ProvRain: Rain-Adaptive Denoising and Vehicle Detection via MobileNet-UNet and Faster R-CNN

Aswinkumar Varathakumaran
School of Computer Science Engineering
Vellore Institute of Technology
Chennai, India
aswinkumar.v9805@gmail.com

Nirmala Paramanandham
School Of Electronics Engineering
Vellore Institute of Technology
Chennai, India
nirmala.p@vit.ac.in

***Abstract*—Provident vehicle detection has a lot of scope in the detection of vehicle during night time. The extraction of features other than the headlamps of vehicles allows us to detect oncoming vehicles before they appear directly on the camera. However, it faces multiple issues especially in the field of night vision, where a lot of noise caused due to weather conditions such as rain or snow as well as camera conditions. This paper focuses on creating a pipeline aimed at dealing with such noise while at the same time maintaining the accuracy of provident vehicular detection. The pipeline in this paper, ProvRain, uses a lightweight MobileNet-U-Net architecture tuned to generalize to robust weather conditions by using the concept of curriculum training. A mix of synthetic as well as available data from the PVDN dataset is used for this. This pipeline is compared to the base Faster RCNN architecture trained on the PVDN dataset to see how much the addition of a denoising architecture helps increase the detection model's performance in rainy conditions. The system boasts an 8.94% increase in accuracy and a 10.25% increase in recall in the detection of vehicles in rainy night time frames. Similarly, the custom MobileNet-U-Net architecture that was trained also shows a 10-15% improvement in PSNR, a ∼5-6% increase in SSIM, and upto a 67% reduction in perceptual error (LPIPS) compared to other transformer approaches.**

***Index Terms*—provident vehicle detection dataset (PVDN), curriculum training, denoising MobileNet-U-Net, night time imagery, rain noise**

I. Introduction

Night driving tends to be vastly more treacherous due to low light levels, glare from oncoming headlights, and inclement weather that impairs visibility as well. Daytime driving differs from nighttime driving because while vehicles can often be identified based on the shape, color, and context of the road, at night drivers primarily rely on very subtle cues that the vehicle is approaching. For example, low illumination might provide limited time to react once a driver perceives dim outlines of headlights far away, but, with distractions, divided attention, and delay in recognition, those few seconds may not be enough to avert catastrophe. Advanced driver-assistance systems (ADAS) and autonomous vehicles (AVs) attempt to mimic and improve upon this human ability, by sensing these cues of oncoming vehicles and warning the driver much earlier than any human might be able to react. In order to examine the issue systematically, the Provident Vehicle Detection at Night (PVDN) dataset [1] was released, which works differently from a typical object detection dataset in that it focuses on driving at night, and utilizes grayscale scenes with detailed labels for oncoming vehicles, bright light sources, and reflective artifacts. The PVDN dataset aims to introduce early detection from light cues instead of relying on full vehicle identity, which ultimately makes the dataset serve as a benchmark that is more aligned with the way human drivers evaluate threats under low-light driving conditions.

Existing datasets hold great promise, but the accompanying pipelines tend to ignore the effects of severe weather. Rain in particular poses distinct challenges for improved low-light driving and detection of contrast features, as it creates a visual clutter with streaking and scattering, introduces distortion through raindrops on the lens, and could obscure weak light patterns due to the combination of the motion of vehicles with the rain. Each of these elements can distort the small, yet critical cues that early detection relies on. Though there exist general rain removal, and image enhancement techniques, they have rarely been tailored for this novel challenge.

The proposed solution in this paper, ProvRain, will help to address this use-case with a lightweight denoising module added to the keypoint detection pipeline while considering the computational limitation of automotive-capable hardware. Nevertheless, beyond pre-processing, a weather-aware curriculum training approach is adopted, where models are first trained on clean nighttime images and are then exposed to increasing synthetically and annotatively severe rainy conditions during training. This curriculum approach will help model robustness, while maintaining detection models that are capable of discriminating small and critical cues such as the reflection of headlights during heavy rain. The ultimate aim is to achieve provident detection like a human - being able to identify hazards sooner, under the same severe conditions that make nighttime driving overall much more hazardous.

II. Related Works

Research surrounding nighttime driving safety has gained momentum in recent years along with the increasing recognition of how much drivers are at an increased risk when driving in dark and/or adverse conditions. To reiterate, although drivers rely on the subtle hints, like the insignificantly faint light on the horizon or slight reflections off of wet asphalt, machine perception systems must also be trained to observe and process these weak signals under unfavorable conditions. And thus, the datasets and architectures developed specifically designed to work in dark, nighttime driving situations have evolved not to apply to broad or general detection during the daytime only, but also anticipate the presence of oncoming vehicles.

The Provident Vehicle Detection at Night (PVDN) dataset [1] has gone on to become the predominantly used benchmark for this task. While causal detection has been demonstrated at scale in recent datasets the PVDN captures the environment of nocturnal driving scenarios and emphasizes the explicit annotation of headlights, taillights and reflective artifacts in driving scenarios. All of the architectures evaluated on PVDN by Ivarsson and Zacke (2023) [2] bare a comparison of six different deep learning models. Convolutional neural networks exhibited a relative overall solid ability in detecting other vehicles and consistently outperformed transformer-based detection by a significant margin, with DenseNet recording the highest accuracy 88.10% accuracy and avg F1 score of 0.83. These examples suggest, that although many researchers have been moving towards transformer architectures for deep learning, the Convolutional neural network still appears to continue to limit its effectiveness to anticipating a localised light pattern in an intrinsically dark, high-noise condition.

Nonetheless, noise and weather continue to be a potentially challenging aspect. Surveys and advances in image deraining [3] indicate the necessity of pre-processing for this type of data. A paper using guided filters showed there was a 12.19% increase in vehicle counts when using pre-processing under heavy rain [4], but this experiment did not truly study the maintenance of light cues at the lower end of the dynamic range. Motion has been examined under video settings, where FastDVDnet provided real-time denoising while maintaining temporal coherence if any small light artifacts were to be tracked [5]. BM3D and other versions indexed to mean Gaussian-threshold [6] still remain as robust single-frame denoising approaches in low-light conditions. Transformer based models have also gained increasing traction in this concept. SwinIR [7] is an image denoiser based on a CNN transformer architecture with windowed self-attention mechanisms. Similarly, Restormer [8] is a transformer based architecture with deep learning which uses channel-wise transposed attention and gated depthwise convolutions to remove noise effectively from images. Nevertheless, pre-processing methods still fail to capture weather corruptions when the weather differs from training data. Robustness studies [9], [10] repeatedly demonstrate how models are under-performing when faced with previously unseen corruptions. Theoretical frameworks such as curriculum learning have been proposed to address this type of problem, where models can begin with easy tasks and progress towards corrupted or complicated versions. Synthetic data generation has shown promise for addressing dataset scarcity in nighttime vehicle detection for this. Synthetic nighttime data generated with CARLA was utilized in tandem with Efficient Attention GAN (EAGAN) for day-to-night style transfer and producing realistic headlight modeling compared to previously failed GAN approaches [11]. This labeling-free augmentation framework, based on mapping daytime annotations directly to style-transferred nighttime images, led to significant improvements in model performance and significantly higher confidence scores than the baseline models trained on daytime only images. Curriculum augmentation and curriculum smoothing [12] have successfully aided the performance of vision models in condition corruptions, but they have not been explored in regards to performance to providential detection.

In conclusion, most of the existing PVDN pipelines adopt a staged runtime consisting of tone mapping or CLAHE preprocessing, lightweight denoising, light-proposal generation (RetinaNet or saliency-based), compact classification, plausibility checks, and temporal tracking. Prior work [13] shows this modular structure can be effective, but other than showing limited focus on weather noise at low-visibility, there is little prior work on this step. This paper continues this line of work by providing a rainfall-aware denoising step as well as while employing a curriculum parcel-driven training step so that reliability is maintained in the very cases early detection is most critical.

III. Methodology

ProvRain enhances existing provident detection pipelines in two significant ways: (i) weather-aware denoising included in the perception stack, and (ii) a tiered architecture designed to maximize performance in the presence of low-light and adverse-weather conditions, designed specifically to ensure that systems on automotive-grade hardware which may have constraints on computing resources can continue to deploy the system without sacrificing detection fidelity whether it be rain or at night.

A. Data Preprocessing

The imagery captured under nighttime driving conditions is characterised by significant degradation due to sensor noise from high ISO settings, strong headlight contrasts, and environmental effects such as rain streaks, water reflections, and haze. In order to deal with this problem, as a first step of the pipeline, pre-processing is proposed to restore visibility before denoising and detection.

The PVDN dataset consists of day and night scenes, with one scene corresponding to a continuous video clip captured by a forward-facing camera on a moving vehicle. Hence, frames in one scene have similar lighting and weather conditions, so it is possible to categorize scenes. Night scenes were manually labeled as either clear or rainy. Only clear nighttime scenes were used for training and validation, and actual rainy nighttime scenes were used only for testing to determine generalization. All frames were resized to the size of the input resolution of the backbone and intensity-normalized to the range $(0, 1)$.

Additionally, frames were tone-mapped and contrast-enhanced using gamma correction followed by Contrast Limited Adaptive Histogram Equalization (CLAHE), defined as:

$$I'_t = \mathrm{CLAHE}\big(\gamma(I_t)\big), \tag{1}$$

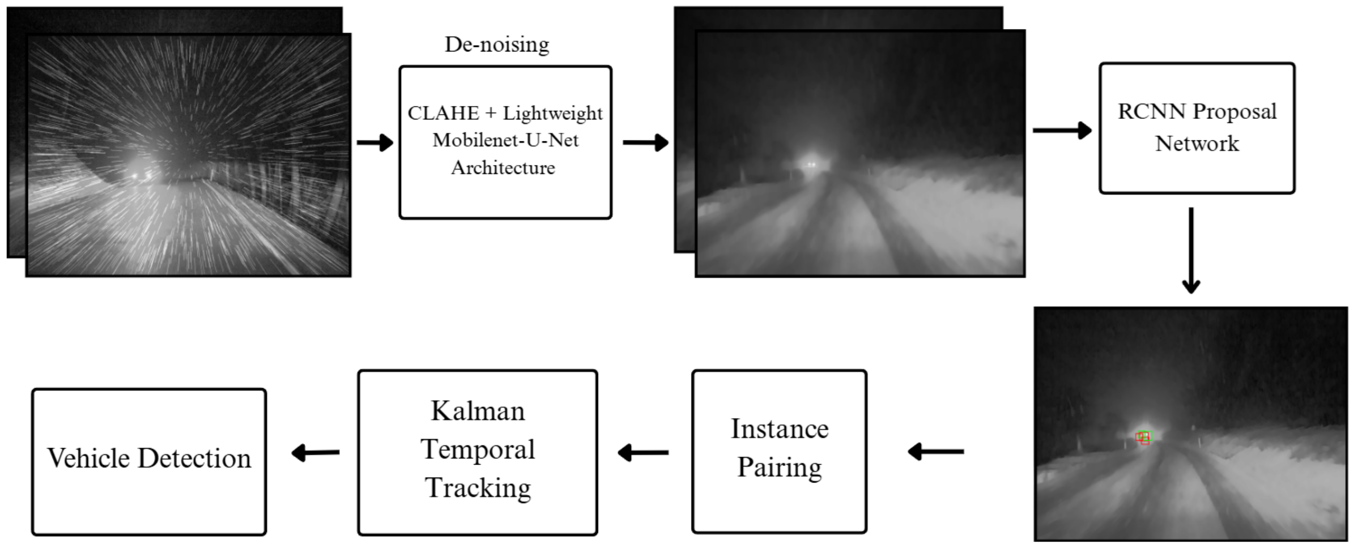


Fig. 1: ProvRain Architecture Diagram

where I_t denotes the raw input frame, $\gamma(\cdot)$ enhances mid-tone details, and CLAHE restores local contrast in dark regions without excessively amplifying noise.

To simulate the degradations caused by rain for supervised learning, rain streaks and Gaussian blur were applied to the clear nighttime image. The rain streaks were created by a gradient with a vanishing point, which originated from the image center. Since the PVDN image is captured by a moving vehicle with a front-mounted camera, this gradient formulation is a very accurate representation of the rain structure caused by perspective during actual driving.

B. Curriculum Training for Weather-Aware Denoising

The denoising network is designed utilizing a MobileNet-U-Net structure, as it is able to maintain a proper trade-off between restoration quality and efficiency, while capturing intricate light detail necessary for accurate vehicle detection as a subsequent stage of the framework. The MobileNet-U-Net is able to leverage the benefits of multi-scale fusion with the help of skip connections in the U-Net structure.

The synthesized degradation of the images made from the clear night scenes was divided into several stages of degradation intensity to aid curriculum learning. This type of learning enables the MobileNet-U-Net model to initially learn the patterns on clean as well as slightly degraded data before being trained on the severe rain artifact data.

The curriculum learning schedule used to train the denoising network is summarized in Table I.

TABLE I: Curriculum Learning Schedule for Denoising Network Training

Stage	Input Data	Noise Intensity	Description
1	Clear nighttime images	0%	Clean baseline initialization
2	Synthetic rain	25%	Light drizzle conditions
3	Synthetic rain	50%	Moderate rain degradation
4	Synthetic rain	75%	Heavy rain and blur artifacts
5	Mixed synthetic rain levels	Variable	Robustness refinement

Training is done sequentially over the stages of the curriculum. Weights trained at stage k are used as the initialization for training at the next stage $k+1$, while the architecture, loss, and optimization parameters are held constant. No data augmentation is done between stages. This sequential noise injection approach allows the network to first stabilize on clean and lightly corrupted images before learning to remove highly rain-corrupted regions, which are high frequency and informative for vehicle detection. Synthetic rain degradation at curriculum stage k is defined as:

$$I_t^{(k)} = J_t + R_t^{(k)}, \tag{2}$$

where J_t denotes the clean background frame and $R_t^{(k)}$ represents injected rain streaks parameterized by streak density, orientation, and opacity. The denoising network g_ϕ is trained to recover the clean image:

$$\hat{J}_t = g_\phi\big(I_t^{(k)}\big), \tag{3}$$

where $\hat{J}_t$ denotes the predicted clean image at time t.

The denoising network is optimized using a composite loss function defined as:

$$L_{\text{total}} = \lambda_{\text{mse}} L_{\text{mse}} + \lambda_{\text{percept}}\big(1 - \text{SSIM}(I_p, J_t)\big) + \lambda_{\text{mask}} L_{\text{mask}}, \tag{4}$$

where $I_p = \hat{J}_t$ is the predicted denoised image.

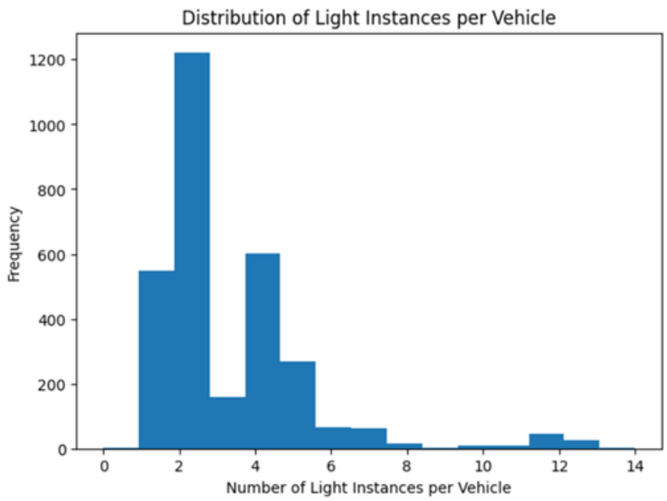


Fig. 2: Distribution of Light Instances per Vehicle

The mean squared error loss enforces pixel-wise fidelity:

$$L_{\text{mse}} = \text{MSE}(I_p, J_t), \tag{5}$$

and the mask loss computes the binary cross-entropy between the predicted rain mask M_p and the ground-truth rain mask M_t:

$$L_{\text{mask}} = \text{BCE}(M_p, M_t). \tag{6}$$

The MSE term promotes the correct reconstruction of image intensity, and the SSIM term preserves local structural consistency between light cues and is crucial for vehicle detection. The mask loss function promotes an explicit suppression of rain streaks and a reduction in false light activation in the detection module.

C. Light Proposal Classification

After post-processing, the system identifies candidate light-like areas. The system primarily identifies lights from the vehicle (and reflections) because they happen earlier and are somewhat more frequent during night time driving conditions. The number of light instances and the ratio of reflection to direct light sources for the PVDN dataset is seen in Fig 2 and Fig 3 respectively. Candidate Proposals $P = \{p_i\}$ are generated by a pruned RetinaNet trained with a singular "light-like" class and set with a very low latency,

$$P = \{\, p_i \mid S(I_t') > \tau \,\} \tag{7}$$

This allows the system to have a wide coverage of headlight, taillight, and reflection artifacts without rejecting weak signals too early.

Each proposal is then categorized into one of three: headlight, taillight, or artifact. A Faster R-CNN classifier (ResNet50) handles each region:

$$y_i = h_\psi(p_i) \tag{8}$$

$$\hat{c}_i = P(y_i \mid p_i) \tag{9}$$

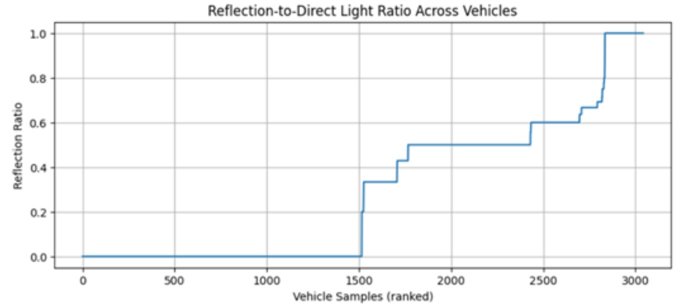


Fig. 3: Reflection-to-Direct Light Ratio Across Vehicles

Where c is the confidence score. This phase is key to eliminate false positives due to roadside lights, reflective signs, or background clutter.

D. Pairing of instances

Valid headlights typically come in the form of symmetric pairs. So, pairing rules are used.

$$\{P\}_{pairs} = \{(p_i, p_j) \mid |y_i - y_j| < \epsilon_y, d_{min} \le |x_i - x_j| \le d_{max}\} \tag{10}$$

The decision-making algorithm integrates all these cues to enable detection. A vehicle is flagged as "likely present" if it passes alignment + spacing test, lights are spreading apart or growing indicating that the vehicle is coming closer and if the classifier's confidence is higher than the required threshold. Because headlights exist across several frames, temporal consistency is also enforced through Kalman filtering:

$$\{\hat{x}\}_{t|t-1} = A\{\hat{x}\}_{t-1|t-1} + Bu_t \tag{11}$$

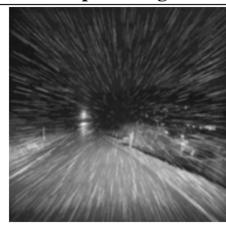
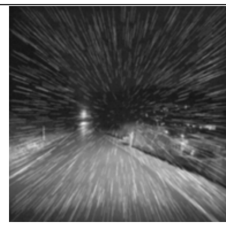
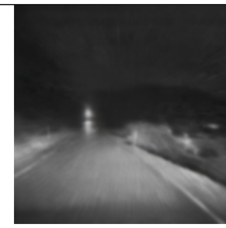
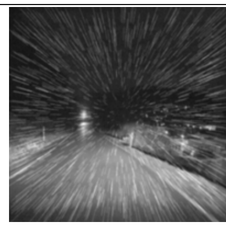
$$P_{t|t-1} = AP^T_{(t-1|t-1)}A^T + Q \tag{12}$$

with Hungarian algorithm-optimized assignments. Optionally, DeepSORT could be used to obtain richer embedding features to improve identity tracking, and to have smoother confidence curves over time. By combining geometric, temporal, and classification signals, the system balances sensitivity and robustness. The entire system architecture is seen in Fig 1.

IV. Observations

The experiments demonstrated a few useful low-light preprocessing techniques, tone-mapping, gamma correction and CLAHE, that improved visibility of vehicle lights but also wrongfully amplified background noise and reflections with the most noticeable impact coming from video captured under rainy conditions. FastDVDnet [5] originally used as a fast video denoiser retained some static low-light characteristics while denoising. It was ineffective at dynamic denoising and removal of rain streaks, and occasional produced rain streak-like patterns that concealed true light sources and provided degraded input for the downstream detection. From these results it was clear that an off-the-shelf denoiser was not viable

TABLE II: Model Inference Results Comparison

Model Inference Results	Input Image	Ground Truth	Model Output
SwinIR			
Restormer			
ProvRain (Ours)			

for the challenging night-rain context. Transformer-based approaches such as SwinIR [7] and Restormer [8] produced visually improved outputs under synthetic degradations; however, SwinIR often failed to suppress rain streaks completely, while Restormer occasionally over-smoothed high-frequency light structures critical for detection, making the frames blurred. In contrast, the proposed lightweight MobileNet-U-Net, consistently preserved salient light cues while suppressing rain streaks. Although the denoiser was supervised using synthetically generated rain due to the lack of paired clean–rainy nighttime data, qualitative inspection on real rainy nighttime scenes shows effective removal of rain streaks without compromising headlight and reflection integrity. This behavior translated into more reliable inputs for the light proposal generation and classification phases by the Faster RCNN.

It was also observed that using light proposals based on simple geometric heuristics together with temporal smoothing through Kalman filtering was effective in generating stable tracks of vehicles. Even without any explicit depth information provided, symmetry and motion coherence were enforced to filter out spurious detections from reflections and artifact noise stemming from rapidly disappearing objects. With curriculum trained custom MobileNet-U-Net removing rain noise, the system also could maintain robust tracks over time, which reflected in detections being earlier and being more trusted. All in all, moving away from generic denoisers towards task-specific lightweight MobileNet-U-Net with curriculum learning was integral to obtaining strong performance under the night-rain conditions without completely sacrificing real-time efficiency.

V. Results And Inference

The methodology in this paper is compared to the original Faster RCNN architecture used on the PVDN dataset. The custom MobileNet-U-Net denoising network, attached to the Faster RCNN pipeline was qualitatively measured against the base model on metrics including recall, accuracy, early warning success and inference time (Table II). ProvRain was trained mainly on Kaggle. The P100 GPU with 16GB VRAM was used, enabling efficient experimentation and faster convergence across platforms. The light custom MobileNet-U-Net denoiser demonstrates the greatest improvement with the suggested pipeline. The night rain images with no denoising had distracting streaks and noise that made automatic detection unstable; the light proposal generator had frequent false positive detections and the classifier reassigned mistaken reflections and rain debris to car lights. Early warnings were therefore less reliable.

When the MobileNet-U-Net denoiser was introduced, the images that were noisy with rain were more easily processed and essential light areas were identifiable. This further improved the proposal generator enabling the classifier to detect headlights and taillights with higher precision. It proved to be much more efficient than other denoising techniques including the SwinIR and Restormer which are equally performative model, however did not capture and retain image sharpness as well after removal of noise(Table II). The peak signal noise ratio (PSNR) score, structural similarity Index (SSIM) and mean absolute error (L1 Loss) were used to compare the models (Table III).

Quantitatively, the custom MobileNet-U-Net denoiser in-

TABLE III: Comparison of Denoising Techniques

Metric	SwinIR	Restormer	ProvRain Denoiser (Ours)
PSNR (dB)	31.45	32.72	36.24
SSIM	0.8874	0.8961	0.941
L1 Loss	0.0108	0.0099	0.0096
MSE	0.00072	0.00051	0.000238
RMSE	0.0268	0.0226	0.0186
MAE	0.0132	0.0120	0.0111
LPIPS	0.1623	0.1389	0.1264

TABLE IV: Inference Results of Proposed Methodology

Metric	Faster R-CNN	ProvRain (Ours)
Proposal Recall (%)	78.28	88.53
Classifier Accuracy (%)	81.74	90.68
Early-warning Success (%)	65.92	88.72
Throughput (FPS)	20	14

creased detection stability and confidence generally with little to no extra cost. By comparison of the raw images to the denoised images, it was visually verified that car lights were still visible in the denoised outputs while rain streaks and other artifacts were suppressed, for the most part. These results demonstrate that for reliable nighttime car detection in rain, task specific lightweight denoising through curriculum learning proves to be advantageous, leading to an uptick in performance from using raw images alone, while maintaining the near real-time performance of the system. As shown in Table IV, the denoising stage significantly improves input quality under rainy nighttime conditions, leading to higher proposal recall, improved classification accuracy, and substantially earlier warnings. While the average throughput decreases from 20 FPS for the baseline Faster R-CNN pipeline to 14 FPS when denoising is enabled, this tradeoff is intentional. The context of advanced driver-assistance systems, where early and reliable detection is often more critical than raw frame rate makes this throughput practical for deployment and reflects a favorable balance between robustness and efficiency.

VI. Conclusions and Future work

The methodology in this paper, ProvRain, presents a lightweight, real-time pipeline for the providential detection of vehicles in severe raining conditions at night. By combining the image preprocessing, the task-specific MobileNet-U-Net with curriculum learning, efficient light proposal generation, classification, and motion tracking, the system is capable of detecting headlights and taillights reliably whilst filtering out rain noise and rain artefacts. The experiments on the PVDN dataset illustrated that the curriculum-trained denoiser can significantly increase detection stability and track persistence, as well as the early warning success when contrasting to using the raw images.

A limitation of this work is the reliance on synthetically generated rain for supervised denoising training, which is necessitated by the lack of paired clean–rainy nighttime driving data; collecting such real-world ground truth remains an open challenge. The denoising stage also introduces additional computation and reduces throughput compared to the baseline, the pipeline maintains near real-time performance, and future work will explore optimization and model compression to further improve inference speed. Generalizing the curriculum learning framework to also be used in additional weather scenarios such as fog or snow could further increase generalizability to different night-driving conditions. Looking into adaptive confidence levels and novel pairing heuristics could lead to even less false positives, and even earlier warnings. Finally, linking the pipeline to utilise multi camera set-ups or monocular depth cues from motion could potentially allow the pipeline to perform more accurate motion estimation without increased computational cost.

References

[1] 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,” *arXiv:2012.15376*, 2020.

[2] C. Ivarsson and J. Zacke, “Night-time vehicle detection based on observable light cues using deep learning,” Student thesis, Dept. of Automatic Control, Lund Univ., Lund, Sweden, 2023.

[3] C. Shi, L. Fang, H. Wu, X. Xian, Y. Shi, and L. Lin, “NiteDR: Nighttime image de-raining with cross-view sensor cooperative learning for dynamic driving scenes,” *IEEE Trans. Multimedia*, vol. 26, pp. 9203–9215, 2024.

[4] B. Setiyono, D. Sulistyaningrum, R. Sulistyaningrum, A. Usadha, and A. Nusantara, “The rain noise reduction using guided filter to improve performance of vehicle counting,” *Int. J. Innov. Comput., Inf. Control*, vol. 16, no. 4, pp. 1353–1370, Aug. 2020, doi: 10.24507/ijicic.16.04.1353.

[5] M. Tassano, J. Delon, and T. Veit, “FastDVDnet: Towards real-time deep video denoising without flow estimation,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 1354–1363, 2020.

[6] Q. Feng, S. Tao, C. Xu, and G. Jin, “BM3D-GT&AD: An improved BM3D denoising algorithm based on Gaussian threshold and angular distance,” *IET Image Process.*, vol. 14, Jan. 2020, doi: 10.1049/iet-ipr.2019.0469.

[7] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, “SwinIR: Image restoration using Swin transformer,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, pp. 1833–1844, 2021.

[8] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, “Restormer: Efficient transformer for high-resolution image restoration,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 5728–5739, 2022.

[9] S. Wang, R. Veldhuis, C. Brune, and N. Strisciuglio, “A survey on the robustness of computer vision models against common corruptions,” *arXiv:2305.06024*, 2023.

[10] S. Schneider, E. Rusak, L. Eck, O. Bringmann, W. Brendel, and M. Bethge, “Improving robustness against common corruptions by covariate shift adaptation,” *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 11539–11551, 2020.

[11] Y. Yang, H. Zhen, Y. Huang, and J. J. Yang, “Enhancing nighttime vehicle detection with day-to-night style transfer and labeling-free augmentation,” *arXiv:2412.16478*, 2024.

[12] S. Sinha, A. Garg, and H. Larochelle, “Curriculum by smoothing,” *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 21653–21664, 2020.

[13] L. Ewecker, E. Asan, L. Ohnemus, and S. Saralajew, “Provident vehicle detection at night for advanced driver assistance systems,” *Auton. Robots*, vol. 47, no. 3, pp. 313–335, 2023.