y de aumentado de datos para la clasificación con ResNet-50 modificada. De acuerdo con los resultados obtenidos de los experimentos se concluye que la metodología propuesta permite reconocer señales de tráfico mexicanas con una exactitud del 95.33%, lo cual es competitivo con los resultados presentados en la literatura. Además, para demostrar la robustez de la propuesta, se presenta una prueba para clasificar imágenes que no contienen señales de tráfico cuyo resultado de exactitud fue de 99.5%.

Palabras clave: Señales de tráfico, red neuronal convolucional, red neuronal convolucional basada en regiones, detección de regiones.

Detection and Classification of Mexican Traffic Signs Using Deep Learning

Abstract. Automatic detection and classification of traffic signs is a task to support the safety of a driver and even to assist in the navigation of a self-driving car. The goal of this article is to present a methodology for the detection and classification of Mexican traffic signs using deep learning. The methodology comprises five stages: 1) the collection of 1284 images of Mexican traffic signs is carried out in an uncontrolled environment, 2) a manual process for the detection of traffic signs in the images is proposed, 3) a convolutional neural

network with the CIFAR-10 set is trained to obtain a broad knowledge of the characteristics of different objects, 4) a region-based convolutional neural network is used to detect the signals, and 5) a process of knowledge transfer and data augmentation is carried out for signal classification with a modified ResNet-50. According to the results obtained from the experiments, it is concluded that the proposed methodology allows recognizing Mexican traffic signs with an accuracy of 95.33%, which is competitive with the results presented in the literature. Also, to demonstrate the robustness of the proposal, a test to classify images that do not contain traffic signs is presented, obtaining an accuracy of 99.5%.

Keywords: Traffic signs, convolutional neural network, region-based convolutional neural network, region detection.

1. Introducción

Las señales de tráfico son los signos visuales utilizados para ofrecerle información a los conductores y peatones que transitan por un camino [1, 2]. Las señales de tráfico se clasifican de acuerdo con sus formas y colores, de manera que sean llamativas para los automovilistas y se les preste atención [3]. De hecho, los países europeos han realizado trabajos para estandarizar las señales de tráfico, lo que dio lugar a la convención de Viena [4]. Sin embargo, a pesar de los esfuerzos, todavía existen variaciones entre las señales de tráfico utilizadas en las diferentes regiones del mundo.

Particularmente en México, de acuerdo con la Secretaría de Comunicaciones y Transportes (SCT), las señales de tráfico se dividen en dos; las verticales que son construidas con placas e instaladas a través de postes, y las horizontales que son las rayas, palabras, símbolos y objetos, aplicados o adheridos sobre el pavimento [5]. A su vez, las verticales se clasifican en tres tipos que son las restrictivas, preventivas e informativas. Las restrictivas indican prohibiciones reglamentarias que regulan el tránsito en las carreteras y son de color rojo y blanco. Las preventivas advierten a los conductores de la existencia de algún problema o peligro en las carreteras y son de color amarillo y negro. Las informativas les otorgan a los conductores datos de ubicación, nombres o kilometrajes y pueden ser azules, verdes o blancas.

Desafortunadamente, cuando las señales de tráfico se ignoran o no se observan debido a condiciones adversas, como puede ser el clima, mala iluminación u oclusiones parciales, se puede causar un accidente. Por lo que, la posibilidad de construir un sistema automático para la detección y reconocimiento de señales de tráfico podría apoyar a los conductores a tener una conducción más segura cuando están cansados, cuando por descuido las ignoran o cuando no las alcanzan a ver.

Debido a que las señales de tráfico se clasifican por sus formas y colores, se pueden aplicar técnicas de visión artificial para detectarlas y reconocerlas en imágenes o vídeos [6]. Actualmente, dichas técnicas se aplican en automóviles autónomos, vigilancia de tráfico, asistencia a conductores o mantenimiento de las carreteras. Actualmente, hay vehículos que cuentan con sistemas de reconocimiento de señales de tráfico, como BMW Series 5 y 7, Ford Focus y Edge, etcétera [7].

En la literatura, se han presentado diversos trabajos sobre detección y reconocimiento de señales de tráfico [1, 3, 8, 9, 10, 11].

Los trabajos se enfocan principalmente en realizar un análisis de los métodos de detección por la forma o color de las señales [12, 13], y en resolver los retos que conlleva una buena detección (variación de iluminación, distorsiones geométricas, figuras similares a la señal, entre otras) [14, 15, 16], junto con la segmentación [17, 18, 19], reconocimiento [20, 21] y rastreo de señales en secuencias de imágenes [22, 23]. Sin embargo, no se encontraron trabajos enfocados a la detección y clasificación del conjunto de señales de tráfico mexicanas, ni tampoco un trabajo que utilice la fusión de una red neuronal convolucional (CNN, por sus siglas en inglés) y una red neuronal convolucional basada en regiones (R-CNN, por sus siglas en inglés) para resolverlo.

Debido a que todavía hay retos por resolver para diseñar un sistema de detección y reconocimiento completamente exitoso, el presente trabajo tiene como objetivo desarrollar una metodología para la detección y clasificación de señales de tráfico mexicanas, adquiridas en condiciones no controladas, mediante técnicas de aprendizaje profundo que incluyen a la CNN y la R-CNN. Las principales contribuciones del trabajo se resumen a continuación:

- Se presenta un set de datos con señales de tráfico mexicanas, de la ciudad de Monterrey, Nuevo León y Ciudad Juárez, Chihuahua, compuesto de 1028 imágenes RGB que contienen un total de 1126 señales de tráfico.
- Se propone y se prueba una metodología para la detección y clasificación de señales de tráfico mexicanas en ambientes no controlados basada en aprendizaje profundo.

El resto del artículo se encuentra organizado de la siguiente manera: en la sección 2, se describen los materiales y los métodos para el desarrollo de la metodología. La sección 3, describe los experimentos y los resultados obtenidos. Finalmente, la sección 4, discute las conclusiones obtenidas del trabajo.

2. Materiales y métodos

La metodología propuesta para resolver el problema de detección y reconocimiento de señales de tráfico se compone de cinco módulos los cuales se muestran en la Fig. 1, cada módulo se describe a continuación.

Fig. 1. Módulos para la detección y reconocimiento de señales de tráfico mexicanas.

Fig. 2. Señales de tráfico mexicanas. (a) Alto. (b) Ceda el paso. (c) Límite de velocidad. (d) No estacionarse. (e) Vuelta prohibida. (f) Glorieta. (g) Niños jugando. (h) Cruce peatonal. (i) Tope. (j) Estacionamiento para discapacitados. (k) Parada de autobús.

2.1. Recolección de imágenes

Para el módulo de recolección de imágenes se obtuvieron muestras de los tres tipos de señales verticales en dos etapas. En la primera, se capturaron señales de tráfico con la cámara de un celular iPhone 8 y un Samsung S9. Las imágenes se obtuvieron en Ciudad Juárez, Chihuahua, e incluyen avenidas y calles transitadas, parques y áreas recreativas públicas, estacionamientos de escuelas y plazas comerciales, así como fraccionamientos y colonias. Al final, se recolectaron 233 imágenes a color con resolución de 900 x 1200 píxeles (iPhone 8) y 675 x 900 píxeles (Samsung S9). Debido a la poca cantidad de imágenes conseguidas, se procedió a la segunda etapa.

En la segunda etapa, se utilizó la herramienta *Street View* de *Google Maps*. Para la recolección se decidió desplazarse en las calles del mapa de Ciudad Juárez, Chihuahua y Monterrey, Nuevo León. Se escogió Ciudad Juárez ya que es la ciudad donde se desarrolló la metodología, mientras que Monterrey fue seleccionada porque tiene condiciones similares a Juárez, y al ser la tercera ciudad más poblada de México, tiene una gran cantidad de señales de tráfico en sus calles y avenidas.

Tabla 1. Cantidades y tipos de señales de tráfico recolectadas.

Tipo	**Total**
Señales Restrictivas	
Alto	294
Ceda el paso	71
Límite de velocidad	129
No estacionarse	217
Vuelta prohibida	92
Señales Preventivas	
Glorieta	59
Niños jugando	91
Cruce peatonal	210
Tope	99
Señales Informativas	
Estacionamiento para discapacitados	87
Parada de autobús	77
Total	1426

Fig. 3. Proceso de detección de regiones de interés con su respectiva etiqueta.

Al final, se logró la recolección de 1051 imágenes de diferentes resoluciones en un rango de 147x148 hasta 1512x2016. Por lo que se cuenta con 1284 imágenes que contienen mínimo un señalamiento y máximo cinco, para un total de 1426 señales de tráfico. Es importante destacar que las señales se adquirieron en diferentes perspectivas, tamaños e iluminaciones, como se muestra en la Fig. 2. En la Tabla 1 se muestra el número de señales obtenidas de cada tipo.

2.2. Detección de regiones de interés

Para cada una de las imágenes recolectadas se realizó un proceso manual de detección de regiones de interés (señales de tráfico). Para comenzar, se definieron las etiquetas para representar cada una de las once clases de interés que corresponden a la primera columna de la Tabla 1. Posteriormente, a cada imagen se le insertó un

1028x12 table

	1 imageFilename	2 Alto	3 Peaton	4 NoEstacionarse	5 CedaPaso	6 ParadaCamion	7 Glorieta	8 LimVelocidad	9 EDiscapacitados	10 NinosJugando	11 Tope	12 VueltaProhibida
486	'C:\Users\Carlos Mendoza\Documents\TrafficSigns\img0486.jpg'	[307,243.07...	[]	[]	[]	[]	[]	[]	[]	[]	[]	[]
487	'C:\Users\Carlos Mendoza\Documents\TrafficSigns\img0487.jpg'	[]	[]	[174,97.5309,63.76...	[]	[]	[]	[]	[]	[]	[]	[]
488	'C:\Users\Carlos Mendoza\Documents\TrafficSigns\img0488.jpg'	[]	[326,156.0140,...	[]	[]	[]	[]	[]	[]	[]	[]	[]
489	'C:\Users\Carlos Mendoza\Documents\TrafficSigns\img0489.jpg'	[]	[]	[119.0843,130.3146...	[]	[]	[]	[]	[]	[]	[]	[]
490	'C:\Users\Carlos Mendoza\Documents\TrafficSigns\img0490.jpg'	[]	[753,357.1517,...	[]	[]	[]	[]	[]	[]	[]	[]	[]
491	'C:\Users\Carlos Mendoza\Documents\TrafficSigns\img0491.jpg'	[497.9775,1...	[]	[]	[]	[]	[]	[]	[]	[]	[]	[]
492	'C:\Users\Carlos Mendoza\Documents\TrafficSigns\img0492.jpg'	[]	[]	[]	[]	[169,136.4775,123....	[]	[]	[]	[]	[]	[]
493	'C:\Users\Carlos Mendoza\Documents\TrafficSigns\img0493.jpg'	[1.1995e+03...	[]	[]	[]	[]	[]	[]	[]	[]	[]	[]
494	'C:\Users\Carlos Mendoza\Documents\TrafficSigns\img0494.jpg'	[147,236.73...	[]	[]	[]	[]	[]	[]	[]	[]	[]	[]
495	'C:\Users\Carlos Mendoza\Documents\TrafficSigns\img0495.jpg'	[]	[]	[270.2097,117.79,97]	[]	[]	[]	[]	[]	[]	[]	[]
496	'C:\Users\Carlos Mendoza\Documents\TrafficSigns\img0496.jpg'	[]	[]	[]	[]	[236,296.658...	[]	[]	[]	[]	[]	[212,178,77,124]
497	'C:\Users\Carlos Mendoza\Documents\TrafficSigns\img0497.jpg'	[]	[]	[]	[]	[]	[]	[]	[]	[]	[]	[216.0449,136.3876,...
498	'C:\Users\Carlos Mendoza\Documents\TrafficSigns\img0498.jpg'	[445,117.30...	[1.1913e+03,11....	[]	[]	[99,122,49,66]	[]	[]	[]	[]	[]	[340,124,57,45]
499	'C:\Users\Carlos Mendoza\Documents\TrafficSigns\img0499.jpg'	[481.2022,9...	[]	[]	[]	[]	[]	[]	[]	[]	[]	[362.99,61,59]
500	'C:\Users\Carlos Mendoza\Documents\TrafficSigns\img0500.jpg'	[]	[]	[]	[]	[204,151.8717,103....	[]	[]	[200.9270,302.1826,...	[]	[]	[]
501	'C:\Users\Carlos Mendoza\Documents\TrafficSigns\img0501.jpg'	[]	[]	[]	[]	[]	[]	[785.6217,540,11...	[]	[]	[]	[]

Fig. 4. Datos obtenidos en la detección de regiones de interés.

rectángulo delimitador para cada una de las señales de tráfico identificadas. A cada clase se le asignó un color diferente de rectángulo como se muestra en la Fig. 3.

Tras haber etiquetado las 1426 señales se creó una tabla con doce columnas, en la primera se guarda la trayectoria de almacenamiento local de la imagen, y en las columnas de la dos a la doce se almacena un vector de 1x4 con un formato [x, y, ancho, alto], que especifica las coordenadas de la ubicación de la esquina superior izquierda y el tamaño del rectángulo delimitador de la región de interés correspondiente. En caso de que no exista la etiqueta de una señal de tráfico en la imagen, se representa como una celda vacía. En la Fig. 4 se muestra un fragmento de los datos obtenidos.

El conjunto de imágenes fue dividido en 80% (1028 imágenes) para entrenamiento y 20% (256 imágenes) para validación. Posteriormente, se utilizaron los datos de la tabla para obtener de cada imagen una copia individual de las señales de tráfico presentes, por lo tanto, se extrajeron 1126 señales para entrenamiento y 300 para validación. Las imágenes se redimensionaron a un tamaño de 224x224 píxeles, debido a que es el tamaño de las entradas de la red ResNet-50 que se utilizará posteriormente.

2.3. Creación de una CNN con CIFAR-10

Para comenzar la clasificación de señales de tráfico se decidió utilizar una CNN, que se compone principalmente de dos partes: en la primera, se utilizan operaciones de convolución y de pooling para generar características profundas de los datos sin procesar; en la segunda, el perceptrón multicapa utiliza las características para asignar una clase a los datos de entrada [24].

Para realizar el entrenamiento de una CNN desde cero se requiere una gran cantidad de imágenes, debido a que solamente así se podrán aprender las diferentes características que representan a un objeto. Además, para implementar la red se necesita una cantidad considerable de recursos computacionales. Debido a que no se contaba con un conjunto de imágenes amplio, se optó por entrenar una CNN con CIFAR-10 [25], que contiene un conjunto de datos amplio, para posteriormente realizar una transferencia del aprendizaje obtenido.

La transferencia de aprendizaje consiste en utilizar el conocimiento de una red previamente entrenada que será aplicado para realizar una nueva tarea. Una ventaja de la transferencia de aprendizaje es que requiere una menor cantidad de datos y recursos computacionales para el entrenamiento [26].

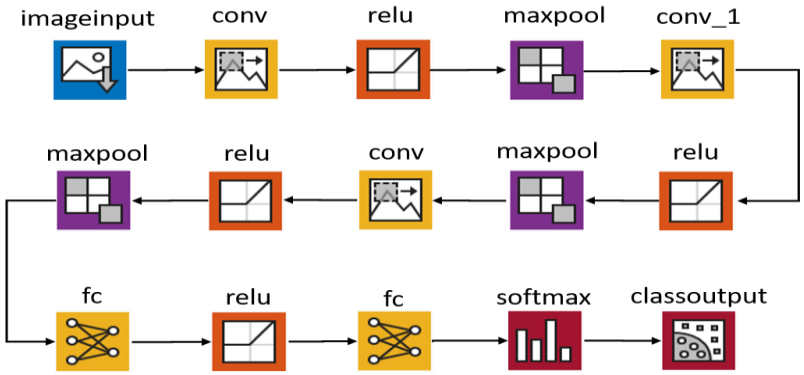

Fig. 5. Arquitectura de CNN para entrenamiento con CIFAR-10.

Se descargó CIFAR-10 con 50,000 imágenes RGB de 32x32 píxeles que incluye aviones, automóviles, pájaros, gatos, venados, perros, ranas, caballos, barcos y camiones. Aun cuando no se incluyen señales de tráfico, es de gran utilidad ya que le permitirá a la CNN aprender conceptos importantes y generalizables para identificar esquinas, bordes, texturas, formas geométricas, entre otros. Para el entrenamiento, y de acuerdo con la experiencia, se diseñó una arquitectura de CNN con 15 capas con una capa de entrada, varias capas de *Convolución*, *ReLU*, *Max Pooling y totalmente conectadas (fully connected)*, una S*oftmax* y una de salida (ver Fig. 5). Además, se utilizó el algoritmo del Descenso del Gradiente Estocástico con Momento (SGDM), porque reduce la oscilación que tiene el SGD en la función de pérdida de la red [15]. En la primera columna de la Tabla 2 se muestra un resumen de los valores establecidos para cada parámetro de la CNN. Una vez que se terminó el entrenamiento, se calculó la exactitud alcanzada por la CNN, que fue de 74.56%. Es importante aclarar que no es necesario alcanzar un 100%, sino la gran cantidad de aprendizaje obtenido con un conjunto de datos tan variado como CIFAR-10.

2.4. Entrenamiento de una R-CNN

El aprendizaje obtenido en el entrenamiento de la CNN con CIFAR-10 fue transferido para entrenar una R-CNN que pudiera aprender a detectar señales de tráfico. Una R-CNN se compone de dos etapas: la primera identifica un subconjunto de regiones en una imagen que puede contener un objeto (señal de tráfico), lo que ayuda a reducir el costo computacional al no buscar en la imagen completa, la segunda realiza la clasificación del objeto identificado [27].

Para el entrenamiento se necesitaron tres entradas: a) la tabla de datos obtenida en la subsección 2.2, b) los valores de los parámetros mostrados en la segunda columna de la Tabla 2 y c) la CNN de la subsección 2.3. Durante el entrenamiento, los pesos de la CNN entrenada con CIFAR-10 son ajustados tomando en cuenta los datos de la Fig. 4. Después, se seleccionan las porciones candidatas a ser señales de tráfico y se etiquetan como positivas o negativas. Las positivas (contienen señales) son aquellas que se superponen con los cuadros generados en la Fig. 4 en un rango de 0.5 a 1, medido por la intersección del cuadro delimitador (coordenadas espaciales) sobre la métrica de

unión. Mientras que las negativas (no contienen señales) son aquellas que se superponen en un rango de 0 a 0.3. Los valores de los rangos fueron seleccionados después de varias pruebas y también se utilizó el SGDM.

La estructura de la R-CNN es idéntica a la que se muestra en la Fig. 5. A la salida de la R-CNN se obtiene el o los objetos detectados enmarcados por un cuadro delimitador (bounding box), la confiabilidad de detección, y la etiqueta de la clase para cada objeto. La confiabilidad genera valores entre cero y uno, y mientras más cerca se encuentre el valor de uno, más confianza hay de que se identificó correctamente la señal de tráfico.

2.5. Creación de una CNN adicional usando como base ResNet-50

Para validar que las señales que clasifica la R-CNN sean las correctas (que detecte una señal de alto y que realmente sea un alto), se aplicó la técnica de transferencia de aprendizaje a la CNN ResNet-50. ResNet-50 es una red residual de 50 capas que fue entrenada con más de un millón de imágenes obtenidas del sitio ImageNet y puede clasificar imágenes en 1000 categorías, tales como teclados, ratones, plumas y muchos animales [28]. Para el entrenamiento, se requirió del repositorio de imágenes de entrenamiento que se creó en la subsección 2.2, el cual contiene 1126 imágenes de señales de tráfico de 224x224 píxeles. Las últimas tres capas de ResNet-50 (Fig. 6(a)), contienen información de cómo combinar las características de las imágenes obtenidas en las capas de convolución para obtener la probabilidad de pertenencia a una clase, el valor de pérdida, y la etiqueta calculada. Para adaptar ResNet-50 al problema de clasificación de señales, se realizó la transferencia de aprendizaje mediante el cambio de las tres últimas capas (ver Fig. 6(b)). Al final, se clasifican las 11 categorías de señales de tráfico, y no las 1000 categorías originales.

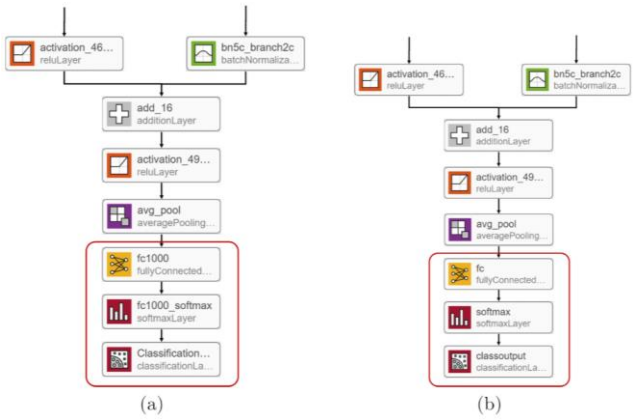

Fig. 6. Cambios realizados a la red ResNet-50. (a) Capas finales originales deResNet-50, (b) Capas finales modificadas a ResNet-50.

Para acelerar el entrenamiento, se congelaron las primeras diez capas al asignarles una tasa de aprendizaje de 0. Adicionalmente, y con el objetivo de evitar un sobreajuste de la red debido a la baja cantidad de imágenes, se realizó un proceso de aumentado de datos.

Tabla 2. Resumen de parámetros y valores utilizados para las redes.

Parámetros	CNN Valor	R-CNN Valor	ResNet-50 Valor
Momentum	0.9	0.9	0.9
InitialLearnRate	0.001	0.001	0.003
LearnRateSchedule	Piecewise	Piecewise	
LearnRateDropFactor	0.1	0.1	
LearnRateDropPeriod	8	100	100
L2Regularization	0.004	0.0001	0.0001
MaxEpochs	40	200	6
MiniBatchSize	128	128	10
Verbose	True	True	False
PositiveOverlapRange		0.5-1	
NegativaOverlapRange		0-0.3	
Shuffle	Once	Once	Every epoch
Validation Data			20%

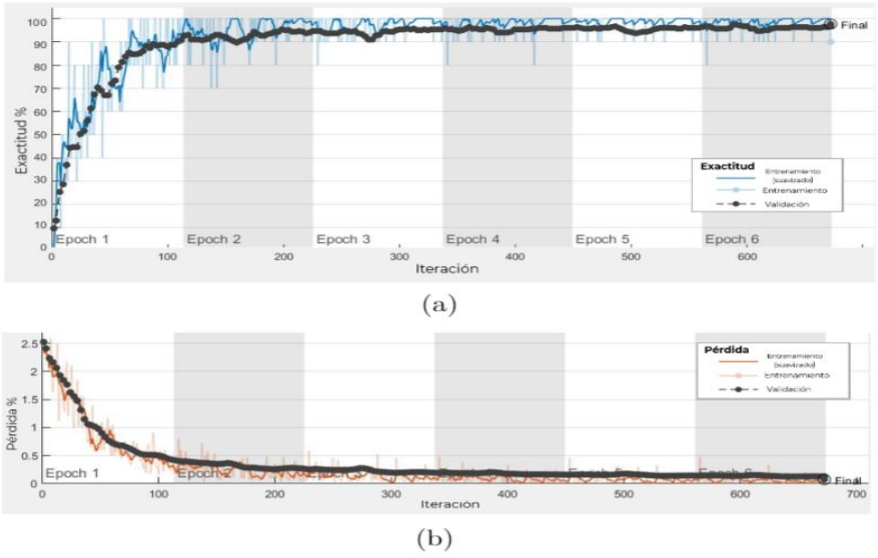

Fig. 7. Gráfica de (a) exactitud y (b) pérdida en el entrenamiento de la CNN.

El aumentado de datos es un proceso artificial para expandir el conjunto original de imágenes de entrenamiento por medio de la aplicación de operaciones de rotación, desplazamiento, y escalamiento, entre otras [29]. Las señales de tráfico fueron rotadas hasta 30 píxeles, y además se obtuvieron espejos verticales y horizontales. La ResNet-50 modificada se entrenó con los valores mostrados en la tercera columna de la Tabla

Tabla 3. Resultados obtenidos con ResNet-50 modificada.

Tipo	Total de señales	Coincidencia positiva	Coincidencia negativa	Confianza media	Exactitud
Señales Restrictivas					
Alto	56	56	0	96.73%	100%
Ceda el paso	13	13	0	99.86%	100%
Límite de velocidad	18	18	0	99.94%	100%
No estacionarse	51	51	0	95.68	100%
Vuelta prohibida	28	25	3	88.72%	89.29%
Señales Preventivas					
Glorieta	7	7	0	81.19%	100%
Niños jugando	14	13	1	93.62%	92.86%
Cruce peatonal	49	48	1	98.41%	97.96%
Tope	21	20	1	99.60%	95.24%
Señales Informativas					
Estac. discap.	29	28	1	89.78%	96.55%
Parada de autobús	14	14	0	97.08%	100%
Total	300	293	7	94.60%	97.67%

2. La gráfica de la exactitud alcanzada durante el entrenamiento se observa en la Fig. 7(a), mientras que la de la pérdida se muestra en la Fig. 7(b).

El entrenamiento se realizó en una computadora HP 15-bs1xx con procesador Intel Core i5 de octava generación y RAM de 8 GB; tomó 288 minutos y 54 segundos y se obtuvo una exactitud del 97.67% tras 672 iteraciones en 6 epochs.

3. Experimentación y resultados

Los experimentos se realizaron para verificar el desempeño de las redes creadas en la sección 2.4 y 2.5. Para la evaluación se utilizaron las medidas de exactitud y confianza media que se muestran en las ecuaciones 1 y 2, respectivamente:

$$\text{Exactitud} = \frac{\text{Coincidencias positivas}}{\text{Núm. de señales de tráfico}}, \tag{1}$$

$$\text{Confianza media} = \frac{\sum \text{Confianza}}{\text{Núm. de señales de tráfico}}. \tag{2}$$

Primero, se analizó la ResNet-50 modificada que se encarga de clasificar la señal de tráfico detectada por la R-CNN. Los resultados de la clasificación de las imágenes del set de validación se muestran en la Tabla 3. Como puede observarse, se obtiene una confianza media total de 94.60% y una exactitud total de 97.67%, los cuáles son valores competitivos con los que se han mostrado en la literatura, por ejemplo, en [3] se menciona una exactitud de 99.75% con una CNN. En la Fig. 8 se muestran ejemplos de la clasificación obtenida con su nivel de confianza correspondiente.

Fig. 8. Resultados de clasificación de señales de tráfico con ResNet-50 modificada.

Tabla 4. Resultados obtenidos con las redes R-CNN y ResNet-50 modificada.

Tipo	Total de señales	Detecciones/ Clasificaciones positivas	Detecciones/ Clasificaciones negativas
Señales Restrictivas			
Alto	56	51	5
Ceda el paso	13	13	0
Límite de velocidad	18	18	0
No estacionarse	51	51	0
Vuelta prohibida	28	26	2
Señales Preventivas			
Glorieta	7	6	1
Niños jugando	14	10	4
Cruce peatonal	49	47	2
Tope	21	21	0
Señales Informativas			
Estac. discap.	29	29	0
Parada de autobús	14	14	0
Total	300	286	14

Posteriormente, se evaluó de forma integral el proceso de detección realizado por la R-CNN y el de clasificación realizado por ResNet-50 modificada.

Fig. 9. Resultados obtenidos en la evaluación de R-CNN y ResNet-50 modificada.

Tabla 5. Resultados obtenidos con las imágenes sin señales de tráfico.

Tipo	Total de señales	Detecciones/ Clasificaciones positivas	Detecciones/ Clasificaciones negativas
Señales Restrictivas			
Alto	200	196	4
Ceda el paso	200	200	0
Límite de velocidad	200	198	2
No estacionarse	200	200	0
Vuelta prohibida	200	199	1
Señales Preventivas			
Glorieta	200	200	0
Niños jugando	200	200	0
Cruce peatonal	200	199	1
Tope	200	200	0
Señales Informativas			
Estac. discap.	200	200	0
Parada de autobús	200	197	3
Total	2200	2189	11

Para realizar la prueba se utilizó los datos de validación que fueran previamente separados de los datos de entrenamiento en la subsección 2.2. Los resultados de la evaluación se muestran en la Tabla 4.

En la Tabla 4, los valores de detección/clasificación negativa incluyen los casos en los que sí hay una señal de tráfico, pero no es detectada y clasificada. Por otra parte, los valores de detección positiva son los casos en los que la señal de tráfico sí es detectada y es clasificada correctamente. Por lo tanto, la exactitud final del sistema es de 95.33%. En la Fig. 9 se muestran cuatro ejemplos de la detección y clasificación de las señales de tráfico con el nivel de confianza correspondiente.

Finalmente, para verificar la robustez de la propuesta, se descargó de la app *Street View* de Google un nuevo conjunto con 62058 imágenes de avenidas, autopistas y centros comerciales ubicados en zonas céntricas de Pittsburgh, Orlando y Manhattan. Sin embargo, lo más importante es que las imágenes no contienen señales de tráfico. De dicho conjunto se seleccionaron aleatoriamente 2200 imágenes, 200 para probar cada una de las 11 clases. El objetivo de la prueba fue identificar los casos en los que se realizan detecciones falsas o un posible comportamiento inesperado en el uso de las redes R-CNN y ResNet-50 modificada. Los resultados obtenidos de la prueba se muestran en la Tabla 5.

Los valores de verdadero negativo hacen referencia al caso en que el sistema infirió que no existen señales de tráfico en la imagen y realmente no se encuentran. Mientras que los valores de falso negativo hacen referencia al caso donde el sistema infirió la existencia de alguna señal de tráfico en la imagen cuando realmente no se encuentra. Como puede observarse en la Tabla 5, la exactitud total de la prueba es de 99.5%, es decir, se cometieron 11 fallos en las 2200 imágenes.

3.1. Discusión

Los resultados obtenidos con la metodología propuesta son competitivos dado que permiten detectar y clasificar un subconjunto de señales de tráfico mexicanas divididas en 11 diferentes clases. Para el caso de la detección y clasificación de señales se observa que la de alto es en la que más se cometen errores, debido a su similitud en las formas y colores con la señal de vuelta prohibida.

También, se detectaron problemas para la señal de glorieta, seguida por la de no estacionarse y tope. Además, es importante resaltar que para las dos señales informativas no se obtuvieron errores.

Por otro lado, para el caso de imágenes que no contienen señales de tráfico, es también la señal de alto en la que más errores se observaron (cuatro), seguida de la señal de parada de autobús y de límite de velocidad. En la clase de señales preventivas es donde se obtienen más éxitos con solamente una equivocación.

4. Conclusiones

En el artículo se presentó una metodología para la detección y clasificación de señales de tráfico mexicanas que utiliza dos CNN y una R-CNN. Las pruebas se realizaron en el conjunto de validación que incluyó 256 imágenes que contienen un total de 300 señales de tráfico mexicanas.

Como resultado de los experimentos se observó que la propuesta obtiene una exactitud de 95.33% para la detección y clasificación de señales, y que además suele no confundirse cuando se le presentan imágenes que no contienen señales de tráfico con una exactitud del 99.5%.

En el futuro, se pretende ampliar el tamaño del conjunto de imágenes y realizar el entrenamiento con unas arquitecturas o detectores distintos con el fin de obtener mejores resultados en una menor cantidad de tiempo. Además, se considera la propuesta de utilizar bases de datos de señales de tráfico internacionales, para poder construir un modelo que pueda en cierta medida generalizar.

Referencias

1. Wali, S., Abdullah, M., Hannan, M., Hussain, A., Samad, S., Ker, P., Mansor, M.: Vision-based traffic sign detection and recognition systems: current trends and challenges. Sensors, 19, pp. 1–28 (2019)
2. Xu, H., Srivastava G.: Automatic recognition algorithm of traffic signs based on convolution neural network. Multimedia Tools and Applications, 79, pp. 11551–11565 (2020)
3. Cao, J., Song, C., Peng, S., Xiao, F., Song, S.: Improved traffic sign detection and recognition algorithm for intelligent vehicles. Sensors, 19, pp. 1–21 (2019)
4. Economic Commission for Europe: Convention on traffic signs and signals. Vienna Convention (1968)
5. Secretaria de Comunicaciones y Transportes: Tipos y Significado de las Señales de Tránsito. http://www.sct.gob.mx/carreteras/direccion-general-de-conservacion-de-carreteras/ publicaciones/senalamiento/ (2019)
6. Jin Y., Yusheng F., Wang W., Guo J., Ren C., Xiang X.: Multi-feature fusion and enhancement single shot detector for traffic sign recognition. IEEE, 8, pp. 38931–38940 (2020)
7. Swathi, M., Suresh, K.: Automatic traffic sign detection and recognition: a review. In: 2017 International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET), pp. 1–6, Chennai (2017)
8. Gudigar, A., Chokkadi, S., Raghavendra, U.: A review on automatic detection and recognition of traffic sign. Multimedia Tools and Applications, 75, pp. 333–364 (2016)
9. Yang, Y., Luo, H., Xu, H., Wu, F.: Towards real-time traffic sign detection and classification. IEEE Transactions on Intelligent Transportation Systems, 17(7), pp. 2022–2031 (2016)
10. De Oliveira, G., da Silva, F., Pereira, D., de Almeida, L., Artero, A., Bonora, A., de Albuquerque, V.: Automatic detection and recognition of text-based traffic signs from images. IEEE Latin America Transactions, 16(12), pp. 2947–2953 (2018)
11. Alghmgham, D., Latif, G., Alghazo, J., Alzubaidi, L.: Autonomous traffic sign (ATSR) detection and recognition using deep CNN. Procedia Computer Science, 163, pp. 266–274 (2019)
12. Yuan, X., Hao, X., Chen, H., Wei, X.: Robust traffic sign recognition based on color global and local oriented edge magnitude patterns. IEEE Transactions on Intelligent Transportation Systems, 15(4), pp. 1466–1477 (2014)
13. Serna, C., Ruichek, Y.: Classification of traffic signs: the European dataset. IEEE, 6(5), pp. 78136–78148 (2018)
14. Liu, C., Chang, F., Chen, Z., Liu, D.: Fast traffic sign recognition via high-contrast region extraction and extended sparse representation. IEEE Transactions on Intelligent Transportation Systems, 17(1), pp. 79–92 (2016)
15. Lee, H., Kim, K.: Simultaneous traffic sign detection and boundary estimation using convolutional neural network. IEEE Transactions on Intelligent Transportation Systems, 19(5), pp. 1652–1663 (2018)
16. Mannan, A., Javed, K., Rehman, A., Babri, H., Noon, S.: Classification of degraded traffic signs using flexible mixture model and transfer learning. IEEE, 7, pp. 148800–148813 (2019)
17. Zhu, Y., Liao, M., Yang, M., Liu, W.: cascaded segmentation-detection networks for text-based traffic sign detection. IEEE Transactions on Intelligent Transportation Systems, 19(1), pp. 209–219 (2018)
18. García, A., Alvarez, J., Soria-Morillo, L.: Deep neural network for traffic sign recognition systems: an analysis of spatial transformers and stochastic optimisation methods. Neural Networks, 99, pp. 158–165 (2018)

19. Liu, C., Li, S., Chang, F., Wang, Y.: Machine vision based traffic sign detection methods: review, analyses and perspectives. IEEE, 7, pp. 86578–86596 (2019)
20. Greenhalgh, J., Mirmehdi, M.: Recognizing text-based traffic signs. IEEE Transactions on Intelligent Transportation Systems, 16(3), pp. 1360–1369 (2015)
21. Luo, H., Yang, Y., Tong, B., Wu, F., Fan, B.: Traffic sign recognition using a multi-task convolutional neural network. IEEE Transactions on Intelligent Transportation Systems, 19(4), pp. 1100–1111 (2018)
22. Yuan, Y., Xiong, Z., Wang, Q.: An incremental framework for video-based traffic sign detection, tracking, and recognition. IEEE Transactions on Intelligent Transportation Systems, 18(7), pp. 1918–1929 (2017)
23. Li D., Jiang H.: On feature selection in network flow based traffic sign tracking models. Computers & Industrial Engineering, 127, pp. 657–664 (2019)
24. Guangle Y., Tao L., Jiandan Z.: A review of convolutional-neural-network-based action recognition. Pattern Recognition Letters, 118, pp. 14–22 (2019)
25. Krizhevsky, A., and G. Hinton.: Learning multiple layers of features from tiny images. Master's Thesis, University of Toronto (2009)
26. Weiss K., Khoshgoftaar T., Wang D.: A survey of transfer learning. Journal of Big Data, 3, pp. 31–40 (2016)
27. Zhang S., Wu R., Xu K., Wang J., Sun W.: R-CNN-based ship detection from high resolution remote sensing imagery. Remote Sensing, 11, pp. 1–15 (2019)
28. Wen L., Li X., Gao L.: A transfer convolutional neural network for fault diagnosis based on ResNet-50. Neural Computing and Applications, 32, pp. 6111–6124 (2019)
29. Shorten C., Khoshgoftaar, T.: A survey on image data augmentation for deep learning. Journal of Big Data, 6, pp. 1–48 (2019)