Pedestrian detection at night by using Faster R-CNN y infrared images

Detección de peatones en la noche usando Faster R-CNN e imágenes infrarrojas

Michelle Galarza Bravo[1,*], Marco Flores Calero[2]

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.

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.

[1,*]Electronic, Automation and Control Engineering Major, Universidad de las Fuerzas Armadas ESPE Sanglquí – Ecuador. Autor para correspondencia ✉: magalarza@espe.edu.ec, ⓘ https://orcid.org/0000-0001-8401-1871

[2]Department of Electrics and Electronics, Universidad de las Fuerzas Armadas ESPE, Sangolquí – Ecuador.
ⓘ https://orcid.org/0000-0001-7507-3325

1. Introduction

Pedestrian detection systems (PDS) are one of the most important technological components that have emerged in recent years with the development of mobile robotics applied to the automotive sector and other similar technologies aimed at vehicular safety [1], which need to operate with high quality standards and have a high efficiency and accuracy, because their goal is to protect human life by preventing collisions from happening [2].

Several reports worldwide indicate that traffic accidents generate high material and human costs [3], where pedestrians have a high accident rate, reaching up to 22% [4]. In the case of Ecuador, road accidents represent more than 10% of deaths due to traffic accidents [5]. Therefore, the detection of pedestrians is a subject of active and challenging research due to the complexity of the road scene, which constantly changes due to several factors. For instance, atmospheric conditions contribute to a low visibility and a permanent change of illumination, occlusions generate incomplete information of the human form, distance impairs the quality of the visual information [1, 6, 7]. At night these mishaps are magnified due to dark environments [1, 2, 8, 9].

On the other hand, due to the recent success of Deep Learning techniques [10, 11], the main objective of this work is to implement a method for the detection of pedestrians at night using visual information in the far infrared and the convolutional neural networks, specifically the architectures of the Faster R-CNN type [9, 11–15] to obtain a competitive system that generates cutting-edge results comparable to those in previous works. Therefore, a new Faster R-CNN architecture is presented at multiple scales, which is evaluated under the test sets of the CVC-09 [16] and LSIFIR [17] databases. The results show improvements especially when detecting pedestrians which are at a distance.

The document is organized as follows. The second section presents the methods and materials used, detailing the previous work carried out in the PDS field, especially deep-learning techniques. Additionally, the proposed design of the new Faster R-CNN architecture for the generation of regions of interest, classification and detection of pedestrians during the night is described, followed by the experimental evaluation for different configurations of the proposed model. Subsequently, in the results and discussion section, the values obtained with respect to the detection quality are displayed on the databases destined to the development of PDS at night. Finally, the last section is devoted to conclusions, recommendations and future work that can be done to improve this proposal.

2. Methods and materials

2.1. Previous works

Currently, there are multiple specialized investigations in the detection of pedestrians at night [1, 2, 7–9, 15, 18–30]. To carry out this process, generally, the work is divided into two parts. The first consists in the generation of ROI, and the second in the classification into pedestrians or background. In this way, it is possible to keep the person located while they remain in the scene.

2.1.1. Generation of ROI over images in the far infrared

For the generation of ROI on infrared images there are several methods, the most popular are: sliding windows [18] that exhaustively search over the whole image in several scales, which means the method requires many computational resources and makes it ineffective for real-time applications. To overcome these drawbacks, new proposals have been created, for example, segmentation by movement, proposed by Chen et al. [19] where regions of local interest are identified using PCA and Fuzzy techniques. Kim and Lee [21] have developed a method that combines image segments instead of thresholds and the low frequencies of far infrared images. Ge et al. [22] have proposed an adaptive segmentation method consisting of two thresholds, one specialized for locating bright areas and another for low contrast areas. Chun et al. [31] apply edge detection to obtain a faster ROI generator.

At present, there are more sophisticated methods that use models of convolutional neural networks and their variants for the generation of new proposals [1, 9, 12, 18]. Thus, the detection of heat points in multispectral resolution using IFCNN (Illumination Fully Connected Neural Network) has been proposed by Guan et al. [8] Vijay et al. [20] add a convolutional neuronal network to the work of Chen et al. [19], for classification. Kim et al. [23] have used cameras in the visible spectrum to detect pedestrians at night using CNN. Other alternatives include the Region Proposal Network or RPN, which is initially focused on locating the ROI by means of a combination of exhaustive search and sliding windows, in three orientations and three scales (9 reference boxes) for each sliding window. Each initial proposal is used to train a completely convolutional network to generate the predictions of the bounding box and the probability scores [12].

2.1.2. Classification of pedestrians on images in the far infrared

The methods developed for the classification can be grouped into two categories: the models based on the manual generation of characteristics [24, 25, 32], and

the models of automatic learning of characteristics using deep learning techniques (DL) [8, 11, 33–38].

In the first case, different manual methods of generating characteristics are used together with a classification algorithm, some examples include: HOG + SVM [26,27], HOG + Adaboost [28], HOG + LUV [39], Haar + Adaboost [29], Haar + HOG and SVM [30]. The second category includes convolutional neural networks (CNN) [2, 8, 11, 34, 38], with their different architectures, such as R-CNN [40], Fast R-CNN [41] and Faster R-CNN [12, 15].

The Fast R-CNN architecture [12, 15] essentially decreases the computational load with respect to CNN, and for this reason the detection time of the R-CNN layer [41] decreases. Consequently, Fast R-CNN together with selective search presents a better detection quality. However, both methods require an external ROI generator and have problems when detecting small objects that, in the context of pedestrians, involve long distances [41, 42].

To remedy these drawbacks, Faster R-CNN [12,15] has been added, including a ROI generator based on fully connected RPN layers which share the feature maps generated by the convolutional network with Fast R-CNN [15]. Therefore, very deep networks can be implemented because the total image passes only once through the CNN stage [15].

Therefore, Faster R-CNN is being widely used to construct PDS [1, 9, 42]. For example, in [1] Faster R-CNN has been used for pedestrian detection in multiple spectra, initially Faster R-CNN has been trained with only color and infrared images, Faster RCNN-C and Faster RCNN-T respectively, using a new model of neural network for training. Subsequently, features have been combined in different stages, creating Early Fusion, Halfway Fusion, Late Fusion and Score Fusion models. Additionally, Wang et al. [9], with reference to Liang et al. [41], combine RPN + BDT to build a pedestrian detection system in multiple spectra. However, it is considered that Faster RCNN does not work very well for the detection of pedestrians, because the feature maps do not present enough information for long-distance pedestrians. For this reason, Feris et al. [43] have proposed a subnetwork for the generation of ROI in multiple scales together with a subnet for

the classification based on Fast R-CNN.

2.2. Pedestrian detection system at night

Figure 1 shows the proposed scheme for the development of the PDS at night, using images taken with infrared illumination and as a Faster R-CNN base architecture together with the VGG16 model [44] where some detailed changes have been developed below.

2.2.1. Generation of ROI over images in the far infrared

Because the original architecture of Faster RCNN [12,15] presents detection problems in the case of pedestrians that are in the distance, the architecture developed in Feris et al is taken into account. [43] Therefore, it has been decided to place two independent region proposal networks (RPN), which have different characteristics, as detailed in Table 2. In both cases, with an approach directed to pedestrians at short distance (RPNCD) and long distance (RPNLD). As shown in Figure 2, RPNLD is powered by the characteristics that are provided by the conv4_3 layer of VGG16 [44], because the grouping networks can discriminate pedestrians that are in the distance, where the more abundant feature maps are beneficial for detecting pedestrians over long distances [6]. Regarding RPNCD, like the original architecture of Faster R-CNN [12], it is fed by the characteristics delivered by the conv5_3 layer, since it extracts the most representative characteristics present in the image. For this reason, it provides excellent results for pedestrians at short distance.

2.2.2. Classification of ROI over images in the far infrared

For the classification stage, the architecture presented in Figure 3 is proposed. As in [43], the option of increasing the resolution of feature maps by applying deconvolution is considered in order to provide better information to the ROI grouping layer. Therefore, the Fast R-CNN part receives the characteristics extracted by the conv4_3 layer of VGG16 [44] as direct input, its deconvolution and the ROI generated by RPNCD and RPNLD as a whole.
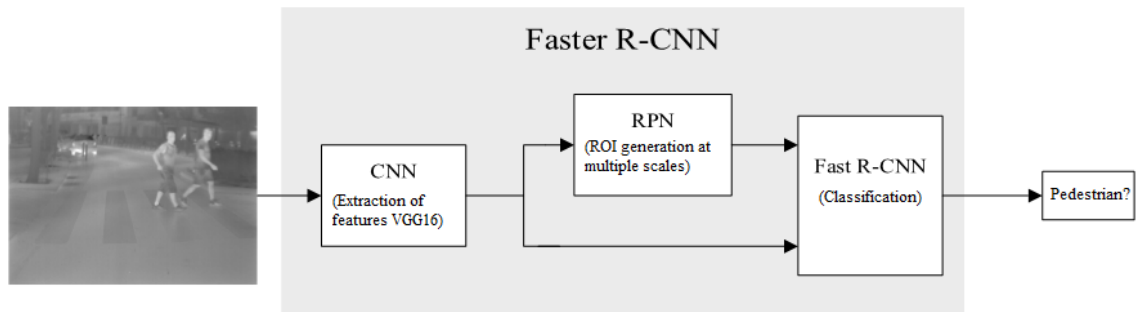
Figure 1. Schematic of the pedestrian detection system at night using Faster R-CNN and images in the far infrared.
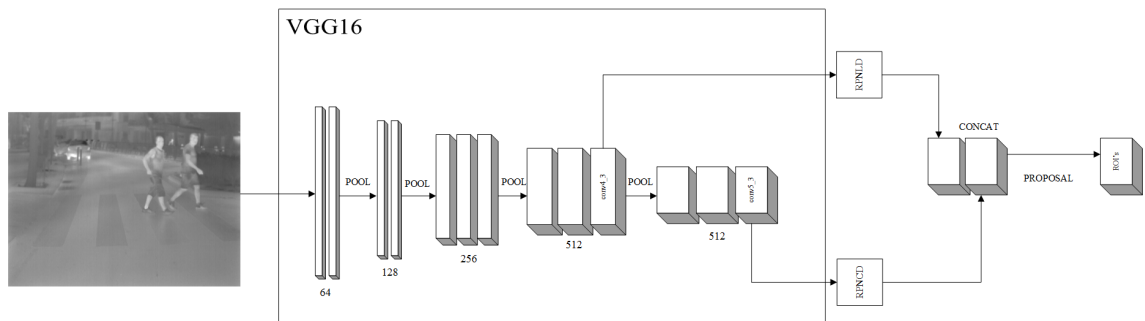

Figure 2. Multiscale RPN architecture based on the VGG16 network [2]. This is the subnetwork responsible for the ROI generation stage.
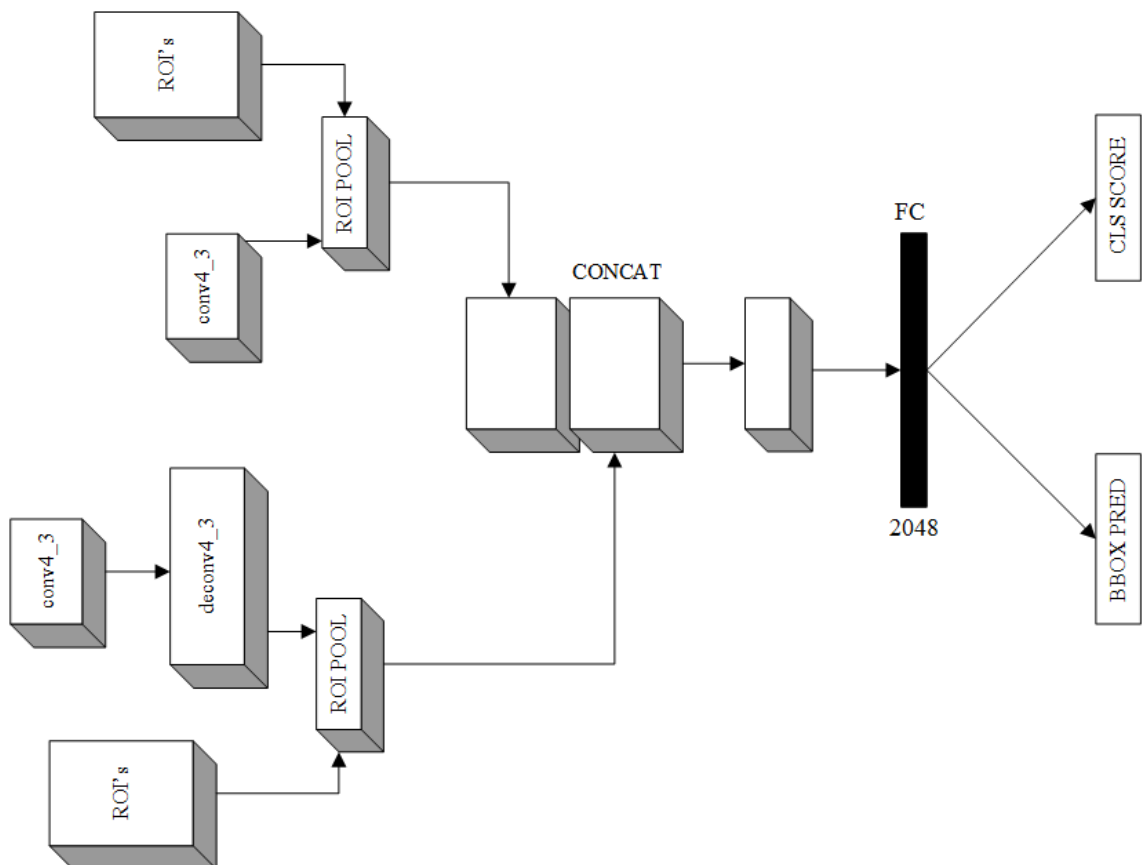

Figure 3. MS-CNN classification architecture [41]. This subnet is intended for the classification stage.

2.2.3. Technical details of the implementation

The learning of the proposed architecture has been developed from the CVC-09 [16] and LSIFIR [17] databases as detailed below:

1. CVC-09 database [16]: It is one of the most used bases for the detection of pedestrians at night. In this case, it was used for the training and testing of the proposal, and later for its validation. Table 1 describes the training and test sets. This database is tagged with pedestrians present in the scene B_{gt}.

Table 1. Content of the CVC-09 database at night

	Positives	Negatives
Training	2200	1002
Test	2283	–

However, in the case of long distances the database presents inconsistencies that have been corrected. Thus, a set of images has been re-labeled to correct these drawbacks and to debug labeling errors.

1. The LSI Far Infrared Pedestrian Dataset database (LSIFIR) [16]: It is another important database for the development of algorithms for pedestrian detection at night. Table 2 describes the training and test sets, with their respective sizes. In this case, like CVC-09, it was used for the training, validation and testing of the proposal.

Table 2. Content of the LSIFIR database. The value in parentheses represents the number of frames that contain pedestrians

	Clasification	Detection
Training	43391 (10209)	2936 (3225)
Test	2205 (5945)	5788 (3279)

In order to train the network, the algorithm initially re-scales the shortest part of the input image to 600 pixels. Regarding the training of the network, this is done through the approximate joint training methodology proposed by Ren et al. [12]. In addition, the weights of each layer belonging to the network are initialized by means of the pre-trained model VGG16, and then fine-tuned by means of the Minchart Stochastic Gradient Descent [45] and the recent Adam optimization algorithm [46] with hyperparameters detailed in Table 3.

As for the RPN, they work independently. Therefore, their training is also independent. The proposals generated by each of them are combined and then labeled using the NMS (Non Maximum Supression) algorithm, where if the IoU (Intersection over Union) index, given by Equation (1), is greater than 0.6, it is a pedestrian, if it is less than 0.3, is labeled as a non-pedestrian, and in case of not fulfilling any of the two conditions, said proposals are excluded from the training.

Immediately after, in the classification stage, NMS is again applied to reduce detection redundancies, applying a threshold of 0.6, where each detection greater than the threshold is labeled as a pedestrian, otherwise a non-pedestrian.

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

Where B_{gt} is the intersection and B_{det} the union, between the actual bounding box annotated in the database CVC-09 [16] or LSIFIR [17] and the result of the bounding box predicted by our model.

Table 3. Training parameters for the proposed model for pedestrian detection at night

Parameter	Value
Learning rate	0,001
Momentum	0,9
RMSProp	0,999
Weight loss	0,0005
Epsilon	10^{-8}
Images by batch	2

2.2.4. Experimental evaluation

To arrive at the proposed model, multiple experiments have been developed, as can be seen in Tables 4 and 5, where the ROI generation subnet and the effects caused by the configuration of the different scales and aspect ratios of RPNCD and RPNLD are analyzed.

For the experiments, the CVC-09 training sets have been used together with LSIFIR for the learning stage of the network and the test sets for the evaluation.

Additionally, the classification subnetwork and the effects caused by deconvolution were analyzed. In Table 5, the results show that applying this strategy allows for an increase in the resolution of the characteristic maps, which causes an increase in the MPA of approximately 6%.

Table 4. Configuration parameters of RPN reference boxes for pedestrians at short and long distance. Results of the ROI generation subnet

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

Table 5. Results obtained by applying deconvolution to the classification subnet

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

3. Results and discussion

Regarding the evaluation of the effectiveness of the proposal, two of the databases representing the reference point were used, aimed at the development of pedestrian detection systems at night using infrared illumination.

3.1. Evaluation protocol

To evaluate the proposed system, the Mean Average Precision (mAP) metrics is proposed, which allows for the measurement of the accuracy of the detector, so that the average accuracy of each detection is calculated for different values of the recall index [12].
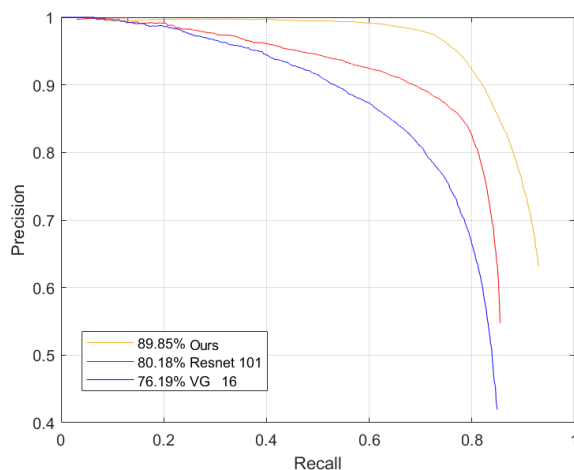
Figure 4. Curve, Precision vs. Recall of the results obtained for different Faster R-CNN network architectures for the pedestrian class, on the combination of the test sets of the CVC-09 and LSIFIR databases.

Additionally, the standard protocol proposed by Dollár et al. [47], that is, the curves that relate the average error rate (miss rate) versus false positives per image (FPPI) will be used in the range of 10^{-2} to 10^{0} FPPI, which is an indicator of specialized accuracy in vehicular topics for pedestrian detection.

3.2. Discussion of results

In Figure 4 are presented the experiments carried out on the test sets of the CVC-09 [16] and LSIFIR [17] databases for different Faster R-CNN network architectures are presented in Table 6. The results have been obtained under the same computational conditions, where it can be observed that this new proposal reaches an MPA of 94.6% in the validation stage, which shows that the learning is superior to that of other proposals. However, it has the disadvantage of requiring a greater computational effort.

Table 6. Results of the tests and validation of the CVC-09 database. Mean average precision (mAP) and image processing per second (fps)

	Validation	Test	
Model	mAP (%)	mAP (%)	fps
VGG16	83,4	76,4	14
Resnet 101	86,1	80,2	8
Ours	94,6	89,9	5
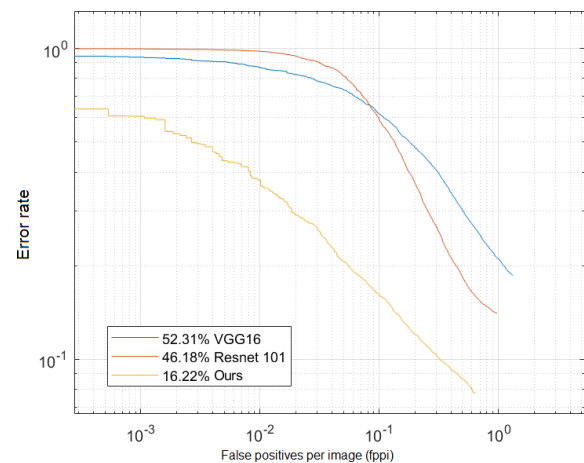
Figure 5. Curves of the average error rates versus FPPI for the different Faster R-CNN network architectures on the combination of the test sets of the CVC-09 and LSIFIR databases.

Thus, it can be seen in Figure 5 that the results of the original models of Faster R-CNN and other models presented by other investigations have been surpassed, as detailed in Table 7.

Table 7. Comparison of average error rates of pedestrian detection systems at night under the CVC-09 and LSIFIR databases

Model	Error rate (%) en 10^{-2}
Latent-SVM HOPE [1]	25,1
Fuzzy C-means CNN [20]	65,4
Ours	16,22

3.3. Processing time

For the experimental evaluation, a computer composed of a GPU with the operating system Linux 16.04, an Nvidia Geforce GTX 1080 Ti card, with 11 GB GDDR5X 352 memory was used. The training time was approximately 5 hours. The average detection time is 170 milliseconds, on images of 640×480 pixels; that is, the system processes 5 images per second.

Figure 6. Examples selected with the results obtained on the combination of the test sets of the LSIFIR and CVC-09 databases, during the night.

- Development of a new DL architecture based on Faster R-CNN together with the VGG16 model for the detection of pedestrians at night using images in the far infrared. The multi-scale RPN network presented better detection specifically for long-distance pedestrians, as shown in Figure 6. Compared to the original RPN architecture, the of RPNCD and RPNLD architecture produced better results. The new architecture increased the mAP from 76.4 to 86%. Additionally, a significant contribution was presented when applying deconvolution to the classification subnet, with the mAP increasing from 86 to 89.9%. However, the deconvolution added in the classification stage increases the computational load. As a result, the network processing is reduced from 10 frames to 5 frames per second.

- Comparison of the performance of the original Faster R-CNN architecture together with the VGG16 and Resnet 101 models, on the CVC-09 and LSIFIR databases, obtaining better results in mAP 9.7% for Resnet 101 and 13.5% for VGG16 Regarding the average error rate, a difference of 29.96% was obtained for Resnet 101 and 36.09% for VGG16.

- Regarding detection, the proposed model demonstrates superior performance with respect to Olmeda et al. [44] and John et al. [14], where the average error rate is reduced by 8.88% with respect to [44] and 49.18% with respect to [14].

- The processing time is 5 frames per second, which makes this proposal a viable method for real-time applications, aimed at vehicular safety.

4. Conclusions and recommendations

4.1. Conclusions

This work presented a method of detecting pedestrians at night using modern artificial intelligence techniques. The following contributions were made:

4.2. Recommendations and future work

To improve the performance of this system it is necessary to include the following recommendations:

- Optimize the proposed algorithm to work in real time, that is, so it is able to process at least 25 frames per second.

- Include a set of features based on multiple spectra for better performance during the day and night.

Acknowledgments

it would have been very difficult to develop this research. In addition, the authors wish to acknowledge the anonymous reviewers who contribute their work for the improvement of this document.

References

[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) Siniestyros 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/10.1561/2000000039. [Online]. Available: http://dx.doi.org/10.1561/2000000039

[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, 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 πhog," *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.jeee.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 multilayer 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, 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.

[48] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. DOI: https://doi.org/10.1109/CVPR.2016.90, pp. 770–778.