IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 31 January 2025, accepted 27 February 2025, date of publication 6 March 2025, date of current version 18 March 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3548837

RESEARCH ARTICLE

Enhanced Nighttime Vehicle Detection for On-Board Processing

LEYRE ENCÍO[1], DANIEL FUERTES[1], CARLOS R. DEL-BLANCO[1], IU AGUILAR[2], CRISTINA PÉREZ-BENITO[2], ALEKSANDAR JEVTIĆ[2], FERNANDO JAUREGUIZAR[1], AND NARCISO GARCÍA[1], (Life Senior Member, IEEE)

[1]Grupo de Tratamiento de Imágenes (GTI), Information Processing and Telecommunications Center, ETSI Telecomunicación, Universidad Politécnica de Madrid, 28040 Madrid, Spain
[2]Ficosa Automotive SLU, 08232 Viladecavalls, Spain

Corresponding author: Leyre Encío (leyre.encio@upm.es)

This work was supported in part by the project Percepción Inteligente para los Vehículos Autónomos y Conectados (InPercept) funded by Centro para el Desarrollo Tecnológico Industrial (CDTI), Spanish Ministerio de Ciencia e Innovación, under Grant PTAS-20211011; in part by European Union, NextGenerationEU, Plan de Recuperación, Transformación y Resiliencia; and in part by Ministerio de Ciencia, Investigación y Universidades (MCIU)/Agencia Estatal de Investigación (AEI)/10.13039/501100011033 of the Spanish Government under Project PID2020-115132RB (SARAOS) and Project PID2023-148922OA-I00 (EEVOCATIONS).

ABSTRACT Nighttime vehicle detection poses significant challenges, particularly in scenarios with limited lighting, where visibility is often compromised. To address this problem, this paper proposes a novel nighttime vehicle detection system that dynamically adapts to extreme lighting conditions, ranging from bright daytime scenarios to challenging nighttime conditions where the vehicle's appearance may be entirely lost. For this purpose, a multi-granularity detection approach is adopted, automatically combining bounding-box and point-based representations depending on the vehicle's visibility. Bounding-box detections, reporting location and size information, are selected when the vehicle appearance is mostly visible, such as in daytime or urban nighttime scenarios with sufficient artificial street illumination. Point-based detections, indicating only location information, are used when the vehicle's appearance is not discernible, such as in rural nighttime scenarios with little or no street illumination. The system is designed as a multi-head neural network built on a shared Hourglass backbone that accepts bounding-box and point-based annotations for training and can automatically predict, depending on the scenario, vehicle bounding boxes or point-based predictions. Extensive evaluations on a combined dataset of BDD100K and PVDN demonstrate that the proposed system achieves higher detection accuracy and robustness compared to existing methods, with mean Average Precision (mAP) scores of 0.7134 on BDD100K, 0.6621 on PVDN, and 0.6814 on the combined dataset. Additionally, a self-acquired dataset, FNTVD, further enhances the evaluation by providing real-world driving conditions. The system also achieves real-time performance at 45.45 FPS, making it suitable for practical applications.

INDEX TERMS Computer vision, convolutional neural networks, deep learning, nighttime detection, vehicle detection.

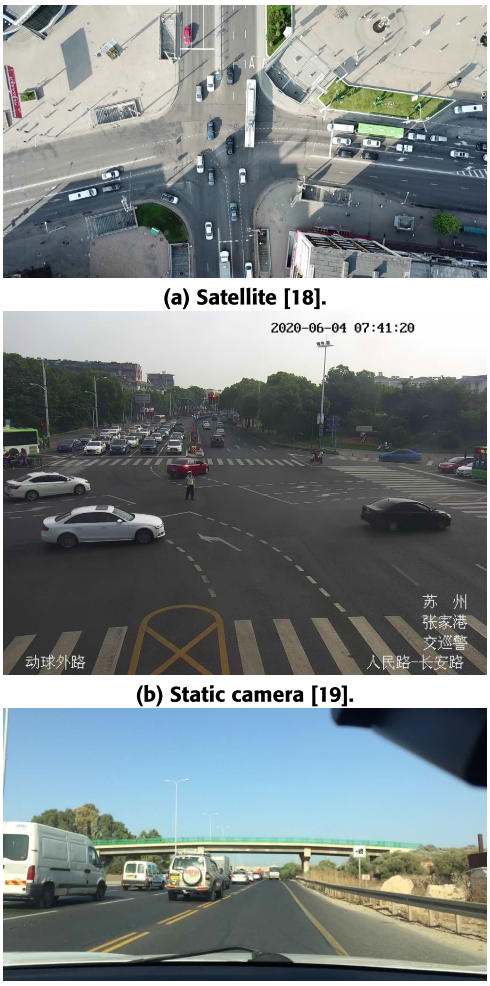
I. INTRODUCTION AND RELATED WORKS

Over the last few years, the total number of vehicles has increased exponentially along with the rise of new and improved technologies. As more vehicles imply more traffic, more accidents, and overall, more road-related problems, Intelligent Transportation Systems (ITS) have been extensively studied. An ITS can be defined as the application of advanced technologies in the field of transportation to improve the efficiency, safety, and sustainability of the transportation network [1]. Therefore, ITS uses a wide range of technologies, such as sensors [2], communication systems [3], and data analytics [4], in order to gather real-time information about traffic flow, road conditions, and other

The associate editor coordinating the review of this manuscript and approving it for publication was Turgay Celik.

relevant factors. This information is then used to optimize the performance of the transportation system by reducing congestion, improving safety, and enhancing the overall user experience. Although there are many applications of ITS, nowadays, one of the most recent and interesting research areas is the design of automated vehicles [5]. Automated vehicles are vehicles that use advanced sensors, artificial intelligence, and communication technologies to navigate without the need for human intervention. These vehicles are capable of sensing their environment and making decisions about speed, direction, and braking based on real-time data. Hence, they have the potential to revolutionize transportation by improving safety, reducing congestion, and increasing mobility. Therefore, problems related to automated vehicles have been largely studied and can be easily found in the literature. Some examples of these significant research topics, each presenting multiple challenges, include driver assistance [6], collision avoidance [7], pedestrian and cyclist recognition [8], [9], and vehicle detection [10], [11].

State-of-the-art works in vehicle detection have studied the application of several types of sensors, such as LiDAR [12], radar [13], and cameras [14], which provide the possibility of implementing multiple scenarios based on the used sensor configurations. Cameras, in particular, have gained significant attention due to their widespread use in various settings, ranging from external cameras, such as stationary traffic cameras [15] or satellite cameras [16] for traffic monitoring and surveillance, up to onboard cameras in vehicles [17]. Some examples of these different scenarios are shown in Fig. 1 (images belong to the aerial cars [18], SEU_PML [19], and BDD100K [20] datasets respectively). However, most of the studies focus on vehicle detection during the day under favorable weather conditions [11], [21]. In comparison, it is difficult to find studies that address images taken in poorly lit or reduced visibility conditions, despite the fact that the chances of accidents increase under these circumstances. According to the Spanish Directorate of Traffic (DGT), 29% of traffic accidents in Spain occur at night, and 40% of mortal victims are from those accidents, despite fewer vehicles being on the road. On the other hand, nighttime situations are not only the most common under challenging conditions but also the most difficult for vehicle detection [22], and even so, most of the available studies are focused on other situations related to unfavorable weather conditions [23], [24], [25]. In nighttime situations, the only visible characteristic of vehicles is often their distorted lights, making it difficult to detect their shape accurately. In night urban scenarios, lighting is typically provided by streetlights and vehicle headlights or taillights. However, streetlights often fail to provide sufficient illumination for effective detection. The challenge becomes even more pronounced in night rural environments, where the level of illumination is extremely low. In such cases, the vehicle's appearance or shape may be nearly impossible to discern, as only the headlights or taillights are visible, often in the form of flashes or complex light patterns. As a result, there



(a) Satellite [18].

(b) Static camera [19].

(c) On-board camera [20].

FIGURE 1. Examples of scenarios using different camera types and settings.

is a lack of effective strategies that satisfactorily address this key but challenging situation. This challenge is not limited to vehicle detection but also extends to pedestrian detection, where low visibility can significantly impact the performance of detection systems. For example, [26] explores a wireless sensor network-based approach for real-time pedestrian detection and classification, emphasizing the importance of integrating multiple sensor modalities to improve detection accuracy in intelligent transportation systems. Reference [27] analyses the performance of four state-of-the-art object detectors for pedestrian and vehicle detection under low illumination conditions. In the analysis, the authors compare the object detectors YOLOv3, SSD, Faster R-CNN, and RetinaNet. These studies underscore the importance of developing detection systems that can handle the challenges posed by urban nighttime environments, where lighting conditions can vary significantly.

Nighttime vehicle detection methods based on computer vision can be categorized into three types: methods based

on prior knowledge, machine learning algorithms, and deep learning techniques. Prior knowledge-based methods rely on pre-existing knowledge or assumptions about the data being analyzed in order to make more accurate, reliable, and effective predictions. In computer vision applications, prior information is based on characteristic features of the appearance of the object. For vehicle detection, the most common characteristics are based on color [28], texture [29], symmetry [30], edge [31] or contour [32]. However, in nighttime scenarios with poor illumination conditions, light beams produce large flashes that tend to overshadow the vehicle. This means that the previously mentioned characteristics are not usually visible. Therefore, prior knowledge methods for nighttime vehicle detection usually focus on headlights [33], taillights [34] or even the color [35] of those lights. In [36] the authors propose an algorithm consisting of headlight segmentation, headlight detection, headlight tracking and pairing and vehicle classification. Reference [32] presents a framework for nighttime vehicle recognition that utilizes both headlights and taillights. Their multi-camera representation groups and reconstructs vehicle lights based on mutual geometric distances between different vehicle components, successfully removing duplicated lights and compensate for missing lights in the detection process. Overall, the main disadvantages of prior knowledge-based methods are their limited applicability for a wide range of real scenarios and the difficulty of acquiring prior information via handcrafted features, leading to a poor performance. On the other hand, as these methods rely on pre-existing knowledge about the data and the problem being solved, they may not be suitable for all types of applications. Lastly, it may be difficult or expensive to acquire the prior knowledge needed to guide the analysis, which may lead to biased or inaccurate results.

Methods based on machine learning usually combine feature descriptors with traditional (shallow) classifiers to obtain highly accurate predictions. An example is given in [35] where the authors compute Haar-like features over gray-level images with the aim of finding vehicle rear lights, delivering those features to a modified version of the Adaboost classifier that uses an active learning approach. Reference [22] proposes a method based on the concept of the Grid of Spatial-Aware Classifiers (GSAC), which are an array of Support Vector Machine (SVM) classifiers that shared the same input image feature vector based on Histogram of Oriented Gradients (HOG). This image descriptor is differently processed by each foveal classifier in the grid, depending on its spatial position in the grid. However, the main disadvantage of machine learning methods is given by their limited performance in real scenarios, due to the use of shallow models with relatively few trainable parameters and the use of general handcrafted features that do not necessarily fit to the final application [37]. Besides, machine learning algorithms that combine feature extraction and classification typically need separate training processes, which means they are not end-to-end trainable.

Unlike previous approaches, deep learning techniques are very complex end-to-end trainable models that automatically learn effective image features tailored to the specific application. Most of the available literature about deep learning models for nighttime vehicle detection is based on advanced object detectors, such as the You Only Look Once (YOLO) family or Region-based Convolutional Neural Networks (R-CNN) family. In [38], a Faster R-CNN approach composed by three main components is suggested for nighttime vehicle detection. The authors propose a deep residual network that automatically extracts features, combined with a spatial attention mechanism that ensures that the network focuses more on the road than the background. Additionally, instead of using the classical Non Maximum Suppression (NMS), they propose a Soft-NMS to minimize the number of missed-detection vehicles. Reference [39] makes use of an SSD architecture using MobileNetv2 as the backbone to perform the feature extraction. They also use an attention mechanism for feature weighting and bottom-top feature fusion strategy to improve the detection performance. Both YOLOv3 and an improved version of YOLOv4 are used in [40], introducing a training process for two new activation functions, called MISH and SWISH that improves the detection performance. In [41], an improved version of YOLOv5 is proposed for real-time detection of vehicles at night. They propose to use shallow high-resolution features and to increase the size of the feature map up to 160×160 (instead of 20×20) to improve the detection of smaller objects. In [42] a combination of YOLOv5 and a tracking algorithm to monitor highway traffic at night. Reference [43] proposes an improved YOLOv5-based method for enhanced detection of small and occluded road vehicle targets, addressing the challenges of detecting vehicles in complex urban environments with limited visibility. Similarly, [44] introduces an improved KSC-YOLOv5 model for night-time vehicle target detection, focusing on optimizing the detection of vehicles under low-light conditions by enhancing the model's ability to capture fine-grained features.

Previous deep learning models have proven effective in nighttime situations of metropolitan areas, where the streetlights lighting up the scene offer a minimum level of illumination that allows vehicle shapes to be distinguishable. However, non-metropolitan nighttime scenarios are typically much worse illuminated, where front and rear lights of a vehicle are frequently perceived as large light flashes that prevent the observation of the vehicle shapes and appearances, as shown in Fig. 2. Under these circumstances, the existing object detectors are not sufficiently accurate, since they assume that the required bounding box annotations for training the underlying neural network models effectively enclose the shape and appearance of the vehicles. As a result, they are not effectively trained leading to wrong predictions. Even more, the use of bounding boxes to annotate the perceived vehicle light flashes seems misleading, since they can occupy large image regions not related to the actual position. Under these extreme but common scenarios, point-base annotation seems more appropriate for representing the vehicle location, since no size, shape, or appearance information can be collected. To address this, [22] introduces

(a) Daytime scenario [20]. (b) High lighting nighttime scenario [20]. (c) Low lighting nighttime scenario [47].

FIGURE 2. Daytime and nighttime images acquired in both urban and non-urban scenarios.

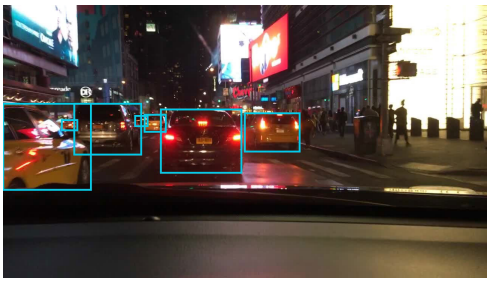
point-based ground-truth annotations, which offer a more practical solution by focusing on the centroid of the object when its boundaries are not easily discernible. However, as the proposed system is solely based on point-based detections, size and location information cannot be inferred in any scenario, even if the vehicle is fully visible.

While significant progress has been made in vehicle detection, existing methods predominantly focus on daytime scenarios or well-lit urban environments, leaving nighttime and rural areas inadequately addressed. This gap is particularly concerning given that nighttime driving poses unique challenges, such as reduced visibility and increased accident rates. Additionally, current state-of-the-art methods, such as Faster R-CNN and YOLO, rely on bounding-box annotations and perform well in well-lit conditions but struggle in low-light rural scenarios where only vehicle lights are visible. This study addresses these limitations by introducing a dual annotation and prediction approach that combines bounding-box and point-based detections, enabling accurate vehicle detection across diverse lighting conditions. Furthermore, the learnable non-maxima suppression mechanism filters out false positives caused by glare or reflections, ensuring accurate detection in chaotic scenes.

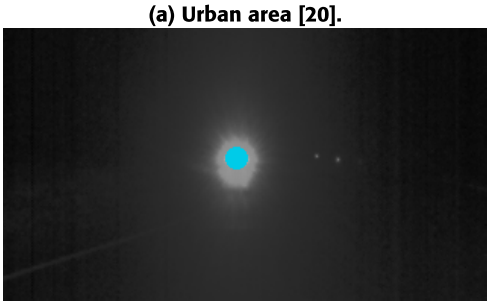
Recent advancements in intelligent transportation systems (ITS) have introduced innovative approaches to optimize network connectivity and traffic management. For instance, the integration of LoRa (Long Range) communication with distributed machine learning has been proposed to enhance network connectivity for green and intelligent transportation systems [45]. This approach leverages low-power, long-range communication to enable efficient data sharing and coordination among vehicles and infrastructure, reducing energy consumption and improving system scalability. Similarly, reinforcement learning-based intelligent traffic signal controllers have been developed to manage heterogeneous traffic in real time [46]. These controllers dynamically adjust signal timings based on traffic conditions, reducing congestion and improving overall traffic flow. While these advancements address critical aspects of ITS, such as connectivity and traffic management, the challenge of accurate vehicle detection in low-light conditions remains understudied. This work focuses on addressing this gap by proposing a robust nighttime vehicle detection system that complements these advancements, contributing to the broader goal of safer and more efficient transportation systems.

This paper proposes a multi-granular visual vehicle detector capable of handling environments with extreme variations in illumination, ranging from bright daytime scenarios to challenging nighttime conditions where the vehicle's appearance may be entirely lost. The multi-granular detection mechanism plays a crucial role by providing either bounding box detection or point-based detection, depending on the visibility of the vehicle's appearance. Bounding box detection is applied when the vehicle's appearance (shape, contour, or structure) is visible, typically in daytime or urban nighttime scenarios with sufficient artificial street illumination. Point-based detection is used when the vehicle's appearance is not discernible, as in rural nighttime scenarios with little to no street illumination, where only vehicle lights (front or rear) are visible. These lights may appear as flashes or complex light patterns that occupy large or disconnected regions in the image, making bounding box predictions impractical. Fig. 2 illustrates these three main scenarios: daytime, urban nighttime, and rural nighttime (images belong to the BDD100K [20], BDD100K [20], and PVDN [47] datasets respectively). In daytime scenarios, the vehicle's appearance is clearly visible (except for potential occlusions), while in urban nighttime scenarios, it is less distinct but still observable. In these cases, the detector predicts a bounding box that includes location and size information. Conversely, in rural nighttime scenarios, the vehicle's appearance is typically not visible, leaving only the lights as observable features. Here, the detector predicts a point-based detection since size information is unavailable. The proposed multi-granular vehicle detector automatically determines the type of prediction (bounding box or point) based on the level of visible vehicle appearance in each scene, providing as much location and/or size information as possible. The underlying neural network can be seamlessly trained using datasets with mixed annotations, including both bounding boxes and points. Fig. 3 illustrates examples of ground-truth annotations corresponding to different illumination levels (images belong to the BDD100K [20] and PVDN [47] datasets respectively).

The proposed multi-granular vehicle detector is designed as a multi-head (multi-branch) neural network built on a shared CNN backbone. The backbone utilizes an Hourglass



(a) Urban area [20].



(b) Non-urban area [47].

FIGURE 3. Examples of a bounding-box annotated nighttime urban scenario and a point-based annotated (blue point) rural scenario.

architecture [48], which extracts high-level semantic feature representations from input images. The multi-head block leverages these feature maps to predict multiple outputs: a) Heatmaps for vehicle location, where duplicate predictions for the same vehicle instance are prevented by a learnable non-maxima suppression module, b) Offsets to refine minor deviations in location predictions, and c) Prediction types (bounding box or point). Additionally, for bounding box detections size information is also predicted.

The main contributions of this paper are summarized below, highlighting the core innovations and features of the proposed system:

- A multi-granular vehicle detector capable of handling environments with extreme illumination variations, from bright daytime scenarios to challenging nighttime conditions (both urban and rural), where the vehicle's appearance may be completely lost.
- Multi-granular capability simultaneously supports bounding box and point-based detections (and annotations for training) based on the illumination conditions and the visibility of vehicle appearances.
- Comprehensive training datasets by means of a combination of two datasets (BDD100K and PVDN) and an effective training procedure to learn a robust model capable of detecting vehicles under diverse illumination conditions, spanning well-lit urban areas to poorly lit rural scenarios, while seamlessly integrating bounding box and point-based annotations.

The proposed system's ability to dynamically adapt between bounding box and point-based predictions offers significant advantages for vehicle detection in varying lighting conditions. In well-lit urban environments, where vehicle shapes, contours, and structures are clearly visible, bounding box predictions provide precise location and size information, enabling accurate detection and tracking. For example, in a daytime urban scenario with sufficient artificial lighting, the system can detect a car's full shape and predict its bounding box, which is essential for tasks like collision avoidance and path planning.

In contrast, in poorly lit rural environments, where only vehicle lights (e.g., headlights or taillights) are visible, bounding box predictions become impractical due to the lack of discernible vehicle shapes. Here, the system switches to point-based predictions, focusing on the centroid of the vehicle lights. For instance, in a dark rural road scenario with minimal street lighting, the system detects the vehicles as a single point, ensuring reliable detection even when the vehicle's shape is obscured. This adaptability allows the system to maintain high detection accuracy across diverse lighting conditions, making it suitable for real-world applications in both urban and rural settings.

It is important to note that the proposed architecture has been designed to meet the requirements of a widely-used automotive embedded system, specifically the TDA4VM by Texas Instruments. This ensures the feasibility of the implementation in a real vehicle and aligning the development with the goals of the project ''Percepción Inteligente para los Vehículos Autónomos y Conectados'' (InPercept) under which this research has been conducted.

II. DEVELOPMENT

This paper aims to develop a nighttime vehicle detection system that utilizes onboard cameras as its primary sensing modality, addressing some of the most significant challenges in nighttime scenarios. The paper introduces the combined use of traditional bounding-box and point-based ground-truth annotations, which represents a noteworthy advantage for nighttime scenarios. Traditionally, bounding boxes or masks are employed for object annotation, but in low-light conditions, the shape of the object may not be clearly discernible due to limited illumination. In such cases, point-based annotations positioned in the centroid of the object prove to be a more suitable alternative. This means that when lighting conditions are sufficient, bounding boxes provide valuable size and location information for detected objects. However, in poorly-lit scenarios, it becomes impossible to extract reliable size information, making point-based annotations a more suitable alternative. Therefore, the paper proposes combining these two approaches to develop a fully functional system capable of accurately predicting all relevant information across different scenarios. This approach not only simplifies the labeling process of images from poorly-lit nighttime scenarios, as it is easier and faster than annotating bounding boxes or masks, but also enhances the system performance for all types of nighttime conditions. Moreover, by leveraging point-based detections, the system eliminates the need for anchors, simplifying the architecture

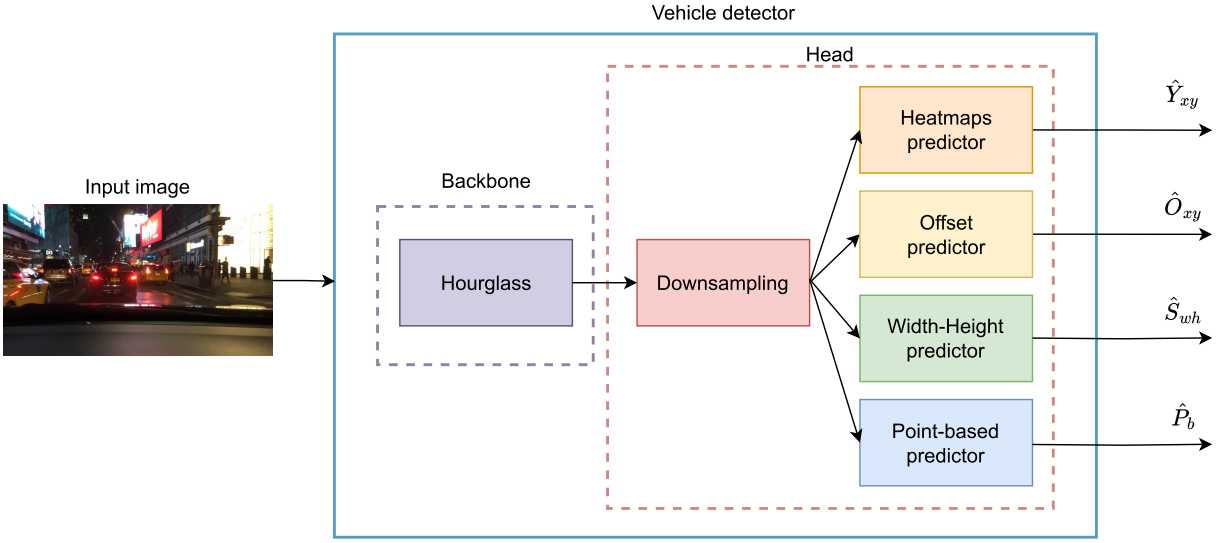


FIGURE 4. System architecture.

and offering significant advantages for the dual approach in adapting to different nighttime scenarios.

Anchors are pre-defined bounding boxes used in most modern object detection algorithms that serve to match objects of different scales and aspect ratios. However, an anchor-free approach simplifies the overall architecture by eliminating the reliance on anchor boxes. This reduction in complexity not only streamlines the model, but also leads to faster inference times and lower computational requirements. Additionally, by removing anchors, object detection algorithms become more flexible for detecting new object categories, since no preprocessing of the training database is required to infer the suitable set of anchor boxes that accommodates to the different and varying sizes and shapes. Moreover, anchor-free methods exhibit superior performance in detecting small objects, which can be challenging for anchor-based approaches due to fixed anchor sizes. Consequently, the system becomes better equipped to identify and localize small objects accurately, which is particularly valuable in various domains such as surveillance or robotics. Hence, an anchor-free detection system not only enhances adaptability but also provides significant advantages in handling challenging scenarios encountered, especially in low-light or nighttime conditions.

The overall architecture is depicted in Fig. 4 and provides a comprehensive visual representation of the system's components and their interconnections. The main module, referred to as the vehicle detector, employs a CNN architecture to process and analyze the input images, generating bounding box or point-based predictions for vehicle detection. This module consists of two submodules: the backbone, which is based on the Hourglass architecture, and the head detection block, which predicts, for each vehicle, the position, the size when applicable, and whether the prediction is based on a bounding box or a point.

A. NIGHTTIME VEHICLE DETECTION SYSTEM

The neural network Vehicle Detector serves as the fundamental module of the proposed vehicle detection system, representing the core component responsible for achieving accurate and robust detection results. The proposed neural network architecture is based on CenterNet [49], which in turn is based on CornerNet [50]. Following their architecture, it comprises two crucial blocks: the CNN backbone and the head detection block (see Fig. 4). The CNN backbone is based on the Hourglass architecture, depicted in Fig. 5, which was originally introduced in [48]. The Hourglass architecture has been selected for its capability to effectively capture small object details and preserving fine-grained features at multiple scales, which is a critical need when detecting distant or partially obscured vehicles at night. It embodies an encoder-decoder design, wherein the input undergoes a series of transformations to derive an n-dimensional embedding, which is subsequently upsampled, reverting to its original shape. This design favors the extraction of the most salient and discriminative features of the input images. These features are commonly referred to as feature maps, playing a pivotal role in enabling the network to effectively perceive and discern objects, with a particular focus on vehicles in this context.

By utilizing the Hourglass network as the backbone, the proposed vehicle detection system benefits from its inherent ability to capture and represent hierarchical features at multiple scales. This characteristic empowers the network to robustly handle the diverse variations in vehicle appearance, shape, and size that may occur across different scenarios. Moreover, the Hourglass network's encoder-decoder architecture enables the network to incorporate both global and local contextual information, facilitating a holistic understanding of the image and enhancing the network's capacity to detect vehicles accurately.

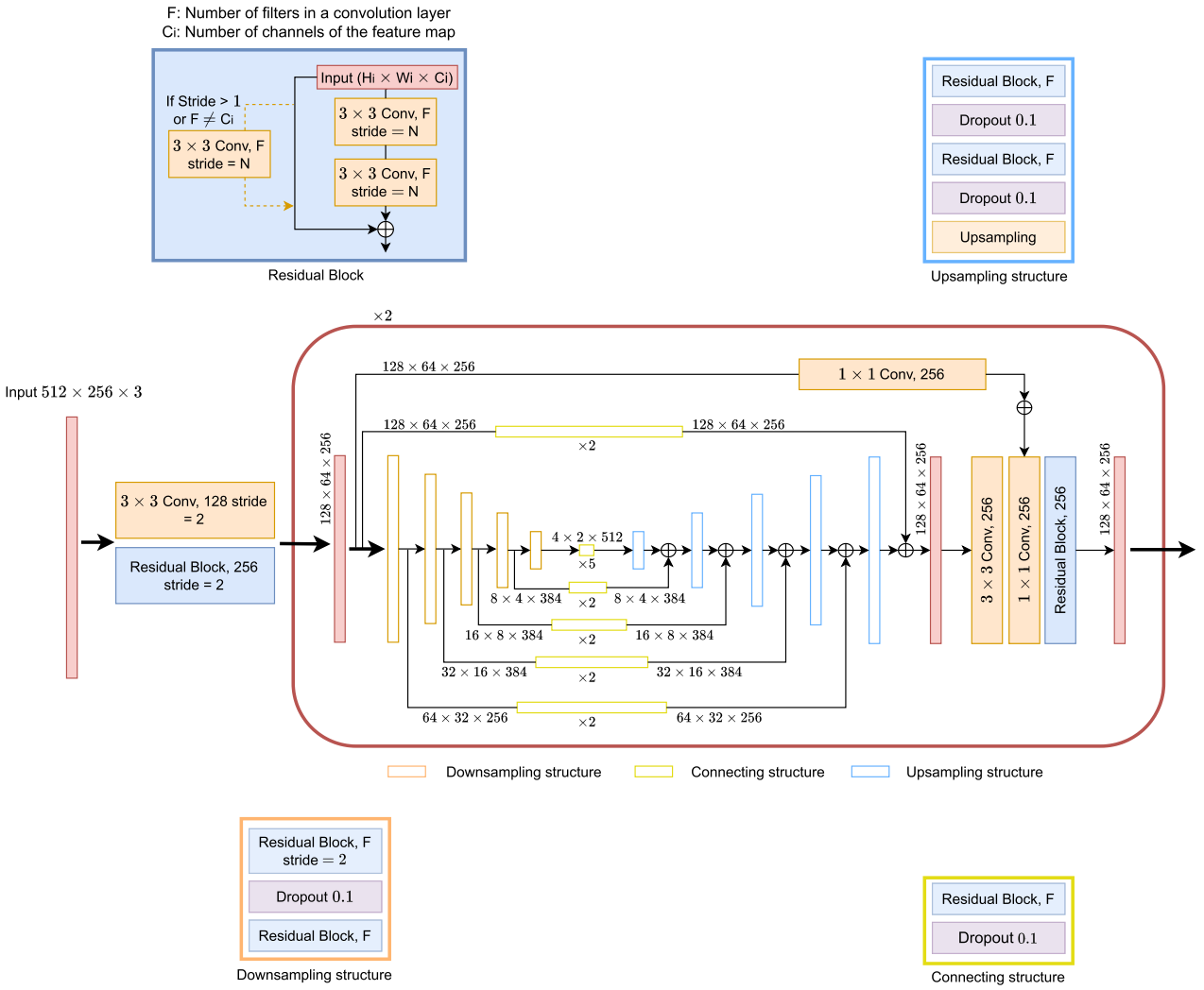


FIGURE 5. Hourglass backbone architecture.

More in detail, the main motivation behind the Hourglass design stems from the necessity to capture information across various scales. The Hourglass architecture is a simple and minimal design that effectively captures features and generates pixel-wise predictions. In order to process and consolidate features across scales, existing approaches independently process the image at multiple resolutions to finally combine the features in the network [51], [52]. However, [48] opted for a different approach, the Hourglass architecture, which employs a single pipeline with skip connections to preserve spatial information at each resolution. This architecture gradually reduces the resolution of the features using a downsampling structure composed of several convolutional layers and max-pooling for downsampling the resulting feature maps. At each max-pooling step, the network is divided into two different branches. One continues the downsampling approach, and the other one applies additional convolutions at the original pre-pooled resolution. This process continues until the lowest resolution of 4×2 pixels is reached, allowing the application of smaller spatial filters that analyze features across the entire image space. To integrate information from different resolutions, the Hourglass design follows the procedure outlined by [51], consisting in a nearest neighbor upsampling from lower resolutions that are element-wise added with the corresponding sets of features of the downsampling part, which has been processed by a connection structure. The Hourglass architecture maintains a symmetric topology, meaning that for every layer from the downsampling part, a corresponding layer exists in the upsampling part. After reaching the output resolution, the network applies two consecutive 1×1 convolutions to generate the final predictions. The output of the backbone is a set of heatmaps, where each heatmap represents the probability of an object's presence at each pixel location.

The head detection block uses the previous feature maps from the CNN backbone to estimate whether the prediction is based on a bounding box or a point, the center-point location of the vehicles, and their size when applicable. The

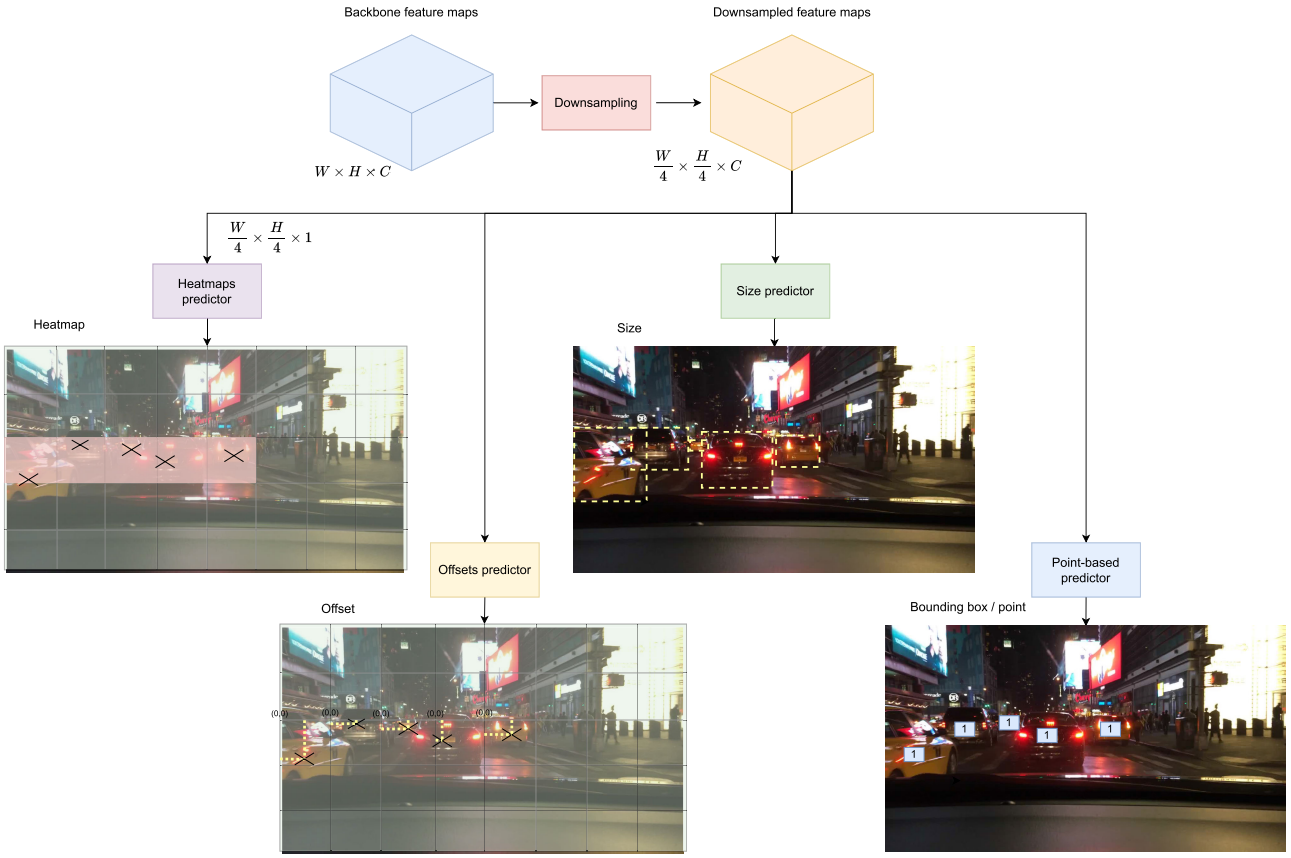


FIGURE 6. Prediction pipeline. The upper branch shows the heatmaps calculation process, while the second branch indicates the offsets prediction. The third branch infers the bounding box size when applicable. Finally, the lower branch predicts whether a bounding box or a point detection is needed.

Head block is composed by five submodules: Downsampling, Heatmap, Offset, Width-Height, and Point-based (See Fig. 6). The Downsampling submodule, downsamples the feature maps to facilitate the subsequent prediction tasks, creating a grid of cells, where each cell corresponds to an element of the downsampled feature map, as illustrated in Fig. 6.

By representing the image as a grid of cells, the network gains the ability to associate specific characteristics extracted from the image with the presence of vehicles in each cell. This is achieved through the utilization of a heatmap-based prediction layer that computes a heatmap from the downsampled feature maps. Each value in the heatmap indicates the likelihood or confidence of a vehicle's presence within the corresponding cell. Values close to 1 mean a high probability of vehicle presence, while values closer to 0 indicate a low probability of encountering vehicles.

To address the issue of multiple predictions per vehicle, since large vehicles can span multiple cells, a non-maxima suppression step is incorporated, which eliminates duplicated predictions. Specifically, the non-maxima suppression is implemented with a 3×3 maxpooling kernel, which selects only one local maximum from every 3×3 neighborhood. This process ensures that each vehicle instance is represented by a single prominent peak in the heatmap, regardless of its size or extent. The combination of downsampling, heatmap generation, and non-maxima suppression exemplifies the system's ability to effectively process the feature maps and produce robust detection outcomes.

In order to precisely determine the location of vehicles within the image, the head detection block incorporates another prediction branch, which builds upon the previous heatmap based predictions. It improves the prediction of the position of each detected vehicle by estimating an offset regarding the top left corner of each cell in the grid. As a result, the reliable and precise locations of vehicles within the image are obtained. This contributes to the system's overall accuracy and reliability, which are crucial for applications such as autonomous driving and advanced driver assistance systems.

To determine whether the network should predict a bounding box or a point for each detected vehicle, an additional prediction branch is introduced in the head detection block. This branch takes the feature maps produced by the Hourglass backbone as input and employs a binary classification mechanism. Using a sigmoid activation function, the branch outputs a probability value for each detection, where a value of '1' indicates the use of a bounding box, and a value of '0' corresponds to a point-based prediction. This approach

ensures that the system dynamically adapts to the lighting conditions, selecting the most suitable representation (bounding boxes for well-lit scenarios and points for poorly-lit environments) thus enhancing the overall detection accuracy and flexibility.

In addition to predicting the location of vehicles, the head detection block is also responsible for inferring the size of each vehicle when applicable. For vehicles detected using bounding boxes, the system calculates the width and height of each box by regressing the object's size based on its coordinates. Specifically, the dimensions are computed as the difference between the top-left and bottom-right corners of the box, represented by $(x_2 - x_1, y_2 - y_1)$. This process is optimized using a standard L1 distance norm, with the resulting size predictions captured in heatmaps of fixed dimensions for all object categories. By simplifying the computation to use a single heatmap for all objects, the system reduces computational complexity while maintaining accurate size predictions for vehicles.

B. TRAINING

The proposed system has been evaluated on an Intel Core i9-13900K @ 5.8GHz CPU along with two NVIDIA RTX 4090 (24GB) GPUs. Pytorch using CUDA in Python was used to implement the model, using PyCharm as the computing platform. A dynamic combination of four distinct loss functions was employed to improve accuracy: a variation of the Focal loss for the predicted location extracted from the non-maxima suppression of the heatmaps [53], the MAE (Mean Absolute Error) loss for the relative coordinate refinement of these location predictions, a Binary Cross-Entropy loss to determine whether the prediction is based on a bounding box or a point, and a standard $\ell 1$-norm loss for regressing the size of the bounding box when applicable.

More in detail, the Focal loss function compares and penalizes the differences between the predicted heatmaps ($\hat{Y}_{xy}$) and the ground-truth heatmaps (Y_{xy}). This function is based on the traditional Binary Cross-Entropy function, but it adds a modulating factor α ($(1 - \hat{Y}_{xy})^{\alpha}$ if the ground-truth value is $Y_{xy} = 1$ and $\hat{Y}_{xy}^{\alpha}$ otherwise) that reduces the penalization of correct predictions. This loss has been further modified to cover the special case of heatmap prediction as in (1).

$$L_k = -\frac{1}{N}\sum_{xy}\begin{cases}(1-\hat{Y}_{xy})^{\alpha}\log(\hat{Y}_{xy}) & \text{if } Y_{xy}=1\\ (1-Y_{xy})^{\beta}\hat{Y}_{xy}^{\alpha}\log(1-\hat{Y}_{xy}) & \text{otherwise}\end{cases} \tag{1}$$

More specifically, the ground-truth, encoded as a heatmap, has been adapted to represent the position of the labelled points as a bi-dimensional discretized Gaussian $(1 - Y_{xy})^{\beta}$, where β is a tunable parameter, which improves the learning process by smoothing the penalization of correct predictions. This bi-dimensional Gaussian representation of the ground-truth is defined in (2), where x, y are the coordinates of the center of each grid cell, g_x, g_y are the annotated coordinates for each vehicle, and σ is a hyperparameter.

$$Y_{xy} = \exp\left(\frac{(x-g_x)^2 + (y-g_y)^2}{2\sigma^2}\right) \tag{2}$$

The expression of the MAE loss is depicted in (3). It calculates the $\ell 1$-norm between the predicted offset ($\hat{O}_{xy}$) and the annotated offset (O_{xy}) and penalizes it in order to minimize it.

$$L_{off} = \frac{1}{N}\sum_{xy}|\hat{O}_{xy} - O_{xy}| \tag{3}$$

The size of each detected object, when applicable, is inferred by regressing the width and height of the bounding boxes using a standard $\ell 1$-norm loss defined in (4). This loss function calculates the difference between the predicted dimensions ($\hat{S}_k$) and the ground truth sizes (S_k) in terms of raw pixel values, without normalization based on the feature map size. By directly comparing the raw pixel coordinates of the predicted and actual bounding box dimensions, the system ensures accurate size predictions that contribute to the overall performance of the detection process.

$$L_s = \frac{1}{N}\sum_{k=1}^{N}|\hat{S}_k - S_k| \tag{4}$$

To determine whether the vehicle should be represented by a point or a bounding box, a Binary Cross-Entropy (BCE) loss, as defined in (5), is employed. This component of the loss function compares the predicted classification ($\hat{P}_k$) (bounding box or point) with the ground truth (P_k), aiming to minimize the discrepancy between them. Specifically, the BCE loss penalizes incorrect classifications by calculating the cross-entropy between the binary prediction and the actual annotation, ensuring the network accurately switches between point-based and bounding-box-based detections depending on the lighting conditions.

$$L_{bp} = \frac{1}{N}\sum_{k=1}^{N}(P_k\log(\hat{P}_k) + (1-P_k)\log(1-\hat{P}_k)) \tag{5}$$

The overall loss or cost function used during training, as shown in (6), combines the previous four loss functions to derive a comprehensive measure of the model's performance. Finally, through backpropagation and gradient descent optimization techniques [54], the model parameters are updated, aiming to minimize this cost function and improve the overall accuracy and precision of the model predictions.

$$L_{det} = L_k + \lambda_{off}L_{off} + \lambda_s L_s + \lambda_{pb}L_{pb} \tag{6}$$

The hyperparameter settings that have been used to train the network, such as the learning rate or the total number of epochs, are listed in Table 1.

III. EXPERIMENTS AND RESULTS

In this section, the datasets used to evaluate the proposed system are first presented. Next, Section III.B gives information on the computed metrics. Section III.C provides numerical

TABLE 1. Description of the hyper-parameters used to train the network.

Hyperparameter	Value
Pre-trained backbone	ExtremeNet
Initial Learning rate	2.5^{-4}
Number of epochs	140
Batch size	32
Image size	512×256
Learning rate epoch 90	2.5^{-5}
Learning rate epoch 120	2.5^{-6}

and visual performance of the system within both datasets, as well as a comparison with other state-of-the-art solutions.

A. DATASETS

To evaluate the proposed system, three different datasets were used. One of them was the BDD100K, which is one of the largest publicly available datasets. The second one was the PVDN, which contains different scenes of dark rural environments at night. The last one, was a new self-acquired dataset called FNTVD (Ficosa Night Time Vehicle Detection) and that has been made publicly available[1] according to the European Data Protection Regulation.

The BDD100K [20] dataset has been used to train and evaluate the performance of the proposed system. This database is the largest public onboard dataset for vehicle detection, containing 100k images with geographic, environmental, and weather diversity. It comprises images and videos captured from the viewpoint of a vehicle's dashboard camera, providing a diverse set of driving scenarios. These scenarios represent real-world driving conditions encountered in urban, suburban, and highway environments. The dataset includes images captured on city streets, showing scenes with buildings, sidewalks, traffic lights, crosswalks, parked cars, and various road markings. These images reflect typical urban driving environments with intersections, pedestrian activity, and complex traffic flow. Additionally, the dataset contains some images depicting residential neighborhoods with houses, driveways, curbs, and local roads, often associated with lower traffic volumes, narrower streets, and slower driving speeds. Furthermore, images in commercial areas showcase scenes with shopping centers, parking lots, commercial buildings, storefronts, and busy intersections. These scenarios involve higher pedestrian activity and various types of vehicles, including delivery trucks and taxis. Finally, other images in the dataset capture highway driving scenarios, featuring multi-lane roads, overpasses, on-ramps, off-ramps, and highway signage. These scenarios represent higher speed driving conditions encountered on expressways and interstates, and they usually contain a lower level of artificial lighting. Regarding weather conditions, the BDD100K dataset includes images captured in different weather conditions, such as rain, fog, snow, or overcast skies. These images represent challenging scenarios where visibility may be reduced, road surfaces may be slippery,

[1]https://www.gti.ssr.upm.es/data/ficosa

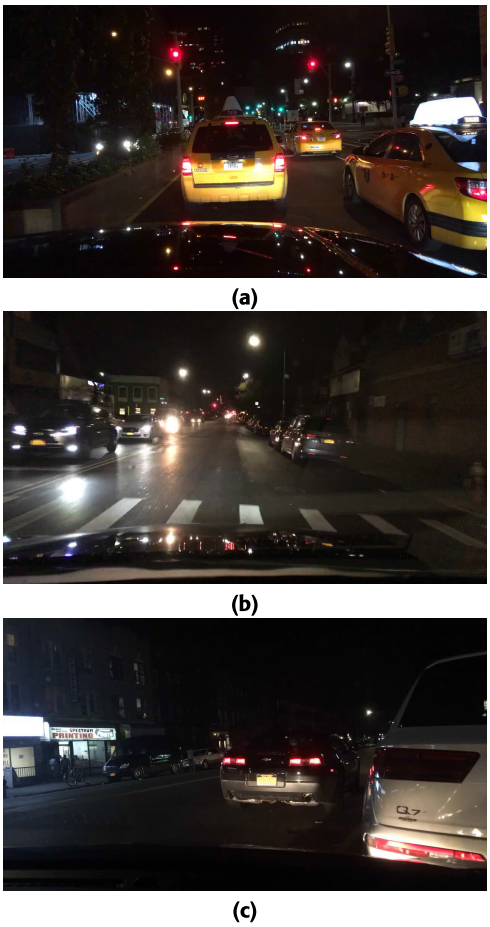


FIGURE 7. BDD100K database nighttime examples. These images depict challenging scenarios characterized by reduced visibility and adverse driving conditions.

and overall driving conditions are impacted. Moreover, the dataset contains images captured during both daytime and nighttime, representing different lighting conditions and visibility levels. Therefore, the dataset provides a diverse collection of images, encompassing a wide range of driving environments and conditions encountered in real-world driving scenarios. However, the nighttime scenes captured in the dataset predominantly feature relatively high levels of artificial lighting, allowing the overall shape of the vehicles to remain discernible. (see Fig. 7).

On the other hand, the PVDN [47] dataset is a novel non-urban dataset that contains roughly 56k greyscale images (see Figure 8). The scenarios depicted within the PVDN dataset include driving situations encountered in rural areas at night. These scenarios may involve open roads, countryside landscapes, and less densely populated regions. Given the rural setting, the dataset likely includes scenes with narrower roads, less artificial lighting, and potentially lower traffic volumes compared to urban or highway environments. As low illumination prevails in these images, perceiving the shape, size, or appearance of objects is usually challenging. Instead, the front and rear lights of oncoming vehicles often manifest as substantial light flashes, obstructing the visibility of the

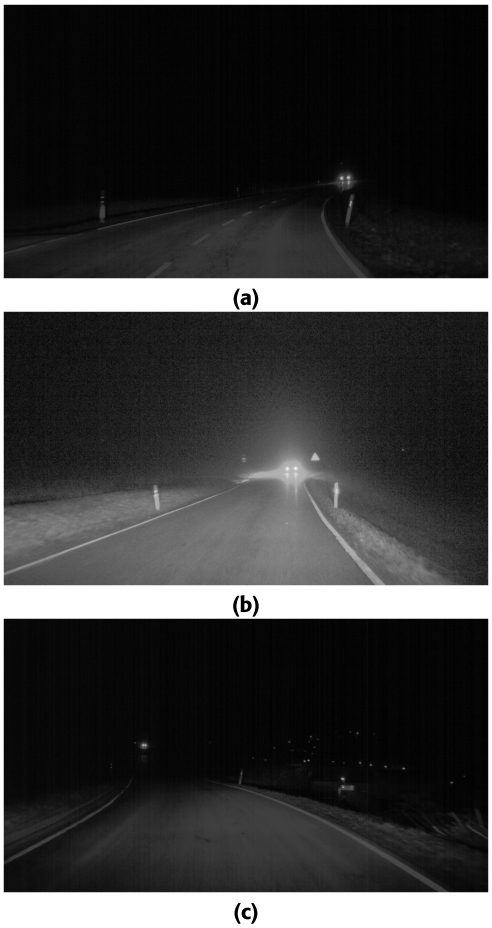


FIGURE 8. Example images of the PVDN database. These scenarios encompass open roads, serene countryside landscapes, and sparsely populated regions, which predominantly exhibit low illumination, making it challenging to perceive the shape, size, or appearance of objects.

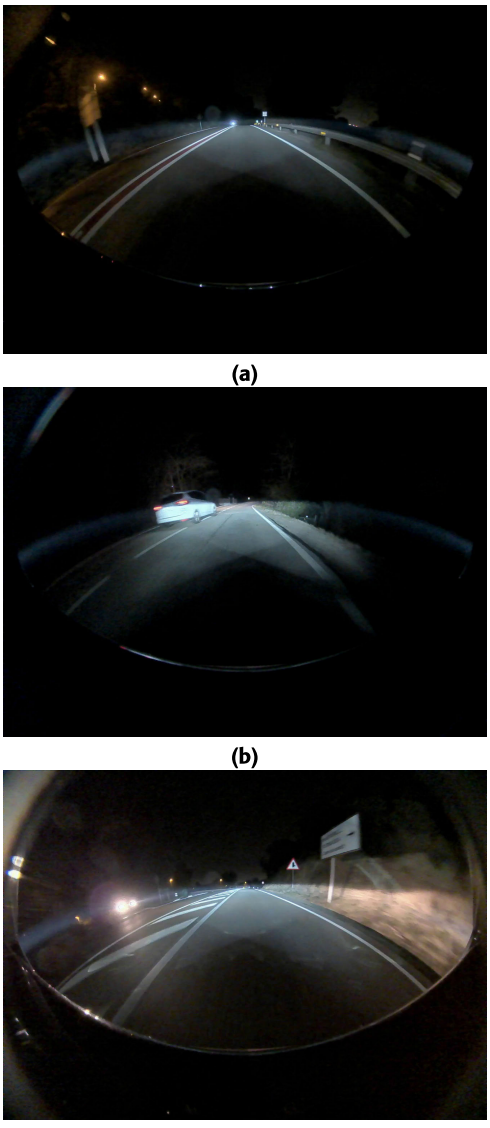


FIGURE 9. FNTVD database examples. These images depict challenging scenarios characterized by reduced visibility and adverse driving conditions. Besides, the dataset is characterized by a high number of small vehicles.

vehicles' shapes and appearances. Additionally, the images within the dataset exhibit a characteristic of very low contrast in most areas, reflecting the overall dim lighting prevalent in the rural nighttime environment. However, there is a distinct area of significantly high contrast focused around the oncoming vehicle, primarily caused by the intense illumination produced by the vehicle's lights. These specific conditions create a unique visual characteristic within the dataset, where the perception of objects, including vehicles, relies heavily on the illumination emitted by their lights. Consequently, the dataset emphasizes the importance of effectively detecting and recognizing vehicles based on the distinctive light patterns they produce.

To ensure that the proposed system can effectively generalize across a broad range of real-world scenarios, the BDD100K and PVDN datasets have been merged for training. This combination allows the system to be exposed to both well-lit urban and highway environments, as well as challenging, poorly-lit rural settings at night. By incorporating the diversity of urban, suburban, and rural nighttime scenarios, the system is able to learn from varying lighting conditions, road layouts, and vehicle appearances without needing to be retrained for different environments. This enhances the system's robustness and adaptability to real-world driving conditions. Merging the BDD100K and PVDN datasets exposes the model to diverse lighting and road conditions, ranging from well-lit urban environments to poorly lit rural areas. While this diversity enhances the model's ability to generalize across different scenarios, it also introduces challenges in ensuring balanced learning. One key challenge is the risk of overfitting to specific scenarios, such as urban environments with abundant artificial lighting, which dominate the BDD100K dataset. To address this, we employ several strategies, such as regularization and SMOTE techniques (Synthetic Minority

Over-sampling Technique) to encourage the model to learn robust features that generalize across diverse conditions. Cross-dataset validation is also performed to evaluate the model's performance on both datasets separately, identifying and addressing any biases in the training process. These strategies collectively help the model balance learning across varied conditions, ensuring robust performance in both well-lit urban environments and challenging rural nighttime scenarios.

The FNTVD dataset captures diverse nighttime driving scenarios characterized by reduced visibility and varying levels of artificial lighting. Recorded using Ficosa's car, equipped with surround-view fisheye cameras that are commonly integrated into modern ADAS systems, the dataset features a 190° field of view and a resolution of 1344×968. This setup enhances peripheral monitoring but introduces geometric distortions that pose challenges for object detection, particularly at the image edges. The dataset includes real-world conditions across urban streets, rural paths, and highways, encompassing dimly-lit roads, sharp curves, and dark backgrounds where glare from headlights complicates detection (see Fig. 9). Notably, it includes a high frequency of small objects and vehicles near the vanishing point, making them particularly challenging to detect and track. These images also feature diverse road elements, such as traffic signs, guardrails, and reflective markers, under varying levels of artificial illumination. The data collection effort spans a broad range of weather conditions, road types, and driving environments, resulting in a comprehensive and dynamic dataset for testing system robustness. All recorded videos have been meticulously annotated using the WebLabel tool [55], ensuring a reliable ground truth for evaluating vehicle detection systems.

B. METRICS

Mean Average Precision (mAP), F-score (F), Precision (P), and Recall (R) have been computed to evaluate the system's performance, as they are widely used evaluation metrics in object detection tasks, and are especially suitable to compare and rank the performance of different detection algorithms.

In bounding-box predictions, True Positives (TP), correctly detected positive entities; False Positives (FP), negative entities incorrectly detected as positive; and False Negatives (FN), positive entities that have been missed, are computed using the concept of Intersection over Union (IoU). However, under poor lighting conditions where point-based detections are employed, IoU is not applicable, and the system switches to use the Euclidean distance (ED) to determine TP, FP, and FN. In these cases, a circular area with a radius equal to the ED is defined around the ground-truth annotation (see Fig. 10). If the predicted object falls within this predefined threshold area, the detection is classified as a TP; conversely, if it falls outside, it is considered a FP. Similarly, if a ground-truth annotation has no associated detection, it is regarded as a FN. Ideally, the area delimited by the ED should not be bigger than those objects. However, since this methodology is applied only in scenarios where accurate information on vehicle size is unavailable, it cannot always be guaranteed. To address this, an experimental analysis to determine the most appropriate ED value, which varies depending on the image resolution, has been conducted. This alternation between IoU and ED ensures that the system adapts to varying lighting conditions and optimally handles both bounding-box and point-based detections.

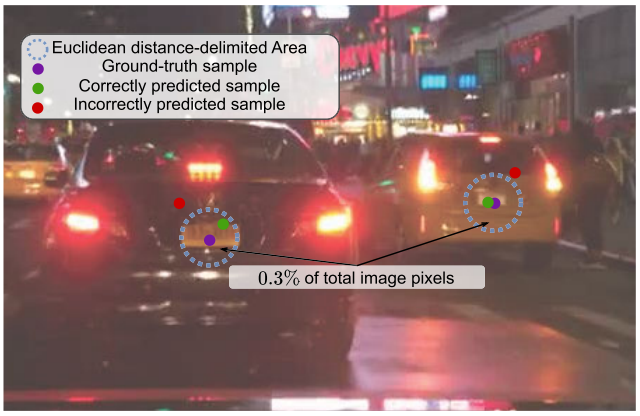


FIGURE 10. TP, FP and FN for point-based mAP computation based on Euclidean distance.

Computational speed, given in frames per second (fps), has also been computed to evaluate the performance of the proposed system, as being capable of achieving real-time performance is crucial for autonomous driving applications.

C. EVALUATION ON ON-BOARD SCENARIOS

1) TESTING THE PROPOSED SYSTEM

The performance results of the proposed vehicle detection system are detailed in Table 2, showcasing its evaluation across the BDD100K (full dataset, daytime-only, and nighttime-only subsets), PVDN, and combined datasets. For the BDD100K dataset, the system achieves a mean Average Precision (mAP) of 0.7134 on the full dataset, which includes both daytime and nighttime images. When evaluated separately, the system achieves a mAP of 0.7212 for daytime images and 0.6929 for nighttime images. The slightly higher performance on daytime images can be attributed to better visibility and clearer vehicle shapes, which facilitate more accurate detection. However, the system maintains strong performance on nighttime images, demonstrating its ability to handle low-light conditions effectively. The F-score, precision, and recall metrics follow a similar trend, with daytime performance being marginally higher than nighttime performance. Notably, the bounding box-point accuracy (Bp-accuracy) remains near-perfect ($\geq$ 0.9998) across all subsets, confirming the system's reliability in determining the correct annotation type (bounding box or point) regardless of lighting conditions.

Regarding the other metrics, the results demonstrate that the proposed system performs robustly across both daytime and nighttime scenarios. While the system achieves slightly higher F-score, Precision, and Recall values for

TABLE 2. mAP, F-score, Precision, Recall, bounding box-point accuracy, and detection rate of the proposed system for the BDD100K, PVDN and combined datasets.

	BDD100K			PVDN	Combined
	Full dataset	Daytime	Nighttime		
mAP	0.7134	0.7212	0.6929	0.6621	0.6814
F-score	0.7554	0.7656	0.7355	0.7374	0.7283
Precision	0.7736	0.7919	0.7434	0.7949	0.7468
Recall	0.7379	0.7411	0.7277	0.6876	0.7106
Bp-accuracy	0.9998	0.9999	0.9998	1.0000	0.9998
Frames Per Seconds (fps)	45.45				

daytime images (0. 7656, 0.7919, and 0.7411, respectively) compared to nighttime images (0. 7355, 0.7434, and 0.7277, respectively), the performance gap is minimal. This indicates that the system is well-suited for nighttime detection, even in challenging low-light conditions. The ability to maintain high performance across both daytime and nighttime scenarios underscores the system's adaptability and robustness, making it a practical solution for real-world applications.

In contrast, the PVDN dataset, which features more challenging rural nighttime scenes with low visibility and sparse lighting, naturally results in slightly lower metrics. The system achieves a mAP of 0.6621, which still demonstrates robust detection performance given the difficulty of the scenarios. The F-score of 0.7374 and precision of 0.7949 suggest that, despite the challenging conditions, the system maintains a high level of accuracy in detecting vehicles, while the recall of 0.6876 reveals the difficulty of identifying all vehicles in low-visibility environments. The Bp-accuracy of 1.0000 reaffirms that the system is highly reliable in determining whether a bounding box or a point annotation should be used.

Finally, for the combined dataset, which integrates both urban and rural nighttime scenarios, the system performs with a mAP of 0.6814, reflecting a strong ability to generalize across varied lighting conditions and scene complexities. The F-score of 0.7283 and precision of 0.7468 show the system's capability to balance accuracy and recall in mixed environments, while the recall of 0.7106 demonstrates that the system consistently detects a significant proportion of vehicles, even when faced with a wide variety of scenarios. The Bp-accuracy of 0.9998 further emphasizes the system's robustness in determining the correct annotation type. Across all datasets, the system achieves a stable frames per second (FPS) rate of 45.45, ensuring real-time processing on our embedded-ready model, making it suitable for real-world applications such as autonomous driving, where quick and efficient detection is crucial.

Overall, while performance metrics are slightly lower in more challenging environments, the system's results are remarkably consistent across diverse datasets. The slight variations between metrics across datasets are small enough to assert that the system maintains steady and reliable performance, regardless of the specific scenario. This demonstrates the system's adaptability and robustness, ensuring effective vehicle detection across a wide range of nighttime conditions.

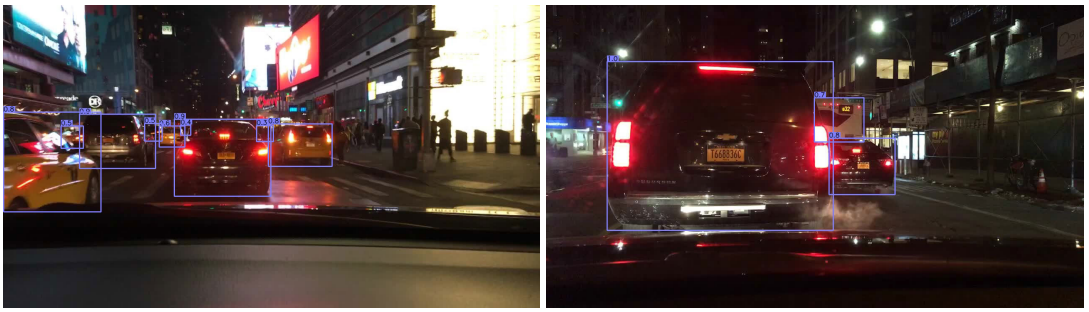
Some examples of the obtained results on critical images are shown in Fig. 11, where Fig. 11.a shows the performance of the system on the BDD100K dataset, Fig. 11.b corresponds to the PVDN dataset, and Fig. 11.c corresponds to the combined dataset. All the presented images contain vehicles from different perspectives, exhibiting different shapes, sizes, and occlusions. Remarkably, the network is capable of detecting all vehicles without any incorrect prediction. This proves the great predictive power and generalization ability of the network in different situations at night.

The performance achieved by the proposed system for nighttime scenarios shows its potential for various real-world applications. It can be attributed to several high-level advantages of the system. First, the use of point-based detections, which fits better the flashlights of the vehicles during the night than the classical bounding boxes. Second, the suppression of the anchors typically employed by other object detection algorithms, which simplifies the network architecture and reduces its computational requirements. And third, the integration of non-maxima suppression within the network architecture, instead of including it in a post-processing step that increases the detection rate of the network.

2) EVALUATION ON A REAL-WORLD SCENARIO: THE FNTVD DATASET

The FNTVD dataset is used in this work as a qualitative benchmark to illustrate the system's performance across diverse nighttime driving scenarios. Due to notable differences between the public datasets and the FNTVD dataset, quantitative comparisons are not applicable. Specifically, the BDD100K and PVDN datasets employ an internal standard camera, while the FNTVD dataset utilizes an externally mounted fisheye camera between the vehicle's headlights. This positioning, combined with the fisheye lens, introduces unique distortions, altered scaling, and varied lighting effects that differ significantly from the BDD100K and the PVDN datasets images.

These differences in camera perspective and resolution affect how the network interprets object scaling and illumination, making direct quantitative performance comparisons unreliable. Therefore, the FNTVD dataset serves as a



(a) Performance on the BDD100K dataset

(b) Performance on the PVDN dataset

(c) Performance on the combined dataset

FIGURE 11. Example of the performance of the proposed system for the BDD100K, PVDN and combined datasets.

qualitative benchmark, offering insights into the model's robustness across varied nighttime conditions.

Examples of the performance of the system on challenging FNTVD images are shown in Fig. 12, highlighting the adaptability of the model in detecting vehicles from multiple perspectives and in varied lighting conditions. It is important to notice that the evaluation has been carried out using the model trained on the combined datasets

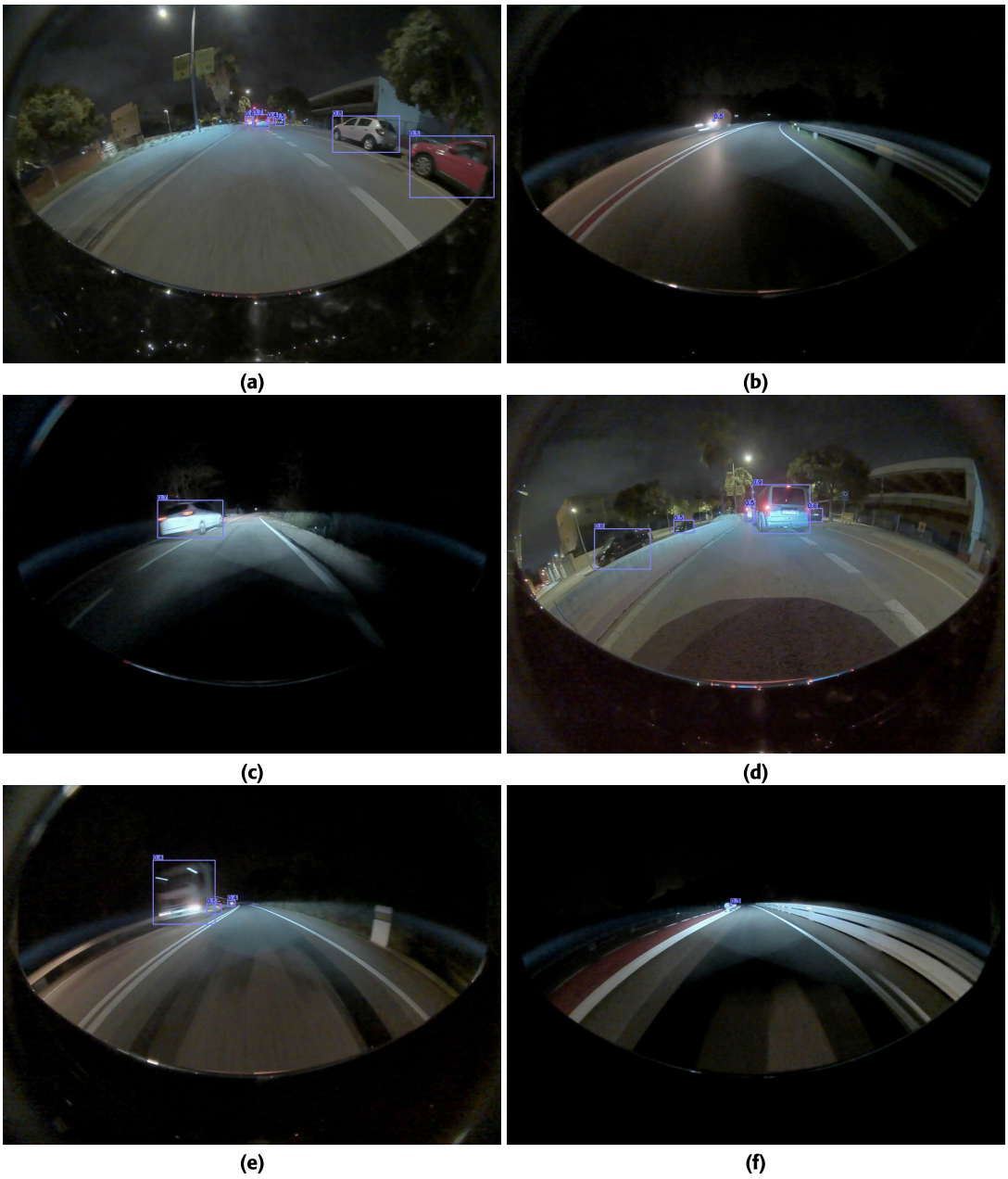


FIGURE 12. Example of the performance of the proposed system trained with the combined database training set and tested with the FNTVD dataset.

TABLE 3. Comparison with state-of-the-art works for the BDD100K, PVDN and combined datasets.

	mAP			Frames Per Second (fps)
	BDD100K	PVDN	Combined	
Proposed system	0.7134	**0.6621**	**0.6814**	45.45
Faster-RCNN [56]	0.7410	N/A	N/A	50.00
RetinaNet [57]	0.7020	N/A	N/A	56.50
Yolo-F [58]	0.6380	N/A	N/A	**86.95**
VarifocalNet [59]	**0.7630**	N/A	N/A	47.60
CornerNet [50]	0.3330	N/A	N/A	10.80

(BDD100K and PVDN). The results illustrate the system's ability to handle partial occlusions and lighting variability effectively, though some instances, as seen in Fig. 12.e, expose challenges in balancing detections of large and

small vehicles, occasionally resulting in duplicate detections. Furthermore, as illustrated in Fig. 12.f, strong light reflections under extremely low-light conditions can occasionally be misinterpreted as vehicles, presenting an additional challenge in these complex environments. These examples highlight the generalization capabilities of the model in real-world nighttime scenarios despite the inherent differences between the public datasets and the FNTVD dataset.

Fisheye cameras, while offering a wide field of view (190° in the FNTVD dataset), pose significant challenges in real-world Advanced Driver Assistance Systems (ADAS) due to inherent geometric distortions, low illumination in poor lighting conditions, and complex light patterns like glare. These distortions, particularly near the edges of the image, complicate object detection and distance estimation, while low lighting scenes further degrade image clarity. To address these issues, the proposed system employs a dual-annotation approach for robust detection in low light (using point-based annotations when vehicle shapes are unclear) and a learnable non-maxima suppression mechanism to filter false positives caused by glare or reflections. Specifically point-based annotations and detections are more robust to geometrical distortions than bounding boxes (since the size information, heavily affected by those distortions, is discarded). The visual examples in Fig. 12 demonstrate the system's ability to detect vehicles in these challenging conditions, highlighting its effectiveness in real-world scenarios. Nonetheless, future work could focus on improving distortion correction algorithms and optimizing computational efficiency to further enhance the utility of fisheye cameras in ADAS applications.

3) COMPARISON WITH THE STATE OF THE ART

The performance of the proposed system is compared with state-of-the-art methods on the BDD100K, PVDN and combined datasets, as shown in Table 3. For the BDD100K, the results are based on the full dataset, which includes both daytime and nighttime images, providing a comprehensive evaluation of the system's performance across different lighting conditions. This comparison highlights the performance of various detection methods on the BDD100K dataset, while also emphasizing the proposed system's unique capability to handle diverse scenarios including challenging nighttime conditions.

The comparison methods selected for this study include some of the most well-known and widely used state-of-the-art (SOTA) algorithms in object detection, as well as some of the latest approaches that have shown promising results in recent literature. Specifically, we compare the proposed system with Faster R-CNN [56], RetinaNet [57], Yolo-F [58], VarifocalNet [59], and CornerNet [50]. These methods were chosen because they represent a diverse range of approaches, including two-stage detectors (e.g., Faster R-CNN), single-stage detectors (e.g., RetinaNet, Yolo-F), and anchor-free detectors (e.g., VarifocalNet, CornerNet). This diversity ensures a comprehensive evaluation of the proposed system's performance relative to the current SOTA.

To ensure a fair and comprehensive comparison, all comparison methods were carefully tuned to their best performance by following the authors' recommendations for hyperparameter tuning, including the learning rate and anchor box configurations, and were evaluated under the same conditions (e.g., dataset splits, evaluation metrics), highlighting the strengths and limitations of the proposed system relative to the state-of-the-art.

While certain methods like VarifocalNet and Faster R-CNN achieve slightly higher mAP scores on the BDD100K dataset (0.7630 and 0.7410 respectively), it is important to note that these methods are optimized for well-lit urban environments and rely heavily on bounding-box annotations. As a result, they are not designed to generalize across different environments or handle point-based ground-truth annotations, which are essential for detecting vehicles in low-light rural scenarios where only flashes of the passing vehicle lights are visible. This limitation restricts their applicability to more conventional urban datasets and prevents them from performing effectively in challenging nighttime conditions.

In contrast, the proposed system achieves a competitive mAP of 0.7134 on the BDD100K dataset, demonstrating its ability to perform well in both daytime and nighttime urban scenarios. More importantly, the system maintains steady performance across the PVDN and combined datasets, with mAP scores of 0.6621 and 0.6814, respectively. The PVDN dataset, which features rural nighttime scenarios with sparse lighting, poses significant challenges for traditional object detection methods due to the lack of visible vehicle shapes and the reliance on point-based annotations. The proposed system's ability to seamlessly switch between bounding-box and point-based predictions based on visibility conditions is a key factor in its success. This adaptability ensures robust performance across diverse environments without the need for retraining or adjustments, a capability that is not achievable by the other methods analyzed.

The inability of state-of-the-art algorithms to provide meaningful results on the PVDN dataset is particularly noteworthy. Methods like Faster R-CNN, RetinaNet, and VarifocalNet fail to learn meaningful predictions on this dataset because they are designed exclusively for bounding-box annotations. Even when point-based annotations are artificially treated as 1×1 bounding boxes, these methods struggle to generalize to the unique challenges of rural nighttime scenarios. This limitation directly impacts their capacity to perform well on the combined dataset, as incorporating such datasets into their training would likely degrade their performance on the BDD100K dataset without adding meaningful improvements in low-light conditions. Consequently, obtaining reliable metrics for these algorithms on the combined dataset is not feasible, further emphasizing their restricted applicability.

In terms of computational efficiency, the proposed system achieves a frame rate of 45.45 FPS, which is sufficient for real-time applications such as autonomous driving. While Yolo-F achieves a higher frame rate (86.95 FPS), this comes at the cost of a significantly lower mAP of 0.6380 on the

BDD100K dataset, indicating a trade-off between detection accuracy and speed. The proposed system strikes a balance between these two factors, prioritizing steady performance across diverse conditions over absolute speed. This makes it particularly suitable for real-world applications where reliable detection in varying lighting conditions is more critical than achieving the maximum possible frame rate. Additionally, while methods like RetinaNet and VarifocalNet also achieve real-time speeds (56.50 FPS and 47.60 FPS, respectively), their inability to handle multiple annotation types and lighting conditions limits their practical application in complex real-world scenarios.

Overall, the results in Table 3 highlight the proposed system's unique advantages, including its robustness to varying lighting conditions, which allows it to perform well across both well-lit urban environments and challenging rural nighttime scenarios. Additionally, the system demonstrates adaptability to multiple annotation types, seamlessly handling both bounding-box and point-based annotations to ensure accurate detection regardless of visibility conditions. Furthermore, the system achieves real-time performance with a frame rate of 45.45 FPS, making it suitable for real-world applications such as autonomous driving and advanced driver-assistance systems (ADAS). These strengths collectively make the proposed system a practical and reliable solution for complex vehicle detection tasks in diverse environments.

4) SYSTEM LIMITATIONS & FUTURE WORKS

While the proposed system demonstrates strong performance across diverse nighttime environments, certain limitations remain. The system achieves 45.45 FPS, sufficient for real-time applications but lower than some state-of-the-art methods due to the Hourglass backbone's computational complexity. Future work could explore lightweight alternatives or model compression to improve efficiency without sacrificing accuracy. Second, distortions caused by fisheye cameras in the FNTVD dataset slightly affect detection accuracy, indicating the need for further refinement to better handle wide field-of-view images. Third, while designed to meet the requirements of embedded platforms like the TDA4VM, further optimization of memory usage and power consumption could improve deployability in resource-constrained automotive environments. Additionally, future work could focus on expanding the system to include other road users, such as pedestrians and cyclists, especially in nighttime conditions. Finally, future lines could address the challenge of tracking excessively small objects near the vanishing point, enhancing the system's ability to detect and monitor distant road users in complex nighttime scenarios.

IV. CONCLUSION

This paper presents a deep learning-based vehicle detection system designed for nighttime scenarios, capable of handling both well-lit urban environments and poorly lit rural areas. The system leverages a dual-annotation approach, using bounding boxes in well-lit conditions and point-based annotations in low-light scenarios, ensuring accurate detection across diverse environments. Built on an Hourglass network backbone, the system incorporates a detection block that predicts vehicle presence through heatmaps and refines predictions using a learnable non-maxima suppression mechanism. The Hourglass architecture, chosen for fine-grained feature capture in extreme lighting, is computationally intensive, limiting frame rates (45.45 FPS). The anchor-free design reduces memory usage, enabling deployment on embedded platforms like the TDA4VM, balancing detection robustness with computational efficiency. Evaluated on the BDD100K, PVDN, and combined datasets, the system achieves competitive mAP scores of 0.7134, 0.6621, and 0.6814. Compared to state-of-the-art methods, the proposed system demonstrates superior adaptability to varying lighting conditions and annotation types, making it a robust solution for real-world nighttime vehicle detection in autonomous driving applications.

REFERENCES

[1] A. Haydari and Y. Yilmaz, "Deep reinforcement learning for intelligent transportation systems: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 1, pp. 11–32, Jan. 2022.

[2] E. Marti, M. A. de Miguel, F. Garcia, and J. Perez, "A review of sensor technologies for perception in automated driving," *IEEE Intell. Transp. Syst. Mag.*, vol. 11, no. 4, pp. 94–108, Winter. 2019.

[3] Z. Lv, R. Lou, and A. K. Singh, "AI empowered communication systems for intelligent transportation systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 7, pp. 4579–4587, Jul. 2021.

[4] L. Zhu, F. R. Yu, Y. Wang, B. Ning, and T. Tang, "Big data analytics in intelligent transportation systems: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 1, pp. 383–398, Jan. 2019.

[5] M. De Ryck, M. Versteyhe, and F. Debrouwere, "Automated guided vehicle systems, state-of-the-art control algorithms and techniques," *J. Manuf. Syst.*, vol. 54, pp. 152–173, Jan. 2020.

[6] S. Sun, A. P. Petropulu, and H. V. Poor, "MIMO radar for advanced driver-assistance systems and autonomous driving: Advantages and challenges," *IEEE Signal Process. Mag.*, vol. 37, no. 4, pp. 98–117, Jul. 2020.

[7] G. Li, Y. Yang, T. Zhang, X. Qu, D. Cao, B. Cheng, and K. Li, "Risk assessment based collision avoidance decision-making for autonomous vehicles in multi-scenarios," *Transp. Res. C, Emerg. Technol.*, vol. 122, Jan. 2021, Art. no. 102820.

[8] A. M. Herrero, C. Pérez-Benito, D. P. Salido, J. S. Gual, and A. Jevtić, "Pedestrian movement prediction based on camera vision and deep learning model," in *Proc. IEEE 26th Int. Conf. Intell. Transp. Syst. (ITSC)*, Sep. 2023, pp. 2233–2238.

[9] S. Ahmed, M. N. Huda, S. Rajbhandari, C. Saha, M. Elshaw, and S. Kanarachos, "Pedestrian and cyclist detection and intent estimation for autonomous vehicles: A survey," *Appl. Sci.*, vol. 9, no. 11, p. 2335, Jun. 2019.

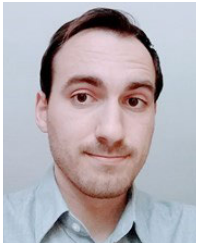
[10] G. Delgado, M. Garcia, M. Nieto, J. A. Ì. de Gordoa, C. Perez, D. Pujol, and A. Jevtic, "Virtual validation of a multi-object tracker with intercamera tracking for automotive fisheye based surround view systems," in *Proc. IEEE 14th Image, Video, Multidimensional Signal Process. Workshop (IVMSP)*, Jun. 2022, pp. 1–5.

[11] X. Hu, X. Xu, Y. Xiao, H. Chen, S. He, J. Qin, and P.-A. Heng, "SINet: A scale-insensitive convolutional neural network for fast vehicle detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 3, pp. 1010–1019, Mar. 2019.

[12] H. Wang, X. Lou, Y. Cai, Y. Li, and L. Chen, "Real-time vehicle detection algorithm based on vision and LiDAR point cloud fusion," *J. Sensors*, vol. 2019, pp. 1–9, Apr. 2019.

[13] A. P. Sligar, "Machine learning-based radar perception for autonomous vehicles using full physics simulation," *IEEE Access*, vol. 8, pp. 51470–51476, 2020.
[14] O. Elharrouss, N. Almaadeed, and S. A. Maadeed, "A review of video surveillance systems," *J. Vis. Commun. Image Represent.*, vol. 77, pp. 103–116, Apr. 2021.
[15] Z. Charouh, A. Ezzouhri, M. Ghogho, and Z. Guennoun, "A resource-efficient CNN-based method for moving vehicle detection," *Sensors*, vol. 22, no. 3, p. 1193, Feb. 2022.
[16] Y. Koga, H. Miyazaki, and R. Shibasaki, "A method for vehicle detection in high-resolution satellite images that uses a region-based object detector and unsupervised domain adaptation," *Remote Sens.*, vol. 12, no. 3, p. 575, Feb. 2020.
[17] A. M. Ali, W. I. Eltarhouni, and K. A. Bozed, "On-road vehicle detection using support vector machine and decision tree classifications," in *Proc. 6th Int. Conf. Eng. MIS*, Sep. 2020, pp. 1–5.
[18] J. Wang, S. Simeonova, and M. Shahbazi, "Orientation- and scale-invariant multi-vehicle detection and tracking from unmanned aerial videos," *Remote Sens.*, vol. 11, no. 18, p. 2155, Sep. 2019.
[19] W. Zhou, C. Wang, J. Xia, Z. Qian, and Y. Wu, "Monitoring-based traffic participant detection in urban mixed traffic: A novel dataset and a tailored detector," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 1, pp. 189–202, Jan. 2024.
[20] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, "BDD100K: A diverse driving dataset for heterogeneous multitask learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2633–2642.
[21] A. I. Maqueda, C. R. del-Blanco, F. Jaureguizar, and N. García, "Structured learning via convolutional neural networks for vehicle detection," *Proc. SPIE*, vol. 10223, May 2017, Art. no. 1022302.
[22] A. Bell, T. Mantecón, C. Díaz, C. R. del-Blanco, F. Jaureguizar, and N. García, "A novel system for nighttime vehicle detection based on foveal classifiers with real-time performance," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 6, pp. 5421–5433, Jun. 2022.
[23] J. Wu, H. Xu, Y. Tian, R. Pi, and R. Yue, "Vehicle detection under adverse weather from roadside LiDAR data," *Sensors*, vol. 20, no. 12, p. 3433, Jun. 2020.
[24] Q. A. Al-Haija, M. Gharaibeh, and A. Odeh, "Detection in adverse weather conditions for autonomous vehicles via deep learning," *AI*, vol. 3, no. 2, pp. 303–317, Apr. 2022.
[25] M. Hassaballah, M. A. Kenk, K. Muhammad, and S. Minaee, "Vehicle detection and tracking in adverse weather using a deep learning framework," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 7, pp. 4230–4242, Jul. 2021.
[26] S. Kumar, S. C. Sharma, and R. Kumar, "Wireless sensor network based real-time pedestrian detection and classification for intelligent transportation system," *Int. J. Math., Eng. Manage. Sci.*, vol. 8, no. 2, pp. 194–212, Apr. 2023.
[27] D.-T. Iancu, A. Sorici, and A. M. Florea, "Object detection in autonomous driving–from large to small datasets," in *Proc. 11th Int. Conf. Electron., Comput. Artif. Intell. (ECAI)*, Jun. 2019, pp. 1–6.
[28] G. N. Swamy and S. Srilekha, "Vehicle detection and counting based on color space model," in *Proc. Int. Conf. Commun. Signal Process. (ICCSP)*, Apr. 2015, pp. 0447–0450.
[29] X. Yang and Y. Yang, "A method of efficient vehicle detection based on hog-lbp," *Comput. Eng.*, vol. 40, no. 9, pp. 210–214, 2014.
[30] R. K. Satzoda and M. M. Trivedi, "Multipart vehicle detection using symmetry-derived analysis and active learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 4, pp. 926–937, Apr. 2016.
[31] S. A. Nur, M. M. Ibrahim, N. M. Ali, and F. I. Y. Nur, "Vehicle detection based on underneath vehicle shadow using edge features," in *Proc. 6th IEEE Int. Conf. Control Syst., Comput. Eng. (ICCSCE)*, Nov. 2016, pp. 407–412.
[32] X. Zhang, B. Story, and D. Rajan, "Night time vehicle detection and tracking by fusing vehicle parts from multiple cameras," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 8136–8156, Jul. 2022.
[33] W. Tsai and H. Chen, "High-accuracy vehicle lamp detection for real-time night-time traffic surveillance," *IET Intell. Transp. Syst.*, vol. 14, no. 13, pp. 1923–1934, Dec. 2020.
[34] T. Nakane, T. Takeshita, S. Tokai, and C. Zhang, "Vehicle rear-lamp detection at nighttime via probabilistic bitwise genetic algorithm," in *Proc. Int. Conf. Cyberworlds (CW)*, Oct. 2019, pp. 117–120.
[35] R. K. Satzoda and M. M. Trivedi, "Looking at vehicles in the night: Detection and dynamics of rear lights," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 12, pp. 4297–4307, Dec. 2019.
[36] T.-A. Vu, L. H. Pham, T. K. Huynh, and S. V. Ha, "Nighttime vehicle detection and classification via headlights trajectories matching," in *Proc. Int. Conf. Syst. Sci. Eng. (ICSSE)*, Jul. 2017, pp. 221–225.
[37] C. Ma and F. Xue, "A review of vehicle detection methods based on computer vision," *J. Intell. Connected Vehicles*, vol. 7, no. 1, pp. 1–18, Mar. 2024.
[38] Y. Guo and M. Zhao, "Nighttime vehicle detection on highway based on improved faster R-CNN model," in *Proc. 2nd Int. Conf. Comput. Vis., Image, Deep Learn.*, Oct. 2021, pp. 168–174.
[39] Z. Chen, H. Guo, J. Yang, H. Jiao, Z. Feng, L. Chen, and T. Gao, "Fast vehicle detection algorithm in traffic scene based on improved SSD," *Measurement*, vol. 201, Sep. 2022, Art. no. 111655.
[40] V. O. O. Castelló, I. S. S. Igual, O. del Tejo Catalá, and J.-C. Perez-Cortes, "High-profile VRU detection on resource-constrained hardware using YOLOv3/v4 on BDD100K," *J. Imag.*, vol. 6, no. 12, p. 142, Dec. 2020.
[41] B. Xiao, J. Guo, and Z. He, "Real-time object detection algorithm of autonomous vehicles based on improved YOLOv5s," in *Proc. 5th CAA Int. Conf. Veh. Control Intell. (CVCI)*, Oct. 2021, pp. 1–6.
[42] J. Azimjonov, A. Özmen, and T. Kim, "A nighttime highway traffic flow monitoring system using vision-based vehicle detection and tracking," *Soft Comput.*, vol. 27, no. 19, pp. 13843–13859, Oct. 2023.
[43] K. Zhu, H. Lyu, and Y. Qin, "Enhanced detection of small and occluded road vehicle targets using improved YOLOv5," *Signal, Image Video Process.*, vol. 19, no. 2, pp. 1–15, Feb. 2025.
[44] J. Guo, M. Shao, X. Chen, Y. Yang, and E. Sun, "Research on nighttime vehicle target detection based on improved KSC-YOLO V5," *Signal, Image Video Process.*, vol. 19, no. 1, pp. 1–11, Jan. 2025.
[45] M. A. A. Khan, H. Ma, A. Farhad, A. Mujeeb, I. K. Mirani, and M. Hamza, "When LORA meets distributed machine learning to optimize the network connectivity for green and intelligent transportation system," *Green Energy Intell. Transp.*, vol. 3, no. 3, Jun. 2024, Art. no. 100204.
[46] R. M. Savithramma and R. Sumathi, "Intelligent traffic signal controller for heterogeneous traffic using reinforcement learning," *Green Energy Intell. Transp.*, vol. 2, no. 6, Dec. 2023, Art. no. 100124.
[47] L. Ohnemus, L. Ewecker, E. Asan, S. Roos, S. Isele, J. Ketterer, L. Müller, and S. Saralajew, "Provident vehicle detection at night: The PVDN dataset," 2020, *arXiv:2012.15376*.
[48] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," in *Proc. Eur. Conf. Comput. Vis.*, Jan. 2016, pp. 483–499.
[49] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," 2019, *arXiv:1904.07850*.
[50] H. Law and J. Deng, "CornerNet: Detecting objects as paired keypoints," in *Proc. Eur. Conf. Comput. Vis.*, Jan. 2018, pp. 765–781.
[51] J. Tompson, A. Jain, Y. LeCun, and C. Bregler, "Joint training of a convolutional network and a graphical model for human pose estimation," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, Dec. 2014, pp. 1799–1807.
[52] J. Tompson, R. Goroshin, A. Jain, Y. LeCun, and C. Bregler, "Efficient object localization using convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 648–656.
[53] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020.
[54] S. Ruder, "An overview of gradient descent optimization algorithms," 2016, *arXiv:1609.04747*.
[55] I. Urbieta, A. Mujika, G. Piérola, E. Irigoyen, M. Nieto, E. Loyo, and N. Aginako, "WebLabel: OpenLABEL-compliant multi-sensor labelling," *Multimedia Tools Appl.*, vol. 83, no. 9, pp. 26505–26524, Aug. 2023.
[56] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
[57] T.-Y. Ross and G. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, 2017, pp. 2980–2988.
[58] Q. Chen, Y. Wang, T. Yang, X. Zhang, J. Cheng, and J. Sun, "You only look one-level feature," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13034–13043.
[59] H. Zhang, Y. Wang, F. Dayoub, and N. Sünderhauf, "VarifocalNet: An IoU-aware dense object detector," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 8510–8519.

LEYRE ENCÍO received the Bachelor of Engineering degree in telecommunication technologies from the Universidad Pública de Navarra (UPNA), in 2020, and the master's degree in telecommunication engineering and the master's degree in signal theory and communications from the Universidad Politécnica de Madrid (UPM), in 2023. In 2022, she became a member of the Grupo de Tratamiento de Imágenes (Image Processing Group), UPM, where she has been involved in several research projects. Her research interests include artificial intelligence, deep learning, computer vision, image processing, machine learning, and reinforcement learning.



DANIEL FUERTES received the Bachelor of Engineering degree in telecommunication technologies and services and the master's degree in signal theory and communications from the Universidad Politécnica de Madrid (UPM), Madrid, Spain, in 2019 and 2020, respectively. Since 2020, he has been a member of the Grupo de Tratamiento de Imágenes (Image Processing Group), UPM, where he has been actively involved in several research projects. His research interests include the areas of artificial intelligence, deep learning, computer vision, machine learning, reinforcement learning, and combinatorial optimization.

CARLOS R. DEL-BLANCO received the degree in telecommunication engineering and the Ph.D. degree in telecommunication from the Universidad Politécnica de Madrid (UPM), in 2005 and 2011, respectively. Since 2005, he has been a member of the Image Processing Group, UPM. Since 2011, he has been a member of the Faculty of the ETS Ingenieros de Telecomunicación. Since 2021, he has been a Professor of signal theory and communications with the Department of Signals, Systems, and Communications. His professional interests include signal and image processing, computer vision, pattern recognition, machine learning, and stochastic dynamic models. He has been actively involved in European projects and national projects in Spain.

IU AGUILAR received the degree in telecommunications engineering from the Universitat Politècnica de Catalunya (UPC), in 2022. He is currently a Vision Systems Engineer with Ficosa, focusing on the development and integration of vision-based algorithms for advanced driver-assistance systems (ADAS). His research interests include computer vision, artificial intelligence, and communications, given their applications to autonomous systems.

CRISTINA PÉREZ-BENITO received the degree in mathematics from the Universidad de Salamanca (USAL), in 2014, and the master's degree in mathematical research and the Ph.D. degree in mathematics from the Universidad Politécnica de Valencia (UPM), in 2015 and 2019, respectively. She is currently a Tech Lead with Ficosa ADAS, leading the research projects focused on vision systems for ADAS. Her research interests include image processing, computer vision, and autonomous driving.

ALEKSANDAR JEVTIĆ received the M.S. degree in electrical engineering from the University of Belgrade, Serbia, in 2005, and the Ph.D. degree in computer science from the Universidad Politécnica de Madrid (UPM), Spain, in 2011. He is currently the Head of research with Ficosa ADAS, an automotive Tier-1 company focusing on developments of ADAS and autonomous driving. His main research interest includes AI and its applications to computer vision and autonomous systems.

FERNANDO JAUREGUIZAR received the degree in telecommunication engineering (six years engineering program) and the Ph.D. degree in telecommunication engineering from the Universidad Politécnica de Madrid (UPM), in 1987 and 1994, respectively. Since 1987, he has been a member of the Image Processing Group, UPM. Since 1991, he has been a member of the Faculty of UPM. Since 1995, he has also been an Associate Professor of signal theory and communications with the Department of Signals, Systems, and Communications. He has been actively involved in European Projects (Eureka, ACTS, IST, ITEA, and EIT-RM) and national projects in Spain. His professional interests include digital image processing, video coding, 3-DTV, computer vision, and design and development of multimedia communications systems.

NARCISO GARCÍA (Life Senior Member, IEEE) received the Ingeniero de Telecomunicación degree and the Ph.D. degree in communications from the Universidad Politécnica de Madrid (UPM), Madrid, Spain, in 1976 and 1983, respectively. Since 1977, he has been a member of the Faculty of UPM, where he is currently a Professor of signal theory and communications. He leads the Grupo de Tratamiento de Imágenes (Image Processing Group), UPM. He has been actively involved in Spanish and European research projects, also serving as an evaluator, a reviewer, an auditor, and an observer for several research and development programs of the European Union. He was a Co-Writer of the EBU proposal, base of the ITU Standard for digital transmission of TV at 34–45 Mb/s (ITU-T J.81). He was an Area Coordinator of the Spanish Evaluation Agency (ANEP), from 1990 to 1992. He was the General Coordinator of the Spanish Commission for the Evaluation of the Research Activity (CNEAI), from 2011 to 2014. He has been the Vice-Rector for International Relations of UPM, from 2014 to 2016. His current research interests include digital video compression, computer vision, and quality of experience. He was a recipient of the Junior and Senior Research Awards of UPM, in 1987 and 1994, respectively; the Spanish National Graduation Award; and the Doctoral Graduation Award.