



# DETECCIÓN DE PEATONES EN LA NOCHE USANDO FASTER R-CNN E IMÁGENES INFRARROJAS

## PEDESTRIAN DETECTION AT NIGHT BY USING FASTER R-CNN Y INFRARED IMAGES

Michelle Galarza Bravo<sup>1,\*</sup>, Marco Flores Calero<sup>2</sup>

### Resumen

En este artículo se presenta un sistema de detección de peatones en la noche, para aplicaciones en seguridad vehicular. Para este desarrollo se ha analizado el desempeño del algoritmo Faster R-CNN con imágenes en el infrarrojo lejano. Por lo que se constató que presenta inconvenientes a la hora de detectar peatones a larga distancia. En consecuencia, se presenta una nueva arquitectura Faster R-CNN dedicada a la detección en múltiples escalas, mediante dos generadores de regiones de interés (ROI) dedicados a peatones a corta y larga distancia, denominados RPNCD y RPNLD, respectivamente. Esta arquitectura ha sido comparada con los modelos para Faster R-CNN que han presentado los mejores resultados, como son VGG-16 y Resnet 101. Los resultados experimentales se han desarrollado sobre las bases de datos CVC-09 y LSIFIR, los cuales demostraron mejoras, especialmente en la detección de peatones a larga distancia, presentando una tasa de error versus FPPI de 16 % y sobre la curva Precisión *vs.* Recall un AP de 89,85 % para la clase peatón y un mAP de 90 % sobre el conjunto de pruebas de las bases de datos LSIFIR y CVC-09.

**Palabras clave:** peatón, infrarrojo, Faster R-CNN, RPN, múltiples escalas, noche.

### Abstract

This paper presents a system for pedestrian detection at nighttime conditions for vehicular safety applications. For this purpose, it analyzes the performance of the Faster R-CNN algorithm for infrared images. The research reveals that Faster R-CNN has problems to detect small scale pedestrians. For this reason, it introduces a new Faster R-CNN architecture focused on multi-scale detection, through two ROI's generators for large size and small size pedestrians, RPNCD and RPNLD respectively. This architecture has been compared with the best Faster R-CNN baseline models, VGG-16 and Resnet 101, which present the best results. The experimental results have been development on CVC-09 and LSIFIR databases, which show improvements specially when detecting pedestrians that are far away, over the DET curve presents the miss rate versus FPPI of 16 and over the Precision vs Recall the AP of 89.85 for pedestrian class and the mAP of 90 over LSIFIR and CVC-09 test sets.

**Keywords:** pedestrian, infrared, Faster R-CNN, RPN, multi-scale, nighttime.

<sup>1,\*</sup>Carrera de Ingeniería en Electrónica, Automatización y Control, Universidad de las Fuerzas Armadas ESPE Sangolquí – Ecuador. Autor para correspondencia ✉: [magalarza@espe.edu.ec](mailto:magalarza@espe.edu.ec), <https://orcid.org/0000-0001-8401-1871>

<sup>2</sup>Departamento de Eléctrica y Electrónica, Universidad de las Fuerzas Armadas ESPE, Sangolquí – Ecuador. <https://orcid.org/0000-0001-7507-3325>

Recibido: 02-05-2018, aprobado tras revisión: 18-06-2018

Forma sugerida de citación: Galarza Bravo, M. y Flores Calero, M. (2018). «Detección de peatones en la noche usando Faster R-CNN e imágenes infrarrojas». INGENIUS. N.º 20, (julio-diciembre). pp. 48-57. DOI: <https://doi.org/10.17163/ings.n20.2018.05>.

## 1. Introducción

Los sistemas de detección de peatones (SDP) son uno de los componentes tecnológicos más importantes que han surgido en los últimos años con el desarrollo de robótica móvil aplicada al sector automotriz y otras tecnologías similares destinadas a la seguridad vehicular [1], las cuales necesitan operar con altos estándares de calidad y tener una alta eficacia y precisión, debido a que su objetivo es proteger la vida humana a través de evitar que suceda un atropellamiento [2].

Varios informes, a nivel mundial, indican que los accidentes de tránsito generan altos costos materiales y humanos [3], donde los peatones tienen un alto porcentaje de accidentabilidad, llegando hasta el 22 % [4]. En el caso de Ecuador, los atropellamientos representan más del 10 % de las defunciones por accidentes de tránsito [5]. Por lo tanto, la detección de peatones es un tema de investigación activo y desafiante debido a la complejidad de la escena vial, la cual cambia constantemente debido a varios factores, por ejemplo, las condiciones atmosféricas contribuyen a una baja visibilidad y a un cambio permanente de la iluminación, las oclusiones generan información incompleta de la forma humana, la distancia perjudica la calidad de la información visual [1, 6, 7]. En la noche estos percances se magnifican debido a los ambientes de oscuridad [1, 2, 8, 9].

Por otra parte, debido al reciente éxito que han presentado las técnicas de aprendizaje profundo (*Deep Learning*) [10, 11], el principal objetivo de este trabajo es poner en marcha un método para la detección de peatones en la noche usando información visual en el infrarrojo lejano y las redes neuronales convolucionales, específicamente las arquitecturas del tipo Faster R-CNN [9, 11–15] para obtener un sistema competitivo que genere resultados de vanguardia comparables a los existentes en los trabajos previos. Por lo tanto, se presenta una nueva arquitectura Faster R-CNN a múltiples escalas, la cual es evaluada bajo los conjuntos de prueba de las bases de datos CVC-09 [16] y LSIFIR [17]. Los resultados evidencian mejoras especialmente al detectar peatones que se encuentran en la lejanía.

El documento está organizado de la siguiente manera. A partir de la segunda sección se presentan los métodos y materiales usados, donde se detallan los trabajos previos realizados en el campo de los SDP en especial técnicas de aprendizaje profundo. Adicionalmente, se describe el diseño propuesto de la nueva arquitectura Faster R-CNN para la generación de regiones de interés, la clasificación y la detección de peatones durante la noche, seguido de la evaluación experimental para distintas configuraciones del modelo propuesto. Posteriormente en la sección de resultados y discusión se exhiben los valores obtenidos respecto a la calidad de detección sobre las bases de datos destinadas al desarrollo de SDP en la noche. Finalmente,

la última sección está dedicada a las conclusiones, recomendaciones y los trabajos futuros que se pueden efectuar para mejorar esta propuesta.

## 2. Métodos y materiales

### 2.1. Trabajos previos

Actualmente, existen múltiples investigaciones especializadas en la detección de peatones en la noche [1, 2, 7–9, 15, 18–30]. Para llevar a cabo este proceso, generalmente, se divide el trabajo en dos partes, la primera consiste en la generación de ROI; y la segunda en la clasificación de las mismas, en peatones o el fondo, de esta manera, se logra mantener localizada a la persona mientras permanece en la escena.

#### 2.1.1. Generación de ROI sobre imágenes en el infrarrojo lejano

Para la generación de ROI sobre imágenes en el infrarrojo existen varios métodos, los más populares son: ventanas deslizantes (*Sliding window*) [18] que buscan exhaustivamente sobre toda la imagen en varias escalas, lo que hace que el método demande muchos recursos computacionales y no ser efectivo para aplicaciones en tiempo real. Para subsanar estos inconvenientes se han creado nuevas propuestas, por ejemplo, segmentación por movimiento, propuesto por Chen *et al.* [19] donde utilizando PCA y técnicas Fuzzy identifican regiones de interés local. Kim y Lee [21] han desarrollado un método que combina segmentos de imagen en lugar de umbrales y las bajas frecuencias de las imágenes en el infrarrojo lejano. Ge *et al.* [22] han propuesto un método de segmentación adaptativo compuesto de dos umbrales, uno especializado para localizar zonas brillantes y otro para zonas de poco contraste. Chun *et al.* [31] aplica detección de bordes para obtener un generador de ROI más rápido.

En la actualidad existen métodos más sofisticados que usan modelos de redes neuronales convolucionales, y sus variantes, para la generación de nuevas propuestas [1, 9, 12, 18]. Así, la detección de puntos de calor en resolución multiespectral usando IFCNN (*Illumination Fully Connected Neural Network*) ha sido propuesta por Guan *et al.* [8]. Vijay *et al.* [20] añaden una red neuronal convolucional al trabajo de Chen *et al.* [19], para la clasificación. Kim *et al.* [23] han usado cámaras en el espectro visible para detectar peatones en la noche usando CNN. Otras alternativas la red de propuestas de región (Region Proposal Network o RPN), inicialmente se centró en localizar las ROI mediante una combinación de búsqueda exhaustiva y ventanas deslizantes, en tres orientaciones y tres escalas (9 cajas de referencia) por cada ventana deslizante. Cada propuesta inicial sirve para el entrenamiento de una red completamente convolucional para generar las

predicciones del cuadro delimitador y los puntajes de probabilidad [12].

### 2.1.2. Clasificación de peatones sobre imágenes en el infrarrojo lejano

Los métodos desarrollados para la clasificación se pueden agrupar en dos categorías: los modelos basados en la generación manual de características [24, 25, 32], y los modelos de aprendizaje automático de características usando técnicas de aprendizaje profundo (DL, *Deep Learning*) [8, 11, 33–38].

En el primer caso se usan distintos métodos manuales de generación de características junto con un algoritmo de clasificación, algunos ejemplos son: HOG + SVM [26, 27], HOG + Adaboost [28], HOG + LUV [39], Haar + Adaboost [29], Haar + HOG y SVM [30]. En la segunda categoría están las redes neuronales convolucionales (CNN) [2, 8, 11, 34, 38], con sus distintas arquitecturas, como son R-CNN [40], Fast R-CNN [41] y Faster R-CNN [12, 15].

La arquitectura Fast R-CNN [12, 15] esencialmente disminuye la carga computacional, respecto a CNN, y por esta razón disminuye el tiempo de detección que presenta la capa R-CNN [41]. En consecuencia, Fast R-CNN junto con búsqueda selectiva, presenta una mejor calidad de detección. Sin embargo, ambos métodos necesitan de un generador de ROI externo y tienen problemas al momento de detectar objetos pequeños que, en el contexto de los peatones, implica largas distancias [41, 42].

Para remediar estos inconvenientes se ha llegado a Faster R-CNN [12, 15] que añade un generador de ROI basado en capas completamente conectadas RPN el cual comparte con Fast R-CNN [15], los mapas de características generados por la red convolucional. Por ende, se puede implantar redes muy profundas debido a que la imagen total pasa una sola vez por la etapa CNN [15].

Por lo tanto, Faster R-CNN está siendo utilizada ampliamente para construir SDP [1, 9, 42]. Por ejemplo, en [1] se ha empleado Faster R-CNN para detección de peatones en múltiples espectros, inicialmente se ha entrenado Faster R-CNN únicamente con imágenes a color e infrarrojas, Faster RCNN-C y Faster RCNN-T respectivamente, utilizando para el entrenamiento un nuevo modelo de red neuronal. Posteriormente se han combinado características en diferentes etapas creando así los modelos Early Fusion, Halfway Fusion, Late Fusion y Score Fusion. Adicionalmente, Wang *et al.* [9], tomando como referencia a Liang *et al.* [41], combina RPN + BDT para construir un sistema de detección de peatones en múltiples espectros. Sin embargo, se considera que Faster RCNN no funciona muy bien para

la detección de peatones, debido a que los mapas de características no presentan la información suficiente para peatones a larga distancia. Por esta razón, Feris *et al.* [43] han propuesto una subred para la generación de ROI en múltiples escalas junto con una subred para la clasificación basada en Fast R-CNN.

## 2.2. Sistema de detección de peatones en la noche

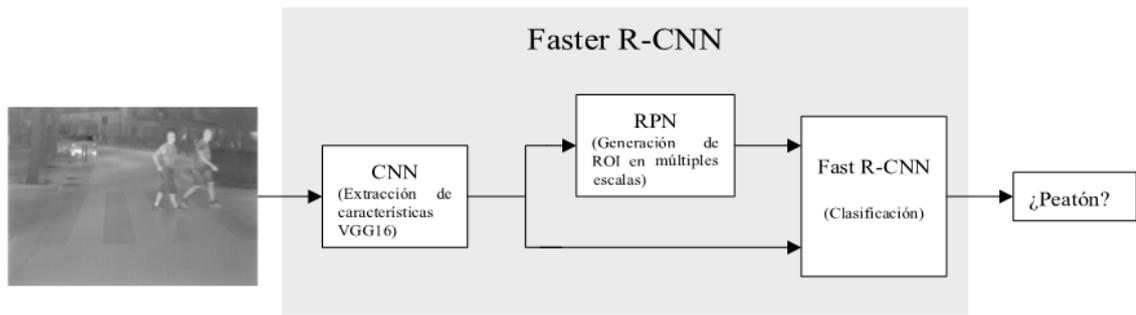
La Figura 1 muestra el esquema propuesto para el desarrollo del SDP en la noche, usando imágenes tomadas con iluminación infrarroja y como arquitectura base Faster R-CNN junto con el modelo VGG16 [44] donde se han desarrollado algunos cambios detallados a continuación.

### 2.2.1. Generación de ROI sobre imágenes en el infrarrojo lejano

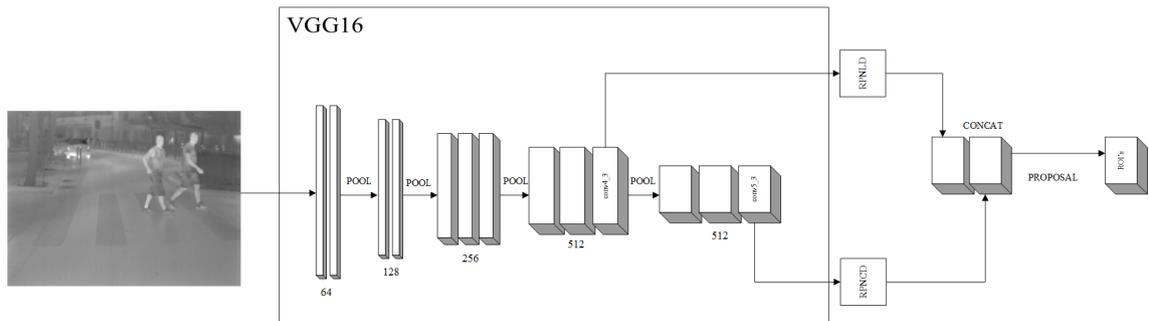
Debido a que la arquitectura original de Faster RCNN [12, 15] presenta problemas de detección en el caso de peatones que se encuentran en la lejanía, se considera la arquitectura desarrollada en Feris *et al.* [43]. Por lo tanto, se ha decidido colocar dos redes de propuestas de región (RPN) independientes, que presentan diferentes características, las mismas que están detalladas en la Tabla 2. En ambos casos, con un enfoque dirigido a peatones a corta (RPNCD) y larga distancia (RPNLD). Como se muestra en la Figura 2, RPNLD es alimentado por las características que son proporcionadas por la capa conv4\_3 de VGG16 [44], debido a que las redes de agrupación pueden discriminar peatones que se encuentren en la lejanía, donde los mapas de características más abundantes son beneficiosos para detectar peatones a largas distancias [6]. En cuanto a RPNCD al igual que la arquitectura original de Faster R-CNN [12] es alimentado por las características entregadas por la capa conv5\_3, ya que extrae las características más representativas presentes en la imagen, por esta razón proporciona excelentes resultados para peatones a corta distancia.

### 2.2.2. Clasificación de ROI sobre imágenes en el infrarrojo lejano

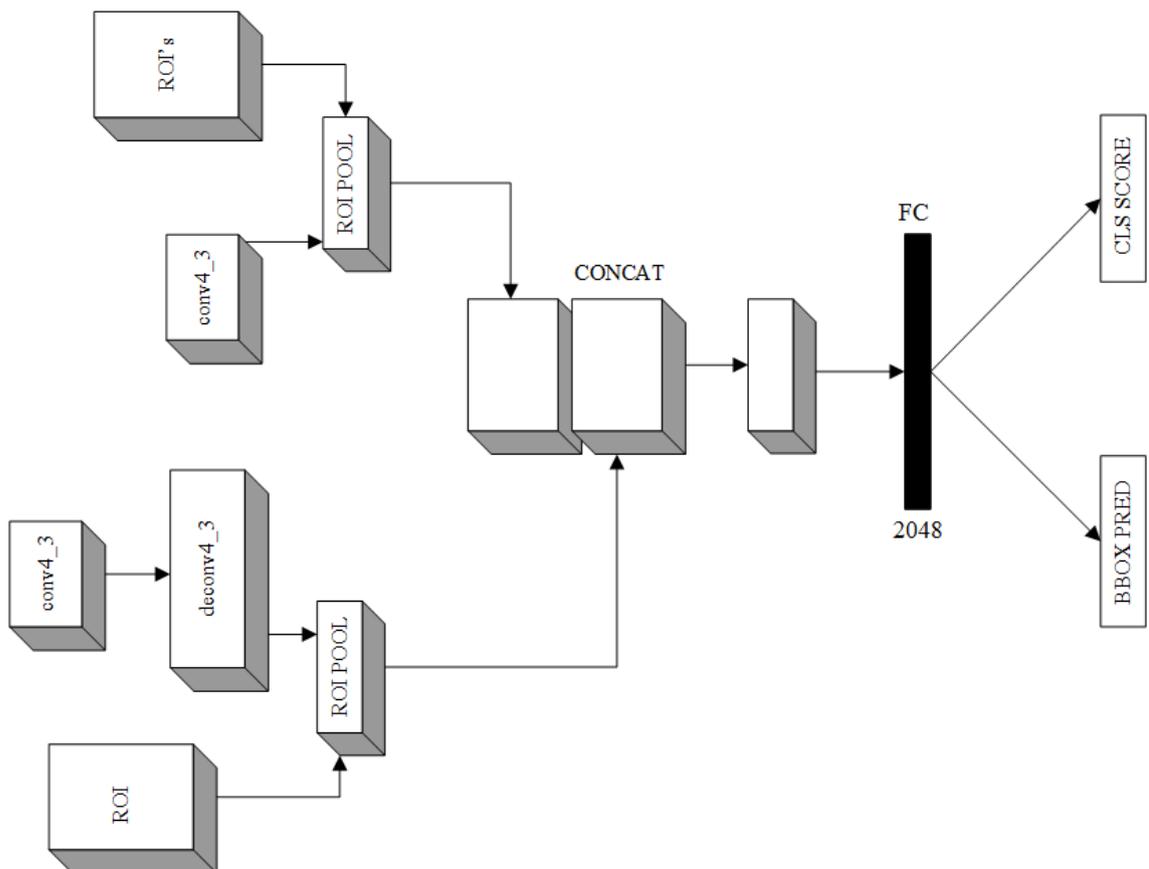
Para la etapa de clasificación se propone la arquitectura que se presenta en la Figura 3. Como en [43] se considera aumentar la resolución de los mapas de características aplicando deconvolución, para proveer mejor información a la capa de agrupamiento ROI. Por lo tanto, la parte de Fast R-CNN recibe como entrada directamente las características extraídas por la capa conv4\_3 de VGG16 [44], su deconvolución y las ROI generadas por RPNCD y RPNLD, en conjunto.



**Figura 1.** Esquema del sistema de detección de peatones en la noche usando Faster R-CNN e imágenes en el infrarrojo lejano.



**Figura 2.** Arquitectura RPN multiscale basada en la red VGG16 [44]. Esta es la subred encargada de la etapa de generación de ROI.



**Figura 3.** Arquitectura de clasificación MS-CNN [30]. Esta subred está destinada para la etapa de clasificación.

### 2.2.3. Detalles técnicos de la implementación

El aprendizaje de la arquitectura propuesta se ha desarrollado a partir de las bases de datos CVC-09 [16] y LSIFIR [17] que se detallan a continuación:

1. La base de datos CVC-09 [16]: Es una de las bases más utilizadas para la detección de peatones en la noche. En este caso se la usó para el entrenamiento y prueba de la propuesta, y posteriormente para su validación. En la Tabla 1 se describen los conjuntos de entrenamiento y de prueba. Esta base de datos viene etiquetada con los peatones presentes en la escena  $B_{gt}$ .

**Tabla 1.** Contenido de la base de datos CVC-09 en la noche

	Positivos	Negativos
Entrenamiento	2200	1002
Prueba	2283	–

Sin embargo, para el caso de largas distancias la base de datos presenta inconsistencias que han sido corregidas. Así, se ha reetiquetado un conjunto de imágenes para corregir estos inconvenientes y depurar los errores de etiquetado.

1. La base de datos LSI *Far Infrared Pedestrian Dataset* (LSIFIR) [16]: Es otra base de datos importante destinada al desarrollo de algoritmos de detección de peatones en la noche. En la Tabla 2 se describen los conjuntos de entrenamiento y de prueba, con sus respectivos tamaños. En este caso al igual que CVC-09 se la usó para el entrenamiento, validación y prueba de la propuesta.

**Tabla 2.** Contenido de la base de datos LSIFIR. El valor entre paréntesis representa el número de fotogramas que contienen peatones

	Clasificación	Detección
Entrenamiento	43391 (10209)	2936 (3225)
Prueba	2205 (5945)	5788 (3279)

Para el aprendizaje de la red, el algoritmo inicialmente reescala la parte más corta de la imagen de entrada a 600 píxeles. En cuanto al entrenamiento de la red, se lo realiza mediante la metodología de entrenamiento conjunto aproximado planteada por Ren *et al.* [12], además, los pesos de cada capa perteneciente a la red son inicializados por medio del modelo pre-entrenado VGG16, para luego ser sintonizado mediante *Minibatch Stochastic Gradient Descent* [45] y el reciente algoritmo de optimización Adam [46] con hiperparámetros detallados en la Tabla 3.

En cuanto a las RPN, estas trabajan de manera independiente. Por lo tanto, su entrenamiento también lo es. Las propuestas generadas por cada una de ellas son combinadas para luego ser etiquetadas mediante el algoritmo de NMS (*Non Maximum Supression*), donde si el índice IoU (*Intersection over union*), dada por Ecuación (1), es mayor que 0,6 es un peatón, si es menor que 0,3 es etiquetado como no peatón, y en caso de no cumplir con ninguna de las dos condiciones, dichas propuestas son excluidas del entrenamiento.

Inmediatamente, en la etapa de clasificación se vuelve a aplicar NMS para reducir redundancias en la detección, aplicando un umbral de 0,6, donde cada detección mayor al umbral se etiqueta como peatón caso contrario no peatón.

$$IoU = \frac{Area(B_{det} \cap B_{gt})}{Area(B_{det} \cup B_{gt})} \quad (1)$$

Donde  $B_{gt}$  es la intersección y  $B_{det}$  la unión, entre el cuadro delimitador real anotado en la base de datos CVC-09 [16] o LSIFIR [17] y el resultado del cuadro delimitador predicho por nuestro modelo.

**Tabla 3.** Parámetros de entrenamiento para el modelo propuesto para la detección de peatones en la noche

Parámetro	Valor
Tasa de aprendizaje	0,001
Momentum	0,9
RMSProp	0,999
Decaimiento de peso	0,0005
Epsilon	$10^{-8}$
Imágenes por batch	2

### 2.2.4. Evaluación experimental

Para llegar al modelo propuesto, se han desarrollado múltiples experimentos, como se puede observar en las Tablas 4 y 5. Donde se analiza la subred de generación de ROI y los efectos que provoca la configuración de las distintas escalas y las relaciones de aspecto de RPNCd y RPNLD.

Para los experimentos se ha hecho uso de los conjuntos de entrenamiento de CVC-09 junto con LSIFIR para la etapa de aprendizaje de la red y los conjuntos de prueba para la evaluación.

Adicionalmente, se analizó la subred de clasificación y los efectos que provoca la deconvolución. En la Tabla 5, los resultados demuestran que al aplicar esta estrategia permite aumentar la resolución de los mapas de características, lo cual provoca un incremento del mAP en un 6 % aproximadamente.

**Tabla 4.** Parámetros de configuración de cajas de referencia RPN para peatones a corta y larga distancia. Resultados de la subred de generación de ROI

Parámetro	RPNCD	RPNLD	mAP (%)	fps
Escala	64, 128, 256	128, 256, 512	86	10
R. Aspecto	1:8, 1:4, 1:2	1:1, 1:2, 2:1		
Escala	64, 128, 256	64, 128, 256	85,6	10
R. Aspecto	1:8, 1:4, 1:2	1:1, 1:2, 2:1		
Escala	32, 64, 128	128, 256, 512	84,5	10
R. Aspecto	1:8, 1:4, 1:2	1:1, 1:2, 2:1		

**Tabla 5.** Resultados obtenidos al aplicar deconvolución a la subred de clasificación

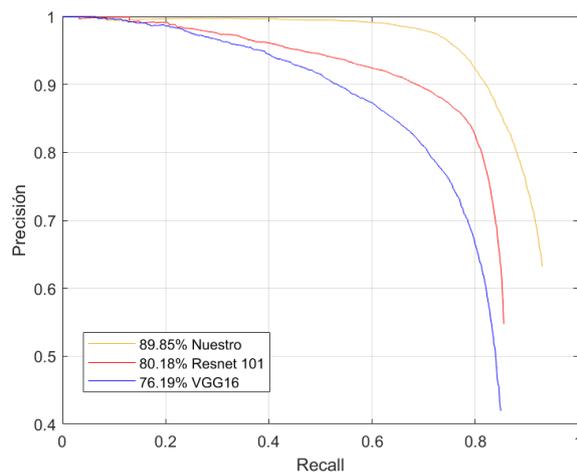
Parámetro	RPNCD	RPNLD	mAP (%)	fps
Escala	64, 128, 256	128, 256, 512	89,9	5
R. Aspecto	1:8, 1:4, 1:2	1:1, 1:2, 2:1		
Escala	64, 128, 256	64, 128, 256	87,8	5
R. Aspecto	1:8, 1:4, 1:2	1:1, 1:2, 2:1		
Escala	32, 64, 128	128, 256, 512	86,8	5
R. Aspecto	1:8, 1:4, 1:2	1:1, 1:2, 2:1		

### 3. Resultados y discusión

En cuanto a la evaluación de la efectividad de la propuesta se utilizaron dos de las bases de datos que representan el punto de referencia, destinadas al desarrollo de sistemas de detección de peatones durante la noche utilizando iluminación infrarroja.

#### 3.1. Protocolo de evaluación

Para evaluar el sistema propuesto se propone la métrica precisión media promedio (mAP que es el *Mean Average Precision*) la cual permite medir la precisión del detector, de manera que se calcula la precisión promedio de cada detección para diferentes valores del índice recall [12].



**Figura 4.** Curva Precisión vs. Recall de los resultados obtenidos para distintas arquitecturas de red Faster R-CNN para la clase peatón, sobre la combinación de los conjuntos de prueba de las bases de datos CVC-09 y LSIFIR.

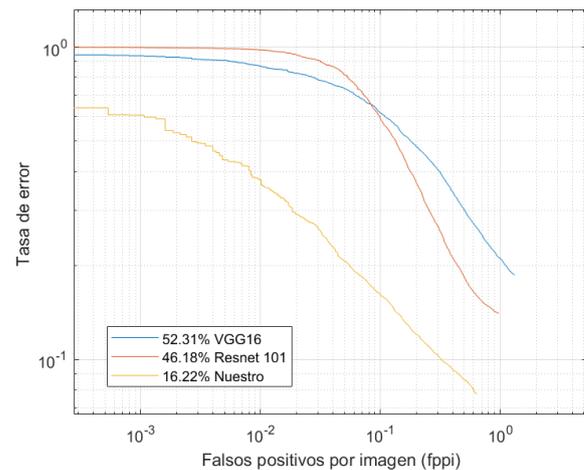
Adicionalmente, se seguirá el protocolo estándar planteado por Dollár *et al.* [47], es decir, se usarán las curvas que relacionan la tasa de error promedio (miss rate) versus los falsos positivos por imagen (FPPI); en el rango de  $10^{-2}$  a  $10^0$  FPPI, que es un indicador de la exactitud especializado en temas vehiculares para la detección de peatones.

#### 3.2. Discusión de los resultados

En la Figura 4 se presentan los experimentos llevados a cabo sobre los conjuntos de prueba de las bases de datos CVC-09 [16] y LSIFIR [17] para distintas arquitecturas de red Faster R-CNN se presentan en la Tabla 6. Los resultados han sido obtenidos bajo las mismas condiciones computacionales, donde se puede observar que esta nueva propuesta alcanza un mAP de 94,6 %, en la etapa de validación, lo que demuestra que el aprendizaje es superior a las otras propuestas. Pero tiene el inconveniente de requerir un mayor esfuerzo computacional.

**Tabla 6.** Resultados de las pruebas y validación de la base de datos CVC-09. Precisión media promedio (mAP) y procesamiento de imágenes por segundo (fps)

	Validación	Prueba	
Modelo	mAP (%)	mAP (%)	fps
VGG16	83,4	76,4	14
Resnet 101	86,1	80,2	8
Nuestro	94,6	89,9	5



**Figura 5.** Curvas de las tasas de error promedio versus FPPI para las distintas arquitecturas de red Faster R-CNN sobre la combinación de los conjuntos de prueba de las bases de datos CVC-09 y LSIFIR.

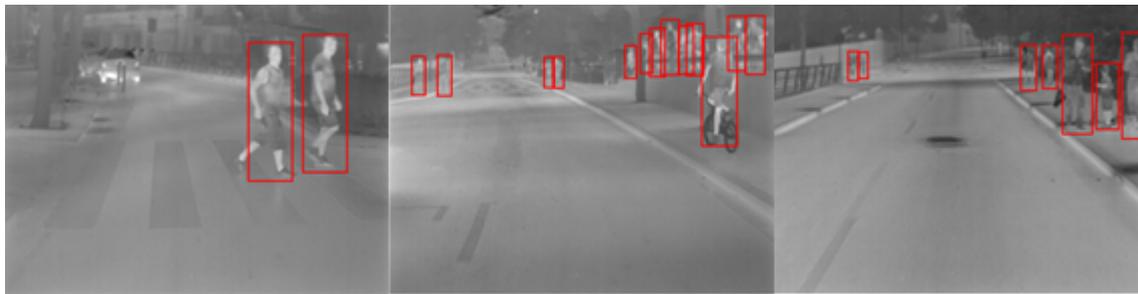
Así se puede observar en la Figura 5 que se ha superado los resultados de los modelos originales de Faster R-CNN y otros modelos presentados por otras investigaciones, como se detalla en la Tabla 7.

**Tabla 7.** Comparación tasas de error promedio de sistemas de detección de peatones en la noche bajo las bases de datos CVC-09 y LSIFIR

Modelo	Tasa de error (%) en $10^{-2}$
Latent-SVM HOPE [1]	25,1
Fuzzy C-means CNN [20]	65,4
Nuestro	16,22

### 3.3. Tiempo de procesamiento

Para la evaluación experimental se usó un equipo compuesto por una GPU con el sistema operativo Linux



**Figura 6.** Ejemplos seleccionados con los resultados obtenidos sobre la combinación de los conjuntos de prueba de las bases de datos LSIFIR y CVC-09, durante la noche.

- Desarrollar una nueva arquitectura DL basada en Faster R-CNN junto con el modelo VGG16 para la detección de peatones en la noche usando imágenes en el infrarrojo lejano. La red RPN de múltiples escalas presentó una mejor detección específicamente para peatones a larga distancia como se puede observar en la Figura 6. En comparación con la arquitectura original de RPN, la arquitectura de RPNCD y RPNLD produjo mejores resultados, la nueva arquitectura incrementó el mAP del 76,4 al 86 %. Adicionalmente, se presentó un aporte significativo al aplicar la deconvolución a la subred de clasificación donde, el mAP incrementó del 86 al 89,9 %. Sin embargo, la deconvolución añadida en la etapa de clasificación incrementa la carga computacional. En consecuencia, la red reduce el procesamiento de 10 fotogramas a 5 fotogramas por segundo.
- Comparar el desempeño de la arquitectura original de Faster R-CNN junto con los modelos VGG16 y Resnet 101, sobre las bases de datos CVC-09 y LSIFIR, obteniéndose resultados superiores en mAP 9,7 % para Resnet 101 y 13,5 % para VGG16. En cuanto a la tasa de error promedio, se obtuvo una diferencia de 29,96 % para Resnet 101 y 36,09 % para VGG16.
- Respecto a la detección, el modelo propuesto demuestra un rendimiento superior respecto a los métodos Olmeda *et al.* [44] y John *et al.* [14], donde la tasa de error promedio es reducida en un 8,88 % respecto a [44] y 49,18 % respecto a [14].
- El tiempo de procesamiento es de 5 fotogramas por segundo, lo que convierte a esta propuesta en un método viable para aplicaciones en tiempo real, destinado a seguridad vehicular.

### 4.2. Recomendaciones y trabajos futuros

Para mejorar el desempeño de este sistema es necesario incluir las siguientes recomendaciones:

- Optimizar el algoritmo propuesto para que trabaje en tiempo real, es decir, sea capaz de procesar al menos 25 fotogramas por segundo.
- Incluir un conjunto de características basadas en múltiples espectros para un mejor rendimiento durante el día y la noche.

16.04, una tarjeta Nvidia Geforce GTX 1080 Ti, con 11 GB GDDR5X 352 bit de memoria. El tiempo de entrenamiento fue de 5 horas aproximadamente. El tiempo promedio de detección es de 170 milisegundos, sobre imágenes de  $640 \times 480$  píxeles; es decir, el sistema procesa 5 imágenes por segundo.

## 4. Conclusiones y recomendaciones

### 4.1. Conclusiones

En este trabajo se ha presentado un método de detección de peatones en la noche usando modernas técnicas de inteligencia artificial, donde se realizaron los siguientes aportes:

## Agradecimientos

Los autores desean expresar sus agradecimientos a los investigadores que han hecho posible las bases de datos de peatones en el infrarrojo, ya que sin esta información habría sido muy difícil desarrollar esta investigación. Además, los autores desean reconocer a los revisores anónimos que contribuyen con su trabajo en la mejora de la redacción de este documento.

## Referencias

- [1] D. König, M. Adam, C. Jarvers, G. Layher, H. Neumann, and M. Teutsch, "Fully convolutional region proposal networks for multispectral person detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, July 2017. DOI: <https://doi.org/10.1109/CVPRW.2017.36>, pp. 243–250.
- [2] D. Olmeda, C. Premebida, U. Nunes, J. M. Armingol, and A. de la Escalera, "Pedestrian detection in far infrared images," *Integrated Computer-Aided Engineering*, vol. 20, no. 4, pp. 347–360, 2013. [Online]. Available: <https://goo.gl/Rss9Qp>
- [3] WHO. (2004) World report on road traffic injury prevention. World Health Organization. [Online]. Available: <https://goo.gl/PBhixd>
- [4] ANT. (2017) Siniestros octubre 2016. Agencia Nacional de Tránsito. Ecuador. [Online]. Available: <https://goo.gl/GoXFX5>
- [5] —. (2016) Siniestros agosto 2017. Agencia Nacional de Tránsito. Ecuador. [Online]. Available: <https://goo.gl/GoXFX5>
- [6] J. Li, X. Liang, S. Shen, T. Xu, and S. Yan, "Scale-aware fast R-CNN for pedestrian detection," *CoRR*, 2015. [Online]. Available: <https://goo.gl/27CMsz>
- [7] J. Yan, X. Zhang, Z. Lei, S. Liao, and S. Z. Li, "Robust multi-resolution pedestrian detection in traffic scenes," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2013. DOI: <https://doi.org/10.1109/CVPR.2013.390>, pp. 3033–3040.
- [8] D. Guan, Y. Cao, J. Liang, Y. Cao, and M. Y. Yang, "Fusion of multispectral data through illumination-aware deep neural networks for pedestrian detection," *CoRR*, 2018. [Online]. Available: <https://goo.gl/AAWJFp>
- [9] J. Liu, S. Zhang, S. Wang, and D. N. Metaxas, "Multispectral deep neural networks for pedestrian detection," *CoRR*, 2016. [Online]. Available: <https://goo.gl/Czc6Jg>
- [10] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing*, vol. 187, pp. 27–48, 2016. DOI: <https://doi.org/10.1016/j.neucom.2015.09.116>, recent Developments on Deep Big Vision.
- [11] L. Deng and D. Yu, "Deep learning: Methods and applications," *Foundations and Trends in Signal Processing*, vol. 7, no. 3–4, pp. 197–387, 2014. DOI: <http://dx.doi.org/10.1561/20000000039>. [Online]. Available: <http://dx.doi.org/10.1561/20000000039>
- [12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems 28*. Curran Associates, Inc., 2015, pp. 91–99. [Online]. Available: <https://goo.gl/5i64rm>
- [13] C. Ertler, H. Posseger, M. Optiz, and H. Bischof, "Pedestrian detection in rgb-d images from an elevated viewpoint," in *22nd Computer Vision Winter Workshop*, 2017. [Online]. Available: <https://goo.gl/L4wB1e>
- [14] C. C. Pham and J. W. Jeon, "Robust object proposals re-ranking for object detection in autonomous driving using convolutional neural networks," *Signal Processing: Image Communication*, vol. 53, pp. 110–122, 2017. DOI: <https://doi.org/10.1016/j.image.2017.02.007>.
- [15] X. Zhang, G. Chen, K. Saruta, and Y. Terata, "Deep convolutional neural networks for all-day pedestrian detection," in *Information Science and Applications 2017*, K. Kim and N. Joukov, Eds. Singapore: Springer Singapore, 2017. DOI: [https://doi.org/10.1007/978-981-10-4154-9\\_21](https://doi.org/10.1007/978-981-10-4154-9_21), pp. 171–178.
- [16] Elektra, *CVC-09: FIR Sequence Pedestrian Dataset*, ElektraAutonomous Vehicle developed by CVC & UAB & UPC, 2016. [Online]. Available: <https://goo.gl/NhYuZ2>
- [17] D. Olmeda, C. Premebida, U. Nunes, J. Armingol, and A. de la Escalera., "Lsi far infrared pedestrian dataset," Universidad Carlos III de Madrid. España, 2013. [Online]. Available: <https://goo.gl/pJTGvj>
- [18] D. Heo, E. Lee, and B. Chul Ko, "Pedestrian detection at night using deep neural networks y saliency maps," *Journal of Imaging Science and Technology*, vol. 61, no. 6, pp. 60 403–1–60 403–9, 2017. DOI: <https://doi.org/10.2352/J.ImagingSci.Technol.2017.61.6.060403>.

- [19] C. Bingwen, W. Wenwei, and Q. Qianqing, “Robust multi-stage approach for the detection of moving target from infrared imagery,” *Optical Engineering*, vol. 51, no. 6, 2012. DOI: <https://doi.org/10.1117/1.OE.51.6.067006>.
- [20] V. John, S. Mita, Z. Liu, and B. Qi, “Pedestrian detection in thermal images using adaptive fuzzy c-means clustering and convolutional neural networks,” in *2015 14th IAPR International Conference on Machine Vision Applications (MVA)*, May 2015. DOI: <https://doi.org/10.1109/MVA.2015.7153177>, pp. 246–249.
- [21] D. Kim and K. Lee, “Segment-based region of interest generation for pedestrian detection in far-infrared images,” *Infrared Physics & Technology*, vol. 61, pp. 120–128, 2013. DOI: <https://doi.org/10.1016/j.infrared.2013.08.001>.
- [22] J. Ge, Y. Luo, and G. Tei, “Real-time pedestrian detection and tracking at nighttime for driver-assistance systems,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 2, pp. 283–298, June 2009. DOI: <https://doi.org/10.1109/TITS.2009.2018961>.
- [23] J. H. Kim, H. G. Hong, and K. R. Park, “Convolutional neural network-based human detection in nighttime images using visible light camera sensors,” *Sensors*, vol. 17, no. 5, pp. 1–26, 2017. DOI: <https://doi.org/10.3390/s17051065>.
- [24] B. Qi, V. John, Z. Liu, and S. Mita, “Pedestrian detection from thermal images with a scattered difference of directional gradients feature descriptor,” in *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, Oct 2014. DOI: <https://doi.org/10.1109/ITSC.2014.6958024>, pp. 2168–2173.
- [25] M. R. Jeong, J. Y. Kwak, J. E. Son, B. Ko, and J. Y. Nam, “Fast pedestrian detection using a night vision system for safety driving,” in *2014 11th International Conference on Computer Graphics, Imaging and Visualization*, Aug 2014. DOI: <https://doi.org/10.1109/CGiV.2014.25>, pp. 69–72.
- [26] J. Kim, J. Baek, and E. Kim, “A novel on-road vehicle detection method using  $\pi$ hog,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 3414–3429, Dec 2015. DOI: <https://doi.org/10.1109/TITS.2015.2465296>.
- [27] K. Piniarski, P. Pawlowski, and A. D. abrowski, “Pedestrian detection by video processing in automotive night vision system,” in *2014 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, Sept 2014, pp. 104–109. [Online]. Available: <https://goo.gl/uxnD6X>
- [28] S. L. Chang, F. T. Yang, W. P. Wu, Y. A. Cho, and S. W. Chen, “Nighttime pedestrian detection using thermal imaging based on hog feature,” in *Proceedings 2011 International Conference on System Science and Engineering*, June 2011. DOI: <https://doi.org/10.1109/ICSSE.2011.5961992>, pp. 694–698.
- [29] H. Sun, C. Wang, and B. Wang, “Night vision pedestrian detection using a forward-looking infrared camera,” in *2011 International Workshop on Multi-Platform/Multi-Sensor Remote Sensing and Mapping*, Jan 2011. DOI: <https://doi.org/10.1109/M2RSM.2011.5697384>, pp. 1–4.
- [30] P. Govardhan and U. C. Pati, “Nir image based pedestrian detection in night vision with cascade classification and validation,” in *2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies*, May 2014. DOI: <https://doi.org/10.1109/ICACCCT.2014.7019339>, pp. 1435–1438.
- [31] Y. Chun-he and D. Cai-Fang, “Research of the method of quickly finding the pedestrian area of interest,” *Journal of Electrical and Electronic Engineering*, vol. 5, no. 5, pp. 180–185, 2017. DOI: <http://doi.org/10.11648/j.jee.20170505.14>.
- [32] J. Baek, J. Kim, and E. Kim, “Fast and efficient pedestrian detection via the cascade implementation of an additive kernel support vector machine,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 4, pp. 902–916, April 2017. DOI: <https://doi.org/10.1109/TITS.2016.2594816>.
- [33] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, “Deep learning for visual understanding: A review,” *Neurocomputing*, vol. 187, pp. 27–48, 2016. DOI: <https://doi.org/10.1016/j.neucom.2015.09.116>.
- [34] H. A. Perlin and H. S. Lopes, “Extracting human attributes using a convolutional neural network approach,” *Pattern Recognition Letters*, vol. 68, pp. 250–259, 2015. DOI: <https://doi.org/10.1016/j.patrec.2015.07.012>.
- [35] P. Sermanet, K. Kavukcuoglu, S. Chintala, and Y. Lecun, “Pedestrian detection with unsupervised multi-stage feature learning,” in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, June 2013. DOI: <https://doi.org/10.1109/CVPR.2013.465>, pp. 3626–3633.

- [36] D. Ribeiro, J. C. Nascimento, A. Bernardino, and G. Carneiro, "Improving the performance of pedestrian detectors using convolutional learning," *Pattern Recognition*, vol. 61, pp. 641–649, 2017. DOI: <https://doi.org/10.1016/j.patcog.2016.05.027>.
- [37] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. Lecun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," 12 2013. [Online]. Available: <https://goo.gl/zNNUCd>
- [38] D. Tomè, F. Monti, L. Baroffio, L. Bondi, M. Tagliasacchi, and S. Tubaro, "Deep convolutional neural networks for pedestrian detection," *Signal Processing: Image Communication*, vol. 47, pp. 482–489, 2016. DOI: <https://doi.org/10.1016/j.image.2016.05.007>.
- [39] J. Cao, Y. Pang, and X. Li, "Learning multi-layer channel features for pedestrian detection," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3210–3220, July 2017. DOI: <https://doi.org/10.1109/TIP.2017.2694224>.
- [40] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, June 2014. DOI: <https://doi.org/10.1109/CVPR.2014.81>, pp. 580–587.
- [41] R. Girshick, "Fast r-cnn," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Dec 2015. DOI: <https://doi.org/10.1109/ICCV.2015.169>, pp. 1440–1448.
- [42] L. Zhang, L. Lin, X. Liang, and K. He, "Is faster r-cnn doing well for pedestrian detection?" in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016. DOI: [https://doi.org/10.1007/978-3-319-46475-6\\_28](https://doi.org/10.1007/978-3-319-46475-6_28), pp. 443–457.
- [43] Z. Cai, Q. Fan, R. Feris, and N. Vasconcelos, "A unified multi-scale deep convolutional neural network for fast object detection," 2016. [Online]. Available: <https://goo.gl/Y4XNZv>
- [44] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations*, 2014. [Online]. Available: <https://goo.gl/98akRT>
- [45] J. Konečný, J. Liu, P. Richtárik, and M. Takáč, "Mini-batch semi-stochastic gradient descent in the proximal setting," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 2, pp. 242–255, March 2016. DOI: <https://doi.org/10.1109/JSTSP.2015.2505682>.
- [46] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," in *ICLR 2015*, 2015. [Online]. Available: <https://goo.gl/so1Da8>
- [47] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 4, pp. 743–761, April 2012. DOI: <https://doi.org/10.1109/TPAMI.2011.155>.