Faster R-CNN for pedestrian detection: a review

Vidushi Singh*, Gaurav Prasad*, Smit Patel*
Department of Electrical and Computer Engineering
University of Waterloo
Email: {vsvidush, g3prasad, sd28pate} @uwaterloo.ca

***Abstract*—Human detection is a basic job in computer vision with various applications in security, surveillance, and autonomous systems. A comprehensive analysis of recent advancements in R-CNN (Region-based Convolutional Neural Network)-based human identification is provided by this review paper. This study examines the theoretical foundations, methods, metrics for evaluating the performance, issues, and potential applications of R-CNN for human identification through a comprehensive review of literature from reputable sources, including academic archives, IEEE Xplore, and Google Scholar. Giving scholars and practitioners a comprehensive review of the most modern methods and techniques in this subject is the study's main objective.**

***Index Terms*—Object detection, computer vision, R-CNN, human detection, deep learning.**

I. Introduction

Pedestrian detection is crucial for applications like computer vision, including autonomous driving, surveillance, and human-computer interaction. Identifying pedestrians is vital for maintaining safety and allows the intelligent systems to comprehend and engage with their surroundings efficiently. Recent developments in Deep Learning techniques like Faster Region-based Convolutional Neural Networks (R-CNN) have demonstrated pedestrian recognition assignments. This section introduces Faster R-CNN, summarizes the significance of pedestrian detection, and describes the paper's organizational framework.

Pedestrian detection systems are crucial components of road safety, this role is even more essential in low-light situations. Sophisticated technologies are required to deal with the complexities of nocturnal environments, and recent research has focused on how deep learning techniques and infrared imaging collaborate to improve detection accuracy [1] [2] [3]. The following literature review will examine how an algorithm named Faster R-CNN can deal with a far-infrared image to recognise pedestrians [1].

Improved pedestrian detection is an important aspect in the automobile industry as the fast changing industry is pursuing the development of better driver assistance systems and ultimately aiming at driver-less cars [1] [4]. The objectives are to increase the safety levels of roads and hence reducing accidents and collision rates, especially at night [1-5].

Nighttime situations present particular difficulties, including decreased visibility and varying illumination [1] [5]. Due to complex elements, innovative methods are needed to improve detection system reliability and resilience [1] [3].

Deep learning techniques like Convolutional Neural Networks (CNNs) have now become a dominant feature in pedestrian detection [1] [2] [6]. Which has led to the improvement of detection accuracy through overcoming the limitations of classical handcrafted techniques by deep learning models with ability to feature extraction from data automatically [2] [6].

In order to overcome the inherent limits of optical imagery under dark conditions, scientists have studied infrared imaging [1] [3]. At night, when detecting pedestrians it is extremely useful that Infrared sensors capture thermal signatures and provide enhanced visibility as well as other relevant things [1] [3].

As we explore the intricacies of pedestrian detection methodologies, the emphasis remains fixed on the Faster R-CNN architecture and relationship with infrared imaging in an effort to reveal important information regarding changing trends in nighttime pedestrian detection.

Fig. 1. Wireless Pedestrian Detection.

II. Literature Review

The analysis of the literature offers a comprehensive overview of the latest developments in pedestrian detection, highlighting the use of novel approaches, deep learning techniques, and infrared imaging to handle problems in a variety of settings. Every publication offers insightful observations that highlight the ever-changing field of study in this important area. Future research directions are indicated for factors such as dataset specificity, computational complexity, and speed vs accuracy trade-offs.

The paper by Liu and Stathaki (2018) titled "Faster R-CNN for Robust Pedestrian Detection Using Semantic Segmentation Network" presents an innovative approach to pedestrian detection by integrating semantic segmentation with the Faster R-CNN algorithm. This integration aims to enhance detection accuracy by reducing false positives caused by hard-value negative samples like tree leaves, traffic lights, and poles [1]. The method boosts the Faster R-CNN architecture by measuring a new sectional approach semantic segmentation architecture, which computes complementary higher-level semantic properties. Multi-resolution maps are different from network layers as they detect pedestrians accurately for unique measurements. Boosted forest deployed for learning the structured features in a cascaded manner to handle hard negatives [1]. Experimental results on the Caltech data pool demonstrate an improved identification accuracy with the help of a semantic network, particularly for small pedestrians in the height range of [50, 80) pixels in height [1]. The semantic network model learned to utilize Caffe-SegNet and initialize VGG-16 architecture learned on ImageNet [1].

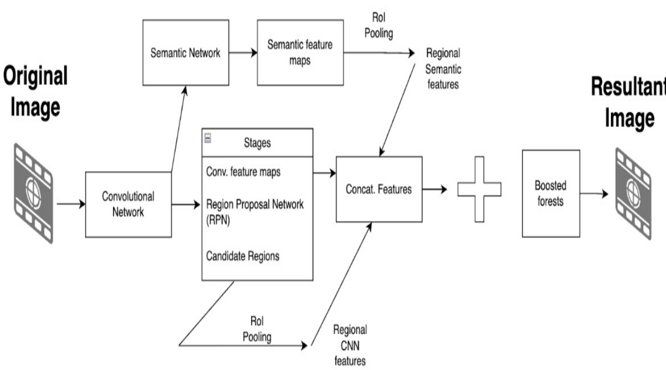

Fig. 2. Semantic segmentation with Faster R-CNN network [1].

Three enhancements to the Faster R-CNN algorithm for pedestrian detection are suggested in this study by Jianwei Ma titled "An Improved Faster R-CNN Algorithm for Pedestrian Detection": Initially, MobileNetv2 is used as the backbone network rather than the conventional VGG16 backbone. Secondly, a more effectively combine shallow and deep feature information, a multi-branch featured pyramid network architecture (M-FPN) was designed. Thirdly, a Region Proposal Network (SE-RPN) is being implemented to minimize background interference characteristics and concentrate on pedestrian attributes. In comparison to the original approach of Faster R-CNN method, the study produces higher results; for the self-built dataset with average accuracy being improved by 6.14%, and the detection speed rose to 27 frames per second. With a detection speed of 39.4 frames per second, the Average Precision (AP) of Caltech dataset was approximately 87% [2].

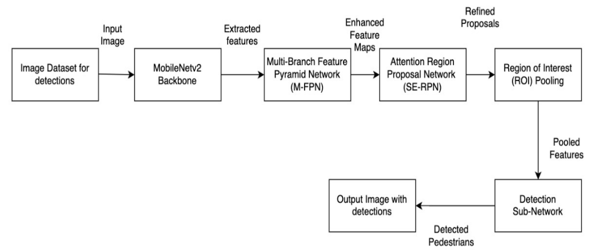

Fig. 3. Improved R-CNN with MobileNetv2 architecture as backbone [2].

The paper by Masita, K.L., Hasan, A.N., and Paul S. focuses on pedestrian detection utilizing an R-CNN object detector with deep learning techniques. The authors have covered a range of technologies for pedestrian detection along with deep learning models such as CNNs and R-CNN, Color Descriptors, Support Vector Machines (SVMs), and feature extraction methods like HOG and the Scale Invariant Feature Transform (SIFT) architecture. Deep learning techniques are performed by utilizing the AlexNet model for feature extraction and transfer learning modifies the weights of Convolutional Neural Networks for their classification task on the datasets. The performance on two different datasets: the Penn-Fudan Pedestrian dataset and the KTH Multiview Football dataset with metrics are measured using Average Precision (AP) and Average Miss Rate (MR). The results show that the large-scale datasets improve the performance of the deep learning model for pedestrian detection [3].

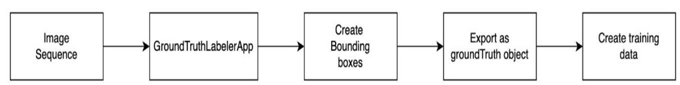

Fig. 4. Detection using Deep Learning (Color descriptors, SVMs etc [3].

The literature study titled "Pedestrian Detection at Night using Faster R-CNN and Infrared Images" by Galarza Bravo and Flores Calero explores the vital field of traffic safety and highlights the critical function of pedestrian detection systems, especially in difficult nighttime circumstances. The study emphasizes the particular difficulties brought on by decreased visibility and dynamic illumination changes at night, while acknowledging the need of precise and effective detection for averting accidents and lowering mortality. Recent studies have concentrated on the integration of cutting-edge technology, particularly deep learning and infrared imaging, to address these issues. Positive trends include the use of the Faster R-CNN algorithm, which effectively detects pedestrians by utilising its Region Proposal Network (RPN) architecture [4].

As seen in the work under evaluation, the introduction of multi-scale techniques shows potential by tackling particular detection issues and improving system robustness. But issues like higher processing costs and dataset restrictions—such as

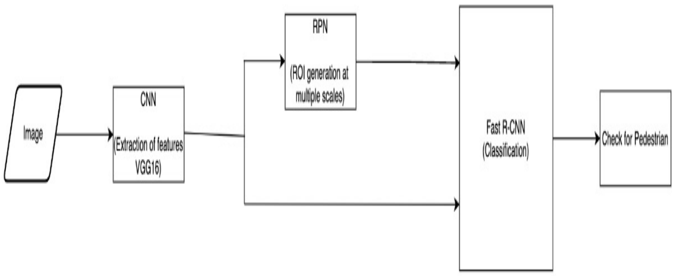

Fig. 5. RPN architecture [4].

a dependency on certain datasets like CVC-09 and LSIFIR highlight the need for more study to provide reliable and useful solutions for nighttime pedestrian identification. In summary, the literature study sheds light on how nighttime pedestrian detection is developing and lays the groundwork for further developments in this crucial area [4].

With an emphasis on Fourier domain adaption, Peggy Joy Lu and Jen-Hui Chuang's study tackles the vital subject of nocturnal pedestrian identification using Faster R-CNN. The problem of insufficient infrared data is addressed by using unsupervised domain adaptation, which is crucial in low-light situations when infrared sensors are necessary. Using Fourier domain adaptation to match the distribution shift between labeled source data (COCO dataset) and unlabeled target data (MI3 dataset), a suggested method improves the item recognition without requiring intricate changes. The approach was compared with the studies of Fourier Domain Adaptation (FDA), an original Domain Adaptation based on Faster R-CNNs (DAF), and the baseline Faster R-CNN. Notable enhancements, varying from 9% to 13%, are noted in several infrared picture channels and dim light situations, proving the effectiveness of the suggested method in resolving data inadequacy issues in surveillance scenarios during the night [5].

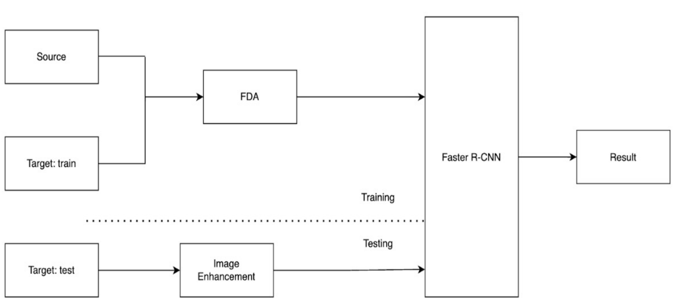

Fig. 6. Fourier domain Adaption [5].

This research proposed novel ways by combining five deep saliency networks with the Faster R-CNN detector to address the difficulty of improving pedestrian recognition under changing illumination conditions. Pixel-level annotations for saliency networks in both color and temperature channels, as well as dataset enrichment with enhanced annotations, are noteworthy contributions. The paper does an in depth analysis of saliency networks, using many metrics, and suggests a unique way to combine color pictures with saliency maps for enhanced detection. The newly installed detectors work better than expected and offer insightful information that will help improve pedestrian identification in a variety of illumination scenarios [6].

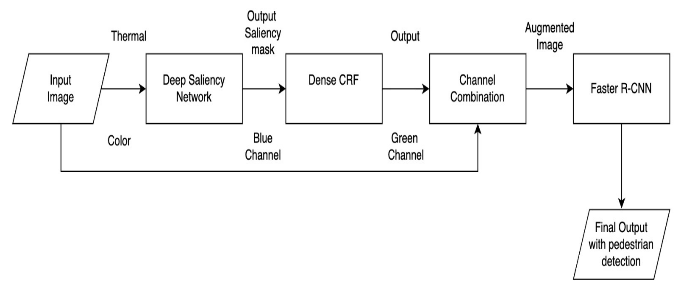

Fig. 7. Saliency networks for illumination condition [6].

This research delves deeply into the benefits of infrared (IR) imaging and tackles the crucial subject of pedestrian identification under difficult lighting conditions. The paper suggests implementing two types of modifications (Type 2 and Type 3) to improve accuracy and speed respectively by utilizing Convolutional Neural Networks (CNNs), notably the highly effective Fast R-CNN architecture. Fast R-CNN Type 2 demonstrates enhanced detection rates, while Type 3 achieves quicker processing, making it more relevant for live applications. The paper effectively illustrates the applicability of these adjustments in low light situations by applying them to infrared photographs. However, there are issues with overfitting to specific datasets, model complexity, and balancing speed versus accuracy that were acknowledged. Difficult lighting can obscure pedestrian features, but the proposed changes help R-CNNs better identify pedestrians even in such challenging situations. While accuracy improves with Type 2, the decreased processing time of Type 3 indicates potential for systems that can respond rapidly [7].

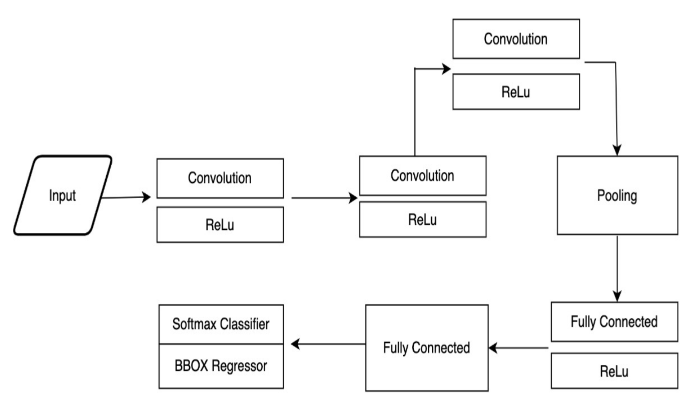

Fig. 8. Modified Faster R-CNN for more accuracy [7].

III. Observation

Overall, all the evaluated publications had many approaches with their own advantages and disadvantages in advanced pedestrian detection. Semantic segmentation included by Liu and Stathaki into Faster R-CNN demonstrated a decrease in miss rates and achieved competitive results with better accuracy [1]. In order to achieve better accuracy and a fair trade-off between speed and precision in pedestrian recognition, Zhao et al. upgraded Faster R-CNN with additions such as MobileNetv2 while using an emphasis on transfer learning for increased flexibility, Masita et al. investigated the R-CNN approaches using the AlexNet model. Specific measures on the model showed an average precision of 52% and 84% on the Penn-Fudan and KTH Multiview Football datasets, respectively [3].

By utilising Faster R-CNN and infrared pictures, Galarza Bravo et al. effectively tackle low-light issues and aid in the detection of pedestrians at night [4]. In order to detect pedestrians at night, Lu and Chuang suggested the adaptation of the Fourier domain. Their results demonstrate notable accuracy gains of between 9% and 13%. Das and colleagues include saliency networks and Faster R-CNN, offering significant perspectives on picture enhancement for pedestrian identification; nonetheless, they do not furnish precise percentages. Although precise percentages are not given, Ullah et al. adapt Fast R-CNN to increase accuracy and speed in infrared pictures, leading to an enhanced detection rate.

IV. Strengths and Setbacks

All the evaluated publications showcased both obstacles and progress in the field of pedestrian detection. Liu and Stathaki's incorporation of semantic segmentation into Faster R-CNN demonstrates superior contextual awareness and competitive accuracy but at the expense of increased computational complexity that slows down inference [1]. A fair trade-off between accuracy and speed is achieved by Zhao et al.'s enhanced Faster R-CNN method, which incorporates MobileNetv2 and M-FPN. However, there may be drawbacks due to the technique's increased processing resource requirements. Although Masita et al. recognize that dataset size affects R-CNN performance, their investigation of R-CNN using the AlexNet model emphasizes the power of transfer learning for adaptation [3].

However, Galarza Bravo et al.'s method of detecting pedestrians at night using infrared pictures and Faster R-CNN shows the promise of deep learning in low light. However, the lack of performance measures prevents a comprehensive evaluation [4]. Although Lu and Chuang's Fourier domain adaptation for night detection was found to be fundamentally useful, it lacks accuracy or percentage accuracy. Das et al.'s incorporation of saliency networks into the pedestrian recognition process provides useful information, but quantitative evaluation is limited by the lack of real-time data [6].

The investigation of Fast R-CNN modifications by Ullah et al. for pedestrian recognition in infrared pictures shows encouraging improvements in speed and accuracy. The paper's wider usefulness is impacted by the understudied generalizability of these adjustments to a variety of datasets [7]. Notwithstanding these drawbacks, the studies together add to the developing field of pedestrian detection by highlighting the necessity of more research and the harmonization of assessment criteria in this area.

V. Conclusion

To sum up, all the evaluated publications include a variety of approaches, strategies, and modifications that together expand the area of pedestrian detection. Every submission offers insightful information that highlights the advantages and tackles particular difficulties in pedestrian detection settings. The incorporation of semantic segmentation, enhancements to Faster R-CNN algorithms, and investigation of diverse detection methods showcase the adaptability of contemporary methodologies.

Nevertheless, there are significant drawbacks, such as the need for processing resources, restrictions on the generalizability to different datasets, and in certain situations, the lack of particular quantitative findings. It becomes clear that uniform assessment measures and benchmarks are required to enable fair comparisons across various approaches.

In spite of these obstacles, all the publications taken as a whole add to our knowledge of pedestrian identification in a variety of difficult situations, including dimly lit environments and a range of settings. The goals of future study should be to investigate the applicability of suggested approaches on a wider variety of datasets, overcome the noted setbacks, and encourage transparency in reporting performance indicators. The continuous endeavor to create reliable and flexible pedestrian detection systems has advanced significantly as a result of these efforts.

References

[1] Liu, T. and Stathaki, T., 2018. Faster R-CNN for robust pedestrian detection using semantic segmentation network. Frontiers in neurorobotics, 12, p.64.

[2] Zhao, Z., Ma, J., Ma, C. and Wang, Y., 2021, November. An improved faster R-CNN algorithm for pedestrian detection. In 2021 11th international conference on information technology in medicine and education (ITME) (pp. 76-80). IEEE.

[3] Masita, K.L., Hasan, A.N. and Paul, S., 2018, November. Pedestrian detection using R-CNN object detector. In 2018 IEEE Latin American Conference on Computational Intelligence (LA-CCI) (pp. 1-6). IEEE.

[4] Bravo, M.G. and Calero, M.F., 2018. PEDESTRIAN DETECTION AT NIGHT BY USING FASTER R-CNN Y INFRARED IMAGES. Ingenius, Revista Ciencia y Tecnología, (20).

[5] Lu, P.J. and Chuang, J.H., 2022, May. Fourier domain adaptation for nighttime pedestrian detection using Faster R-CNN. In 2022 IEEE International Symposium on Circuits and Systems (ISCAS) (pp. 3155-3159). IEEE.

[6] Das, A.J., Choudhury, S., Saikia, N. and Rajbongshi, S., 2023. Pedestrian detection in thermal and color images using a new combination of saliency network and Faster R-CNN. International Journal of Computing and Digital Systems, 13(1), pp.1-1.

[7] Ullah, A., Xie, H., Farooq, M.O. and Sun, Z., 2018, November. Pedestrian detection in infrared images using fast RCNN. In 2018 Eighth International Conference on Image Processing Theory, Tools and Applications (IPTA) (pp. 1-6). IEEE.