Revista mktDescubre - ESPOCH FADE
N° 17 Junio 2021, pp. 95 - 105
Revista mktDescubre - ESPOCH FADE
Registro IEPI Título Nº: 3232-12

ISSN en Línea: 2602-8522
Latindex Folio: 27399
Directorio Folio único 20003 Catálogo
Periodo: Julio - Diciembre 2020

PROPUESTA DE AUTOMATIZACIÓN DEL AFORO VEHICULAR A PARTIR DE IMÁGENES VIALES USANDO REDES NEURONALES CONVOLUCIONALES

Patricio Moreno Vallejo
✉ xavier.moreno@espoch.edu.ec
Escuela Superior Politécnica de Chimborazo - Ecuador

María Vallejo Sanaguano
✉ maria.vallejo@espoch.edu.ec
Escuela Superior Politécnica de Chimborazo - Ecuador

Gisel Bastidas Guacho
✉ gis.bastidas@espoch.edu.ec
Escuela Superior Politécnica de Chimborazo - Ecuador

RESUMEN

El conocimiento del flujo vehicular de una vía es un factor importante al momento de la planificación vial. Por lo tanto, en el presente estudio se propone un modelo de aprendizaje profundo basado en la arquitectura de redes neuronales Faster-RCNN para contabilizar el aforo vehicular sin la necesidad de supervisión humana. La metodología de investigación experimental utilizada permitió probar y optimizar el modelo candidato con el algoritmo de gradiente descendente para localizar vehículos y determinar el aforo vehicular en imágenes viales. El modelo propuesto recibe como entrada una imagen representada en una matriz y da como salida la imagen con los vehículos localizados. Una vez entrenado el modelo con imágenes de autopistas de China, se evaluó el rendimiento del mismo con imágenes de vías del Ecuador captadas por cámaras del ECU911 que se encuentran disponibles en la web. Como resultado de la evaluación se obtuvo un modelo funcional de aprendizaje profundo para la automatización del aforo vehicular con un 95% de certeza. A pesar de que el modelo fue entrenado con imágenes de otro país, los resultados muestran que el modelo se adapta de forma favorable a la realidad del Ecuador con un tiempo de inferencia óptimo de 0.28 segundos.

Palabras clave: Aforo Vehicular, Redes Neuronales Convolucionales, Inferencia, Clasificación.

ABSTRACT

The knowledge of the vehicular flow of a road is an essential factor in road planning. Therefore, in the present study, a deep learning model based on the Faster-RCNN neural network architecture is proposed to get the vehicle capacity of a road without the need for human supervision. The experimental research methodology applied allowed testing and optimizing the candidate model with the gradient descent algorithm to locate vehicles and determine the vehicle capacity in road images. The proposed model receives as input an image represented in a matrix and outputs the image with the localized vehicles. Once the model was trained with highway images in China, its performance was evaluated with pictures of roads in Ecuador which were captured by cameras of the ECU911 available on the web. As a result of the evaluation, a functional deep learning model was obtained for automating vehicle counting with an accuracy of 95%. Even though the model was trained with images from another country, the results show the model adapts favorably to the reality of Ecuador with an optimal inference time of 0.28 seconds.

Keywords: Vehicular counting, Convolutional Neural Network, Inference, Classification.

Fecha recepción: enero 2021
Fecha aceptación: junio 2021

1. INTRODUCCIÓN

Hoy en día el tránsito vehicular se ha incrementado considerablemente en las vías de pequeñas y grandes ciudades ocasionando congestiones vehiculares, por lo que, es necesario analizar el flujo de los vehículos mediante herramientas automatizadas que permitan apoyar la toma de decisiones para mejorar el tránsito vial. El aforo vehicular es utilizado para gestionar el tránsito en una vía determinada, sin embargo, muchas de las veces esta actividad se realiza de forma manual usando personal humano, lo cual es costoso y demorado. Existen diversos sistemas tecnológicos que permiten realizar dicha tarea con diferentes niveles de precisión y de forma temporal o permanente. Estos sistemas se pueden clasificar en sistemas instalados en pavimento que son los más intrusivos y sistemas elevados que no son intrusivos. Un sistema instalado en pavimento, por ejemplo, es el sistema de manguera neumática. Este sistema realiza el conteo de vehículos mediante la detección de la presión que ejercen los neumáticos sobre una manguera que se coloca en la vía. Otro sistema es el de lazos inductivos, el cual crea un campo magnético que permite detectar vehículos cuando se encuentran cerca. Por otra parte, dentro de los sistemas elevados se pueden encontrar sistemas que usan laser infrarrojo, radar a hiperfrecuencia, ultrasonido, etc. Adicionalmente, se pueden encontrar sistemas que realizan el aforo vehicular mediante el análisis de video, estos sistemas tienen la ventaja que se pueden usar cuando se desea aforos tanto temporales como permanentes. En este caso, el sensor viene a ser una cámara de video que simplemente debe captar la vía en la cual se desea hacer el aforo, pero a diferencia de otros sensores, su calibración es más sencilla. Sin embargo, para su funcionamiento es necesario un software especializado que analice los marcos de video para extraer información referente a los vehículos que transitan por la vía. Algunos sistemas utilizan técnicas tradicionales de visión por computadora que tienden a requerir una calibración que en ciertos casos puede llegar a ser compleja (Moreno, Bastidas & Moreno, 2020). En la actualidad, la calibración se sigue haciendo de forma manual, sin embargo, también se han creado técnicas experimentales que usan aprendizaje profundo como AutoCalib (Bhardwaj et al, 2018). Esta técnica extrae puntos característicos de los vehículos captados por las cámaras para posteriormente obtener los parámetros de calibración. Por otro lado, Xia et al. (2016) propone un método de bucle virtual para el conteo de vehículos usando un modelo de mixtura gaussiana y algunas operaciones morfológicas que mejoran la calidad de los vehículos detectados. Biswas et al. (2017) crea un framework denominado OverFeat para el conteo vehicular automático que se basa en una combinación de técnicas de aprendizaje profundo y de aprendizaje de máquina, sin embargo, los autores no indican los fotogramas por segundo a los que puede trabajar el framework, lo cual es importante al momento de realizar conteos en tiempo real a través de video. Zhu et al. (2018) utiliza imágenes aéreas de ultra alta resolución captadas por un drone para entrenar una red neuronal profunda que solo funciona con imágenes aéreas. Trivedi, Sarada & Dhara (2018) utilizan videos de vías disponibles en Youtube para procesarlos usando técnicas clásicas de visión por computadora como son la detección de bordes, métodos de distancia euclidiana, operaciones morfológicas, y desenfoque gaussiano con el fin de detectar vehículos. Asha & Narasimhadhan (2018) utilizan el algoritmo YOLO, propuesto por Redmon et al. (2016), para el conteo de los vehículos, el cual es conocido por su rapidez en la detección de objetos en tiempo

real. Liu et al. (2020) realizan el conteo del número de vehículos en una región de interés (ROI) mediante la detección, rastreo y contabilización de las trayectorias de los vehículos en movimiento. Para lograr este conteo utilizan el método de suavizamiento de distancias de Mahalanobis y Faster R-CNN como detector de objetos.

Por otra parte, el Ecuador cuenta con un sistema de videovigilancia del Servicio Integrado de Seguridad ECU911 que tiene alrededor de 4779 cámaras distribuidas en el territorio nacional. Estas cámaras generan miles de horas de video que en muchos de los casos solo se utilizan cuando existe un incidente. Los videos captados no siempre están listos para ser analizados por un computador, por lo que es necesario realizar un preprocesamiento que permita homogeneizar las imágenes que pueden corresponder al mismo lugar, pero verse diferentes al ser captadas por distintos dispositivos o bajo diversas circunstancias climáticas.

Adicionalmente, pueden existir imágenes captadas por cámaras en diferentes posiciones de la vía, es decir, algunas cámaras pueden captar toda la vía mientras otras solo captan una sección de la vía por estar ubicadas en una posición perpendicular o a un costado de la vía. Por otro lado, las imágenes también pueden tener varios vehículos u otros objetos que dificultan la localización de los objetos de interés. En el presente estudio se propone un modelo de automatización del aforo vehicular mediante la aplicación de técnicas de preprocesamiento y aprendizaje profundo, particularmente, se propone un modelo de Redes Neuronales Convolucionales (CNN) para localizar y clasificar vehículos en una vía de tal forma que se pueda contabilizar el aforo vehicular.

En este artículo, en la sección 2 se presentan los materiales y métodos usados en el presente estudio. En la sección 3, se describen los resultados de la evaluación del modelo propuesto. En la sección 4, se presenta la discusión de los aspectos más relevantes del estudio. Finalmente, en la sección 5 se presentan las conclusiones.

2. MATERIALES Y MÉTODOS

En esta sección se presenta la metodología usada para construir una propuesta de software para la automatización del aforo vehicular en imágenes viales mediante el uso de una red neuronal convolucional. El presente estudio es experimental debido a que se dispone de un conjunto de variables sobre las cuales se tiene control para realizar diversas pruebas utilizando software especializado. El estudio también es transversal dado que se parte de un conjunto de imágenes de vías captadas por una cámara de video en un tiempo determinado y tiene un enfoque cuantitativo puesto que durante el estudio se usan datos numéricos entre 0 y 255 que representan imágenes las cuales se utilizan para probar la hipótesis, es decir, probar un modelo candidato de redes neuronales convolucionales que se aproxime a una función objetivo desconocida la cual recibe como entradas imágenes de vías y da como salida la imagen con los vehículos localizados. El modelo se aproxima a la función ideal usando el algoritmo de gradiente descendiente el cual permite inferir las funciones para detectar los objetos que corresponden a vehículos y determinar sus cuadros delimitadores. Un cuadro delimitador es el rectángulo más pequeño que completamente rodea a un objeto, en este estudio se determinan cuadros delimitadores de los vehículos, como se muestra en la Figura 1.

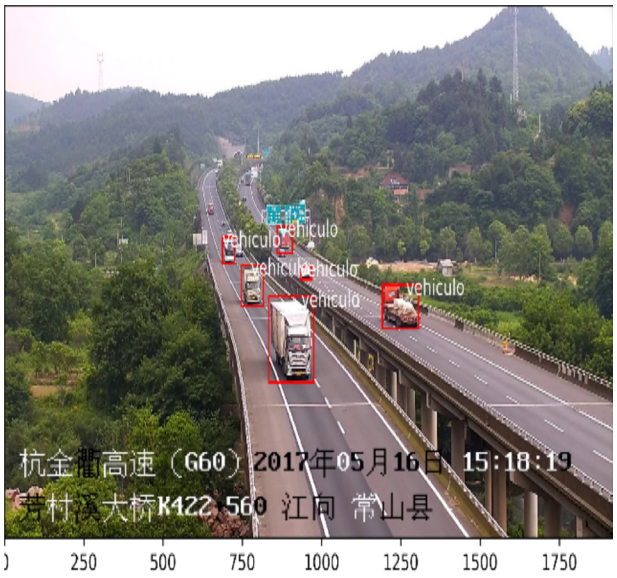

Figura 1. Cuadro delimitador de vehículos en color rojo.
Fuente: (Song et al, 2019)
Elaborado: por los autores

Además, se realiza un análisis cuantitativo del rendimiento del modelo propuesto para determina la efectividad del mismo. Por otro lado, el alcance de la investigación es explicativo debido a que se busca expresar el comportamiento de las variables dependientes que corresponden a las coordenadas (X, Y) de las esquinas opuestas de los cuadros delimitadores de los vehículos con respecto al conjunto de variables independientes que son cada uno de los pixeles de las imágenes de las vías. Asimismo, durante la investigación se utiliza el método inductivo dado que se entrena el modelo partiendo de un conjunto de datos específico que corresponde a vías captadas por cámaras de videos en China para después probarlo en un conjunto de datos de vías de otros países como Ecuador.

Datos

Los datos utilizados en esta investigación fueron obtenidos a partir de fuentes de información secundarias. El conjunto de datos para entrenar el modelo propuesto contiene 11,129 imágenes viales de un circuito cerrado de vigilancia (CCTV) de autopistas de la ciudad de Hangzhou, China. El CCTV está compuesto de 23 cámaras de vigilancia (Song et al, 2019). Las imágenes tienen una resolución de 1920x1080 pixeles y contienen desde vehículos livianos hasta vehículos pesados como buses y camiones. Adicionalmente, las imágenes cuentan con 57,290 anotaciones realizadas manualmente de cuadros delimitadores que localizan todos los vehículos en cada imagen. En la Figura 2 se muestra una imagen contenida en el conjunto de datos y los diferentes vehículos que se pueden encontrar en las vías.

Debido a que las imágenes de alta resolución requieren muchos recursos computacionales al momento de optimizar un modelo, las imágenes fueron preprocesadas cambiando su resolución a una resolución más baja de 416x740 pixeles. Además, para acelerar la convergencia del modelo a un punto óptimo, se removieron imágenes poco frecuentes como son las que contenían cuadros delimitadores con un ancho mayor a 300 pixeles, ya que estas imágenes correspondían a cámaras que enfocan la vía de muy cerca provocando que los vehículos se vean extremadamente grandes.

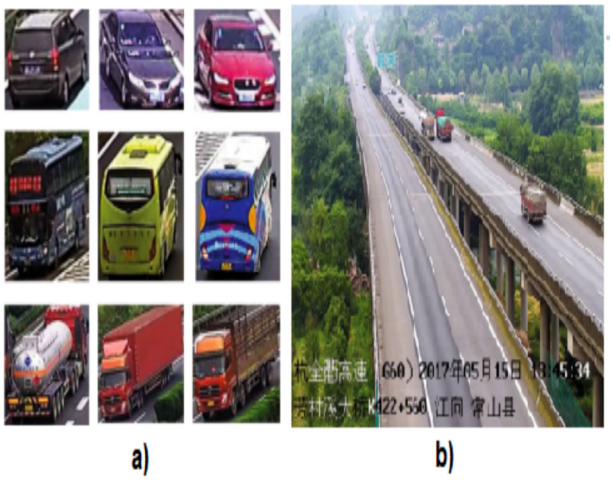

Figura 2: Conjunto de datos: a) Categorías de vehículos que se encuentran en las imágenes viales: carros, buses, camiones; b) Imagen vial capturada por el circuito de vigilancia.
Fuente: (Song et al, 2019)
Elaborado: por los autores

Por otro lado, es conocido que los datos no normalizados pueden provocar problemas al momento de la optimización de un modelo de aprendizaje profundo haciendo que el entrenamiento sea más lento e inestable. Razón por la que se normalizan los datos aplicando el centrado de pixeles en las imágenes para escalar los valores de los pixeles de las imágenes RGB de tal forma que el valor promedio de los pixeles en cada canal sea 0. Para realizar esta normalización, se obtuvo primeramente el promedio de los valores de los pixeles en cada canal: [115.92, 124.71, 120.50] y después se substrajo dichos valores de cada pixel por canal. Esto también ayuda para normalizar el contraste e iluminación de las imágenes.

Modelo

En el presente estudio se propone un modelo de red neuronal convolucional de tipo extremo a extremo que permite la localización de vehículos en una escena vial para contabilizar el aforo vehicular. El modelo se basa en la arquitectura Faster R-CNN planteada en (Ren et al, 2017), que es la última mejora de sus predecesoras (Girshick, 2015; Girshick et al, 2014). La arquitectura del modelo propuesto se muestra en la Figura 3. El modelo recibe como entrada los valores numéricos de una imagen de resolución 416x740 que está compuesta de tres canales donde cada canal contiene los valores que indican la cantidad del color primario Rojo, Verde y Azul de cada pixel basado en el modelo de color RGB. Por lo tanto, las entradas del modelo son imágenes representadas por 923,520 valores numéricos o características. Estas entradas alimentan al modelo que inicialmente utiliza transferencia de conocimiento al incluir en la arquitectura los cinco bloques convolucionales de la red VGG (Simonyan & Zisserman, 2015) inicializados con los pesos de ImageNet para extraer un mapa de características, el cual contiene solo la información más relevante de los datos de entrada. Luego, este mapa de características se usa para alimentar a una red de propuestas de regiones conocida como RPN por sus siglas en inglés Region Proposal Network. La RPN es la columna vertebral del modelo propuesto debido a que esta subred aprende a identificar las regiones en donde se encuentran los objetos de interés que en este caso son los vehículos. La RPN está compuesta por una capa convolucional de 512 canales con filtro de 3x3 y función de activación RELU; esta capa actúa como una ventana deslizante que va buscando información en todo el mapa de características, posteriormente, esta capa se conecta a otras dos capas convolucionales donde la primera es una capa convolucional conocida como capa de clasificación que tiene filtro de 1x1, con función de activación Sigmoid y 12 canales que corresponden al número de cajas de ancla; y la segunda es una capa convolucional conocida como capa de regresión que tiene filtro de 1x1, con función de activación lineal y 48 canales que corresponden al número de cajas de ancla multiplicado por 4. Las 12 cajas de ancla que se mencionan se obtienen debido a que se usan 4 cajas de escalas diferentes: 24, 48, 96, y 128; y cada una de estas cajas con 3 razones diferentes: 1:1, 1:2, y 2:1, dando un total de 4x3 = 12 cajas de ancla. Profundizando un poco más en las últimas dos capas de arquitectura de la RPN, se puede mencionar que la capa de clasificación se encarga de determinar la probabilidad de que cada una de las 12 cajas de anclaje contenga el objeto de interés, es decir, contenga un vehículo. Por otra parte, la capa de regresión se encarga de encontrar las coordenadas óptimas de los cuadros delimitados de cada caja

en caso de que contengan un vehículo. Cabe recalcar que en caso de que no se detecte un objeto de interés dentro de una caja de anclaje, esta no será considerada. En cambio, si se detecta un objeto de interés, la capa de regresión retornará las coordenadas de los cuadros delimitadores que se conocen como las propuestas de la RPN. Estas propuestas son concatenadas con el mismo mapa de características que inicialmente se usa como entrada de la RPN. Posteriormente, se usan dos capas densamente conectadas de 4096 neuronas, las cuales permiten extraer los pares de coordenadas (x1, y1) y (x2, y2) que corresponden a las esquinas opuestas del rectángulo que delimita la localización del vehículo en la imagen de entrada usando regresión. Finalmente, usando las coordenadas de salida, se grafican los cuadros delimitadores en la imagen de entrada y una vez localizados los vehículos se contabiliza el aforo vehicular.

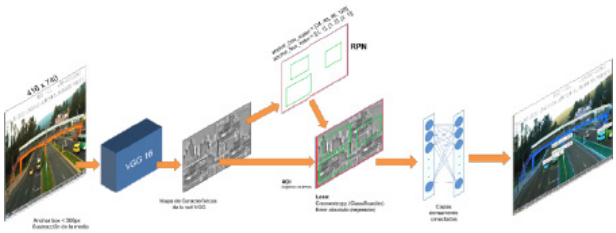

Figura 3: Modelo propuesto para la automatización del aforo vehicular.
Fuente: propia
Elaborado: por los autores

Para optimizar el modelo se utiliza ADAM (Kingma & Ba, 2015) con la función de perdida entropía cruzada binaria para la RPN, entropía cruzada categórica para la clasificación final y la diferencia absoluta para las secciones de regresión que obtienen las regiones de interés (ROI) y los cuadros delimitadores de los vehículos. Adicionalmente, se utiliza la técnica de dropout, la cual aleatoriamente selecciona un porcentaje de neuronas y las ignora durante el entrenamiento del modelo. De esta forma se evita el sobreajuste del modelo a los datos de entrenamiento y se puede adaptar a nuevos datos.

El modelo propuesto fue implementado en el lenguaje de programación Python usando las librerías de Tensorflow y Keras. Debido a que las imágenes y los mapas de características son representados como matrices multidimensionales o tensores, fue necesario disponer de la librería numpy para realizar operaciones matemáticas con los tensores. Se utilizó también la librería de OpenCV para preprocesar las imágenes.

3. RESULTADOS

En esta sección, se evalúa el rendimiento del modelo de redes neuronales convolucionales para la automatización del aforo vehicular propuesto en el presente estudio.

El modelo propuesto se entrenó durante 85 épocas pasando 1000 imágenes viales en cada época. Para la optimización se utilizó una tarjeta gráfica NVidia K80 que permitió que cada época se ejecute en promedio en 823 segundos. Al final, el entrenamiento del modelo tomó 19.43 horas.

La Figura 4 muestra la evolución de la pérdida causada por la RPN al modelo propuesto durante el proceso de entrenamiento. En la primera época la pérdida empieza en 0.338 y va reduciendo con el paso de las épocas hasta finalizar con una pérdida de 0.064. En la Figura 5 se puede observar la pérdida de la última subred que tiene una evolución decreciente similar a la de la subred RPN hasta terminar con una pérdida de 0.045.

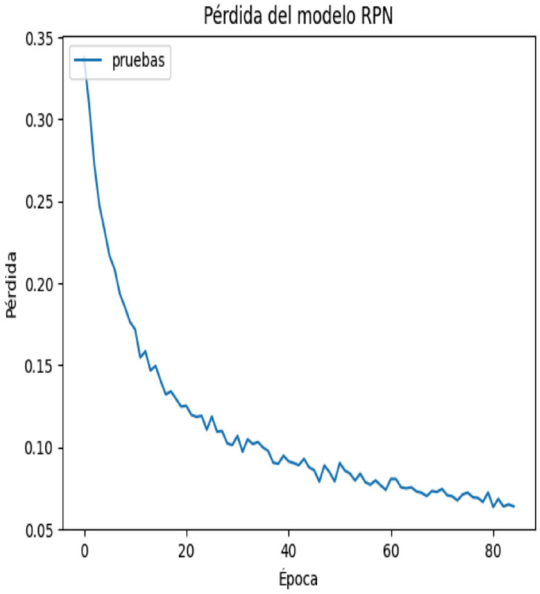

Figura 4: Pérdida de la RPN según el número de épocas.
Fuente: Propia
Elaborado: por los autores

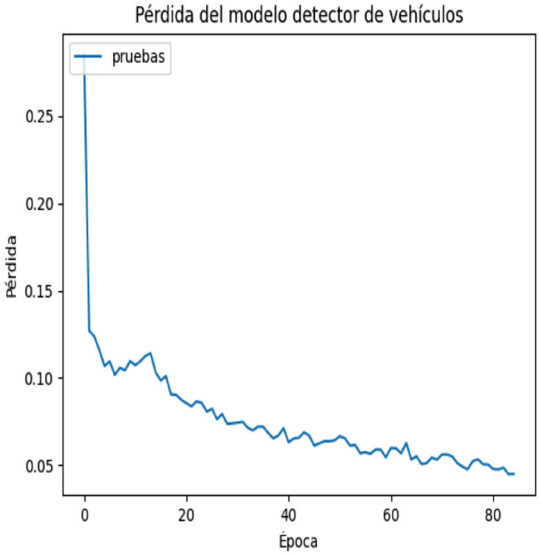

Figura 5. Pérdida de la subred densamente conectada según el número de épocas.
Fuente: Propia
Elaborado: por los autores

En la Figura 6 se observa la pérdida total del modelo que incluye las pérdidas causadas por las diferentes subredes del modelo tanto en los procesos de clasificación como de regresión. La pérdida total tiene una evolución decreciente al igual que las pérdidas de las subredes que conforman el modelo propuesto, llegando a una pérdida total final de 0.410. De esta forma se puede observar que la optimización del modelo es estable al observarse que la pérdida va decrementando conforme se va avanzando de época a época. Por cuestiones de recursos y tiempo no fue posible optimizar por más épocas el modelo propuesto, sin embargo, acorde a los resultados obtenidos se puede observar el mejoramiento del modelo con el paso de las épocas y la disminución considerable de la pérdida total.

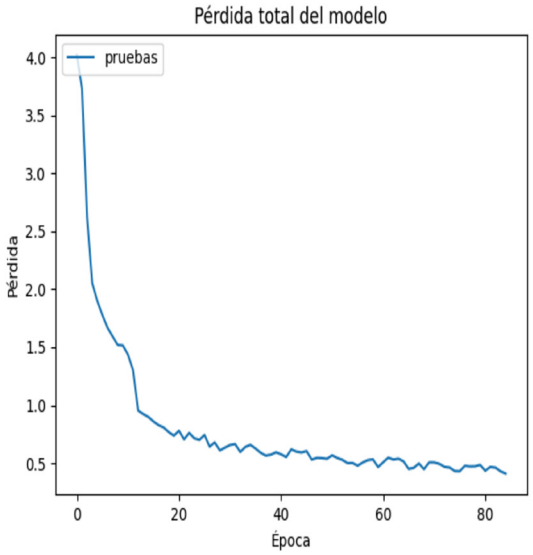

Figura 4: Pérdida de la RPN según el número de épocas.
Fuente: Propia
Elaborado: por los autores

La Figura 7 presenta la evolución de la certeza del modelo en lo que respecta a la localización de objetos de interés en los mapas de características de las imágenes de entrada. La mejor certeza se alcanza en la época 84 con un valor del 95%.

Adicionalmente, para evaluar el modelo propuesto, se recopilaron imágenes

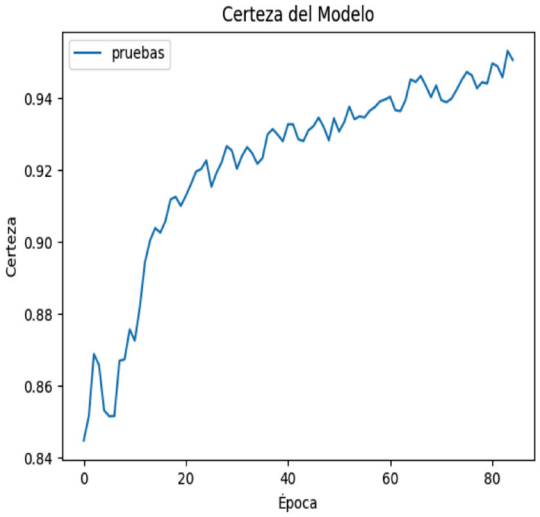

Figura 7. Certeza del modelo propuesto.
Fuente: Propia
Elaborado: por los autores

públicas disponibles en internet que contienen escenas de vías del Ecuador captadas por cámaras de videovigilancia del servicio Integrado de Seguridad del ECU911. Las imágenes recopiladas se usaron como entrada del modelo propuesto y de esta forma se pudo probar el mismo, obteniendo resultados referentes a la localización de vehículos y aforo vehicular. Como se observa en la Figura 8 el modelo es capaz de detectar la mayoría de los vehículos a pesar de ser una imagen completamente nueva que no fue presentada al modelo durante la etapa de entrenamiento y optimización del modelo, lo cual indica que el modelo propuesto es robusto al momento de predecir automáticamente el aforo vehicular en escenas viales nuevas.

Por otro lado, se implementó una aplicación de escritorio la cual usa el modelo propuesto y permite realizar la inferencia de la localización y conteo de vehículos en escenas viales captadas a través de imágenes. El tiempo promedio que le toma al sistema hacer la predicción es de 0.28 segundos, esto significa que el modelo propuesto puede ser aplicado en sistemas de video en tiempo real dado que se podría hacer aproximadamente 3 inferencias por segundo en un video.

Figura 8: Localización de vehículos usando el modelo propuesto en una imagen captada por el ECU911
Fuente: Propia
Elaborado: por los autores

4. DISCUSIÓN

La presente investigación busca automatizar el aforo vehicular utilizando videos que capturen escenas de vías mediante la propuesta de un modelo de redes neuronales de extremo a extremo que permita realizar inferencias de la localización de vehículos en un tiempo óptimo. Mediante el uso de redes convolucionales y redes densamente conectadas se logra optimizar un modelo que es capaz de realizar inferencias en un tiempo promedio de 0.28 segundos en imágenes nuevas. A diferencia de otros trabajos como el de Song et al. (2019), se ha realizado un análisis de la robustez del modelo en imágenes de entrenamiento y pruebas con diferentes

distribuciones, es decir, se ha realizado pruebas usando imágenes captadas por cámaras de video de vías del Ecuador que no tienen ninguna relación con imágenes de autopistas Chinas que se usaron en el entrenamiento del modelo. Algunos estudios han utilizado como base para la localización de vehículos el algoritmo YOLO (Redmon et al, 2016) como son los estudios de Asha & Narasimhadhan (2018) y Song et al. (2019). Sin embargo, el algoritmo Faster-RCNN puede alcanzar un mejor rendimiento o un rendimiento similar en métricas de precisión al algoritmo YOLO según Benjdira et al. (2019). En otros trabajos (Asha & Narasimhadhan , 2018; Liu et al, 2020; Xia et al, 2016) solo se presentan resultados teóricos sin una aplicación para los usuarios finales, por lo tanto, en el presente trabajo se propone una aplicación que permite cargar las imágenes y realizar la inferencia usando el modelo propuesto de forma transparente para el usuario final. Como trabajo futuro se puede mejorar la certeza del modelo modificando los hiperparámetros o cambiado la subred VGG por una arquitectura más actual como la ResNet (He et al., 2016).

5. CONCLUSIONES

- En este artículo, se propone un modelo para automatizar el aforo vehicular a partir de imágenes de escenas viales, el cual se basa en la arquitectura Faster R-CNN. El modelo propuesto aprovecha la transferencia de conocimiento al usar los pesos de una red VGG pre-entrenada para clasificar objetos. Además, usa la RPN para determinar las regiones de interés e inferir las coordenadas de los cuadros delimitadores de los vehículos en el mapa de características de una imagen vial. Este modelo alcanzó una certeza del 95% en la localización de vehículos y automatización del aforo vehicular. Adicionalmente, el modelo mostró robustez al alcanzar una certeza similar cuando fue probado con imágenes viales de Ecuador que no fueron utilizadas durante el entrenamiento del modelo. El tiempo de predicción promedio de 0.28 segundos es mínimo, por lo que el modelo propuesto podría ser usado en sistemas de tiempo real.

6. REFERENCIAS BIBLIOGRÁFICAS

1. Asha, C. S., & Narasimhadhan, A. V. (2018). Vehicle Counting for Traffic Management System using YOLO and Correlation Filter. 2018 IEEE International Conference on Electronics, Computing and Communication Technologies, CONECCT 2018. https://doi.org/10.1109/CONECCT.2018.8482380
2. Benjdira, B., Khursheed, T., Koubaa, A., Ammar, A., & Ouni, K. (2019). Car Detection using Unmanned Aerial Vehicles : Comparison between Faster R-CNN and YOLOv3. 2019 1st International Conference on Unmanned Vehicle Systems-Oman (UVS), 1–6.
3. Bhardwaj, R., Tummala, G. K., Ramalingam, G., Ramjee, R., & Sinha, P. (2018). AutoCalib: Automatic traffic camera calibration at scale. ACM Transactions on Sensor Networks, 14(3–4), 1–27. https://doi.org/10.1145/3199667

4. Biswas, D., Su, H., Wang, C., Blankenship, J., & Stevanovic, A. (2017). An Automatic Car Counting System Using OverFeat Framework. Sensors, 17(7), 1535. https://doi.org/10.3390/s17071535
5. Girshick, R. (2015). Fast R-CNN. Proceedings of the IEEE International Conference on Computer Vision, 2015 Inter, 1440–1448. https://doi.org/10.1109/ICCV.2015.169
6. Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 580–587. https://doi.org/10.1109/CVPR.2014.81
7. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Identity mappings in deep residual networks. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9908 LNCS, 630–645. https://doi.org/10.1007/978-3-319-46493-0_38
8. Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings. https://arxiv.org/abs/1412.6980v9
9. Liu, Z., Zhang, W., Gao, X., Meng, H., Tan, X., Zhu, X., Xue, Z., Ye, X., Zhang, H., Wen, S., & Ding, E. (2020). Robust Movement-Specific Vehicle Counting at Crowded Intersections.
10. Moreno Vallejo, P. X., Bastidas Guacho, G. K., & Moreno Costales, P. R. (2020). Estudio de factibilidad del uso de modelos de redes neuronales artificiales en la automatización del aforo y clasificación vehicular del transporte público. ConcienciaDigital, 3(3), 528–540. https://doi.org/10.33262/concienciadigital.v3i3.1355
11. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016-December, 779–788. https://doi.org/10.1109/CVPR.2016.91
12. Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031
13. Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, 1–14.
14. Song, H., Liang, H., Li, H., Dai, Z., & Yun, X. (2019). Vision-based vehicle detection and counting system using deep learning in highway scenes. European Transport Research Review, 11(1), 51. https://doi.org/10.1186/s12544-019-0390-4
15. Trivedi, J., Sarada Devi, M., & Dhara, D. (2018). Vehicle Counting Module Design in Small Scale for Traffic Management in Smart City Dimensional reduction View project Intelligent Transportation System

Using Computer Vision View project Vehicle Counting Module Design in Small Scale for Traffic Management in Smart City. https://doi.org/10.1109/I2CT.2018.8529506

16. Xia, Y., Shi, X., Song, G., Geng, Q., & Liu, Y. (2016). Towards improving quality of video-based vehicle counting method for traffic flow estimation. Signal Processing, 120, 672–681. https://doi.org/10.1016/j.sigpro.2014.10.035
17. Zhu, J., Sun, K., Jia, S., Li, Q., Hou, X., Lin, W., Liu, B., & Qiu, G. (2018). Urban Traffic Density Estimation Based on Ultra High Resolution UAV Video and Deep Neural Network. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 4968–4981.