IEEE Transactions on Automation Science and Engineering (T-ASE) paper, presented at ICRA 2026, Vienna, Austria.
Cite as T-ASE paper.

Robustness of Panoptic Segmentation for Degraded Automotive Cameras Data

Yiting Wang[1,2], Haonan Zhao[1], Daniel Gummadi[1], Dianati Mehrdad[1,3], Debattista Kurt[1], Valentina Donzella[2]

***Abstract*—Precise situational awareness is essential for the safe deployment of artificial intelligence in real-world applications, particularly in assisted and automated driving (AAD) systems. Among perception techniques, panoptic segmentation is a promising technique to identify and categorise objects, impending hazards, and drivable space at a pixel level. While panoptic quality might be affected by automotive camera data quality, a comprehensive understanding and modelling of their relationship remains underexplored. Motivated by such a need, this work proposes a unifying pipeline to evaluate the robustness of panoptic segmentation models for automotive cameras, correlating it with 8 traditional image quality metrics (IQA). The proposed pipeline begins by generating a novel degraded dataset, D-Cityscapes+, featuring 19 realistic automotive degradation types at varying severity levels, including novel models for darkness and snowfall conditions with veiling effect. Evaluations on 14 state-of-the-art segmentation model backbones yielded key insights: 1) large-particle degradations (e.g., lens droplets, heavy snow) severely degrade segmentation performance, increasing uncertainty and edge-concentrated segmentation errors; 2) Transformer-based models outperform CNN models under adverse conditions; however, longer processing time, a higher number of parameters, and computational cost are limiting their real-world deployment; 3) Frequency-based IQA metrics, such as CW-SSIM, strongly correlate with segmentation performance, serving as reliable predictive tools. 4) visual enhancements via restoration do not coherently benefit downstream segmentation tasks, underscoring the need for perception-specific restoration techniques. The benchmark and code:https://github.com/Warwick-Jocelyn/BRPS.**

***Note to Practitioners* — Reliable situational awareness is vital for autonomous systems operating under adverse conditions and potentially degraded sensor data. This study shows that while Transformer-based architectures outperform traditional CNNs in robustness, their long processing time (often <2 FPS) and high computational costs (often >500 GFLOps) limit real-world deployment. We also find that simply increasing the quantity of training data without considering 'noise coverage' offers very limited improvement in model robustness under extreme weather conditions, due to the long-tail effect. These findings underscore the practical value of degraded datasets such as D-Cityscapes+, which simulate 19 realistic conditions at multiple severity levels. Moreover, frequency-based image quality metrics such as CW-SSIM are shown to correlate strongly with perception performance, offering a practical diagnostic tool for benchmarking segmentation robustness. The presented insights provide practitioners with a valuable benchmark, predictive metrics, and clear guidelines for developing robust, real-time-capable segmentation models suitable for challenging automotive environments. Future research could also explore multimodal sensing integration to enhance overall system robustness.**

***Index Terms*—Autonomous driving, perception sensor, panoptic segmentation robustness, image quality metrics.**

[1]WMG, University of Warwick.[2]School of Engineering and Materials Science, Queen Mary University of London. [3] Queen's University of Belfast

The work was partially supported by the Centre for Doctoral Training to Advance Deployment of Future Mobility Technologies (CDT FMT) High-Value Manufacturing CATAPULT. Corresponding author:Yiting.Wang.1@warwick.ac.uk.

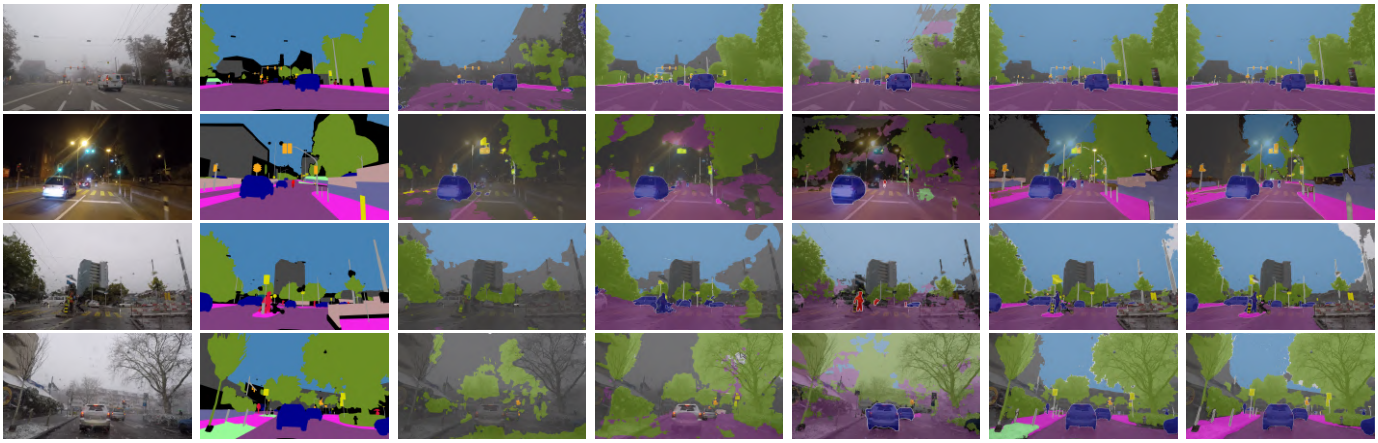
Fig. 1. Visual examples of the newly proposed degraded dataset (D-Cityscapes+) with 19 types of degradation, from top to bottom, are categorised as unfavourable light, adverse weather, internal sensor noises, motion blur, and distortion artefacts.

I. Introduction

Automotive cameras, a prevalent type of perception sensors, are indispensable in systems requiring real-time situational awareness, such as assisted and automated driving (AAD) and video surveillance [1]–[3]. The performance of perception algorithms in these systems is directly influenced by the quality of camera data. Degradation in image quality due to environmental factors, such as lighting variations, weather conditions, and internal sensor noises, can significantly hinder system accuracy and reliability, posing risks to both user safety and overall system functionality [3]–[7].

Amongst various perception tasks, panoptic segmentation stands out for its ability to provide comprehensive scene understanding by simultaneously segmenting countable objects (e.g., cars and pedestrians) and background regions (e.g., sky and buildings) [8], [9]. This granularity of information is essential for making well-informed and precise decisions and for responding to potential hazards in AAD functions. Unlike object detection, semantic segmentation, or instance segmentation, panoptic segmentation offers unique insights into both instance-level and semantic-level features [8]. Despite the advantages of this technique, degradation factors

IEEE Transactions on Automation Science and Engineering (T-ASE) paper, presented at ICRA 2026, Vienna, Austria.



Input GT (a) (b) (c) (d) (e)

Fig. 2. Visual panoptic segmentation results in the real-world ACDC dataset. From top to bottom: fog, night, rain and snow. (a)-(e) represent P-Deeplab (ResNet), P-Deeplab(HRNet), EfficientPS, Oneformer(Swin-L), Oneformer(ConvNeXt-L) respectively.

can significantly impair the visual quality and performance of panoptic segmentation models in the real world. (see Fig. 1, Fig. 2) . Additionally, the even greater degradation can be observed in the more lightweight models, where critical object classes like vehicles fail to be segmented altogether in heavy snow scenarios (see Fig. 5 in the suppl.). Addressing these challenges is essential to ensure the reliability of AAD systems under diverse operating conditions.

While robustness research has extensively examined tasks like object detection and semantic segmentation under degraded conditions, systematic evaluations of panoptic segmentation under AAD perception degradation remain scarce [10]–[12]. Existing robustness studies often overlook key factors, such as dark environments, which are among the primary causes of road accidents [12], [13]. Moreover, the lack of diverse and high-quality degraded datasets limits the scope of prior research, as synthetic datasets often fail to replicate real-world conditions convincingly [11]. *This gap underscores the need for a comprehensive and systematic investigation into the robustness of panoptic segmentation specifically for automotive cameras*. It raises an intriguing inquiry: What inherent noise-related artefacts cause the most severe degradation of perception functions (e.g., panoptic segmentation), and why do certain model architectures demonstrate superior robustness under these degradations?

To address these challenges, this work introduces a unified pipeline to evaluate the robustness of panoptic segmentation models, correlating their performance with image quality metrics (see Fig. 3). Our pipeline includes the following key components: firstly, a realistic degraded dataset, D-Cityscapes+, is synthesised, featuring 19 types of degradation (e.g., low light, adverse weather, and internal sensor noise) simulated at three severity levels. This dataset ensures a uniform distribution of degraded frames and includes newly proposed degradation models for extremely low-light and snow conditions with veiling effects. Secondly, the robustness of 14 state-of-the-art panoptic segmentation architectures, including CNNs (e.g., feature pyramid networks), transformers (e.g., Swin Transformer) and transformer-upgraded CNNs (e.g., ConvNeXt) is systematically analysed across various degradation scenarios. The performance drop, feature corruption, uncertainty, and error distribution are evaluated. Lastly, performance variations are correlated with eight image quality metrics (containing spatial and frequency domain information), identifying metrics such as CW-SSIM that strongly predict segmentation robustness. Additionally, real-world adverse weather scenarios are considered to validate the findings.

Overall, this study is the first to provide both a qualitative and quantitative analysis of the impact of real-world degradation factors on panoptic segmentation for automotive cameras. The main contributions of this work are: **(I) A novel degraded dataset (D-Cityscapes+)** featuring ***19*** automotive-relevant degradation models at multiple severity levels, including two new configurable models for realistic dark and snowy conditions; **(II) A unified evaluation framework** that systematically quantifies the impact of different degradations on ***14*** state-of-the-art panoptic segmentation architectures with and without restoration. It reveals architecture-specific vulnerabilities and characterises uncertainty distributions and feature corruption, identifying the critical trade-off between robustness and real-time performance; **(III) New predictive insights** showing that frequency-based image quality metrics (e.g., CW-SSIM) correlate strongly with segmentation robustness, offering predictive tools for evaluating and improving network robustness.

II. Related Work

Robustness of Automotive Cameras Automotive camera noise factors and perception robustness are widely investigated topics. For example, Ceccarelli et al. discuss the common camera failures during the imaging process and give a quantitative analysis via object detection [13]. Dong et al. simulated common corruptions in cameras and Lidar for 3D object detection [11]. These researchers simulate the degraded paired datasets via simple picture editing that is not designed specifically for automotive cameras, leading to unsatisfactory fidelity for automotive applications *(see supp. materials)*.

A summary of recent robustness research on degraded camera data and the relationship with different perception tasks, such as image classification, detection, and segmentation [11],

IEEE Transactions on Automation Science and Engineering (T-ASE) paper, presented at ICRA 2026, Vienna, Austria.

TABLE I
COMPARISON OF RELATED METHODS IN TERMS OF PERCEPTION TASKS, IMAGE SIGNAL PROCESSING (ISP), COMPRESSION (JPEG), MUD DROP (MUD), SEVERITY LEVELS, AND THE CORRELATION STUDY BETWEEN THE IMAGE QUALITY AND THE PERCEPTION PERFORMANCE IN AUTOMOTIVE (AUTO.).

Paper	Perception Task	Dataset	Noise	ISP	Jpeg	Dark	Mud	Severity	correlat	Auto.
[12]'21	semantic segmentation	synthetic (paired)	✓	x	✓	x	x	x	x	x
[14]'22	panoptic segmentation	real-world (unpaired)	x	x	x	✓	x	✓	x	✓
[11]'23	3D object detection	synthetic (paired)	✓	x	x	x	x	✓	x	✓
[15]'24	3D semantic segmentation	synthetic (paired)	✓	x	x	x	x	✓	x	x
Ours	**panoptic segmentation**	**synthetic + real**	✓	✓	✓	✓	✓	✓	✓	✓

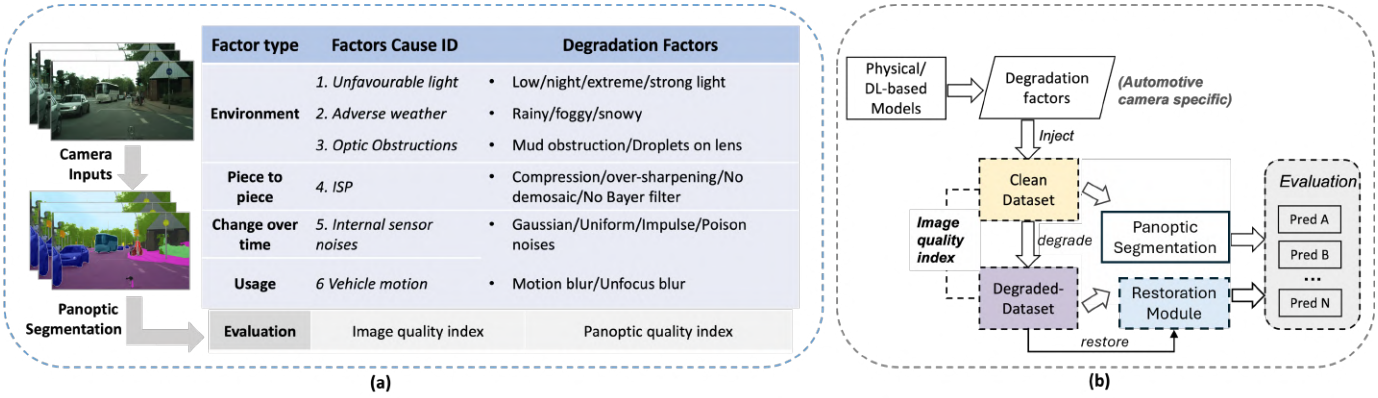


Fig. 3. (a) shows the 19 noise factors. 4 groups of factor types are selected following the P-diagram [16]. The 6 Factors Cause IDs are the currently most commonly seen noise factors impacting perception in automotive cameras, from which 19 specific noise factors that have the potential to be rendered are selected (see Tab. II for the implementation details). (b) illustrates the proposed unifying degradation impact pipeline, which consists of applying the noise factors to the clean dataset, panoptic segmentation, restoration and evaluation.

[17]–[20] is listed in Tab. I). Most of these works focus on natural images; for instance, 15 types of corruption are synthesised in the ImageNet-C dataset [10]. As the robustness of perception is crucial for the safety of real-world applications, research is also emerging on the robustness of automotive camera data, especially under adverse weather [16]. For example, Wang et al. generated 11 types of image degradation and tested the impact of one panoptic segmentation model [21]. Differently from previous work, this paper: **1)** considers a wider range of noise factors, including unfavourable light conditions and pollutant particles (mud and stain) on the camera lens; **2)** uses state-of-the-art noise models that are designed specifically for automotive applications to reduce the simulation to the real-world gaps for fairer robustness validation; **3)** analyses qualitatively and quantitatively the effect of noise factors, including metrics degradation, model uncertainty, and error distribution.

Limitations of Existing Degraded Automotive Camera Datasets Various degradation-driven benchmarking datasets are available to facilitate the robustness of research and to improve the algorithm performance under degradation conditions; they can be divided into real-world captured datasets and synthetic augmented datasets. Many real-world datasets are captured with labels for specific tasks. However, they have several limitations in comprehensive robustness research.

1) Lack of labels and paired data for panoptic segmentation under various noisy conditions. **2)** Uneven distributions of noise factors, resulting in reduced generality and reliability, e.g., only 0.2% of the images are collected in foggy conditions in the BDD100kk dataset [22]. **3)** Scarcity of extreme degradation levels [14], e.g., ACDC only has small rain and snow without raindrops or snowflakes. As an alternative, image synthesis methods are proposed to use physical models [23]–[27], deep learning-based methods [27], [28] and virtual environments [29]–[31] to generate degraded camera datasets. Except for some off-the-shelf generation models, challenges exist in generating datasets with adequate quality for adverse weather and unfavourable light conditions. Platforms such as CARLA might support the option of generating various environmental conditions using the embedded functions, but they have very limited image fidelity tailored for automotive camera applications. Traditional techniques, on the other hand, involve complicated model formulations and hand-crafted parameters. The deep learning (DL)-based algorithm, which offers more flexibility, is data-dependent and lacks simple solutions to regulate the severity of degradation. Additionally, the performance of DL-based approaches is inconsistent (e.g., GAN-based methods failing to capture structural integrity [32]). Therefore, for the degraded data generation in this research, a combination of both traditional and DL-based approaches is selected, with the aim of generating a degraded dataset with better quality.

Panoptic Segmentation State-of-the-art DL-based panoptic segmentation models can be divided into top-down, bottom-up, single-path, and other techniques [33]–[40]. Well-established methods fuse the existing popular instance and semantic segmentation information. For example, feature pyramid networks are used in the panoptic FPN and the UPSNet to

improve multiscale feature learning ability [33], [34]. The lightweight EfficientNet is used in the EfficientPS to reduce the number of parameters [35]. The atrous convolution and dilated convolutions are used in the Panoptic Deeplab to capture multiscale context information [36]. The YOSO model leverages a single-stage, anchor-free design with efficient feature representation and an optimised architecture, enabling it to achieve real-time panoptic segmentation performance [41]. In addition to the CNN-based architectures, recently, ViT-based architectures have been proposed with improved performance, benefiting from the self-attention mechanisms and long-range correlation learning ability [38]–[40], [42]–[44]. For example, Mask2Former and OneFormer propose universal image segmentation strategies that can achieve better performance on multitasks (e.g., semantic, instance, and panoptic segmentation) by training a single model with a single dataset [39], [43].

III. Analysis of Selected Noise Factors

In the context of this paper, *degradation factor* is defined as any external (e.g., adverse environmental conditions) or internal factors (e.g., electronic noise) that may influence the quality of the generated sensor data, with a specific focus on the effects of such factors on the perception algorithms' performance. Reliable evaluation of sensor perception robustness necessitates a comprehensive assessment of noise models. The implementation details can be seen in Tab. II.

A. Rationale of the Selected Degradation Factors

Previous works often focus on natural-image degradations with limited noise types, overlooking automotive-specific conditions such as ISP, low-light conditions, and mud (see Tab. I). To systematically define relevant degradation factors for automotive applications, we follow the established P-diagram framework, considering environmental influences, component-level variability (piece-to-piece), degradation over time, and effects resulting from system usage [16], [45]. Guided by these principles, we categorise six fundamental degradation causes commonly encountered by automotive cameras, deriving 19 specific noise factors (see Fig. 3). *Smog* has high visual similarity and optical characteristics closely resembling fog, so its effect are considered aligned with results presented for fog. Sensor malfunctions are covered extensively through classical random noise patterns (Gaussian, impulse, uniform, Poisson). However, more factors could be considered explicitly in future research if additional differentiation proves necessary.

B. Off-the-shelf Models

Many common image noise models have been recently implemented based on relatively mature theoretical physical models; therefore, some well-established off-the-shelf models have been used to generate some of the identified noise factors: strong light [46], Droplets on lens [13], JPEG compression [47], over-sharpening [47], no Bayer filter [13], no demosaic [48], internal sensor noises [47], motion blur [47] and unfocus blur [47]. Apart from these, noise factors for automotive applications in mud obstructions and No Demosaic are also chosen in this application [16], [48]. More theoretical background regarding off-the-shelf noise models and their effect on automotive cameras can be seen from supp. materials. As for the weather conditions, since there are several augmented multi-weather cityscape datasets available, we choose the most suitable ones based on the following considerations: **1) Rain Model**. Rain Cityscapes [23] generated rainy images following the physical rules considering the depth information, while it fails to simulate the photometry of the rain streak or the field of view needed for the cameras. The new rain rendering method [25] was proposed with the advantage of pre-defining the required rain rate (mm/hr) and producing more vivid rain streaks for generating realistic rainy images, which has been chosen as our rain model. **2) Fog Model**. Foggy Cityscapes [26] uses the scattering model to augment the foggy images. As the original depth map provided from the dataset is incomplete and discontinuous (with random 'holes' missing the depth values), depth denoising and guided filters are used to obtain smooth transmission maps [26]. However, the simple filter operation fails to capture the boundaries between different semantic objects, which leads to the invasion of simulated fog physical simulation. An improved version called Foggy Cityscapes-DBF is therefore proposed [26] with a dual-reference cross-bilateral filter for better adherence to semantic boundaries in the scene and hence chosen as the fog model in this research.

C. Proposed Degradation Factors Model

1. Novel unfavourable light models. In the literature, there is no unified definition of unfavourable light conditions; therefore, this work proposes an intuitive definition of three types of unfavourable low light conditions from darker to brighter and novel approaches tailored to each specific lighting condition. Specifically, low light illustrates the uniform darkening of images, night light represents urban driving areas with vibrant street lights, and extreme light illustrates compound darker illumination compared with the night light. This distinct categorisation ensures a nuanced and accurate representation of various real-world lighting conditions (see first row in Fig. 1). This paper abandons the commonly used gamma correction, which is effective in generating different brightness for indoor static objects but inadequate for simulating real-world dark conditions characterised by uneven light distribution and pixel saturation [49]. Therefore, the low-light images are generated via $I^1_{low} = EZ_{dce}(I, \theta)$ to keep the saturated pixels while darkening the other areas. EZ_{dce} is the reversed curve estimation-based image darkening method naming EC-Zero-DCE [49], θ is the parameters of the model. For nighttime images, the darkening process alone cannot produce the flare and glare features that are invariably present in nighttime photographs. Thus, we generate the night light images N using cycleGAN ($I^1_{night} = GAN_{cyc}(I)$) with the subsequent cycle consistency loss function $L_{cyc}(G, F) = E_{I\sim p_{data}(I)}\|F(G(I)) - I\|_1 + E_{N\sim p_{data}(N)}\|G(F(N)) - N\|_1$ [32]. $G()$ and $F()$ are the generators for clean-to-night and night-to-clean, respectively. Furthermore, the night light model is insufficient for extremely dark images since there remains

IEEE Transactions on Automation Science and Engineering (T-ASE) paper, presented at ICRA 2026, Vienna, Austria.

TABLE II
DETAILS IN TERMS OF THE CATEGORY ID, DEGRADATION FACTOR NAMES, SYNTHETIC IMPLEMENTATION STEPS, AND CONFIGURATIONS FOR THE SELECTED NOISE FACTORS. FOLLOWING PREVIOUS WORK [13], [25], [26], 3 SEVERITY LEVELS ARE CHOSEN FOR EACH DEGRADATION CONDITION.

ID	Degradation	Implementation Details	Configs.
1	**Low light**	Apply the inverse of the DL-based curve estimation image enhancement method (EC-Zero-DCE [49])	The pre-trained model from EC-Zero-DCE [49]
1	**Night light**	Retrain CycleGAN [32] with the BDD100K [22]	PyTorch with 2k epochs for training
1	**Extreme light**	Apply the combination of the model of CycleGAN [32] + EC-Zero-DCE [49]	Apply the low-light model effect to the nightlight model
1	Strong light	Apply the Python package from Imagecorruptions [46]	Brightness corruptions severity levels = {1,3,5}
2	Rainy	Apply rain rendering method [25] with a physical simulator and accurate rain photometric modelling	Three pre-defined severity levels with rain rate (mm/hr) = 50,100,200
2	Foggy	Apply Foggy-DBF datasets with the the atmospheric scattering model from [26]	The attenuation coefficient = 0.005, 0.01, 0.02 with the visibility ranges(m) = {600,300,150}
2	**Snowy**	Improve the Snow Cityscapes [27] by **1)** synthesising diverse snow masks **2)** adding the veiling effect [50]	Snow mask originally generated using Photoshop from [27] and divided into small,medium, and large
3	Mud obstruction	Mud model from [16] for random mud masks generation	Three kernel sizes {12, 24, 36} and intensity 0.7
3	Droplets on lens	Apply the PIL library to blend the raindrops into the image to simulate the droplets on the lens on camera lens	The raindrop masks from [13] with severity level = {2,3,4}
4	Compression	Apply JpegCompression() in the imgaug library [47]	compression levels = {20, 50, 80}
4	Over-sharpening	Apply imgaug.augmenters.Sharpen(alpha, lightness) in the imgaug library [47]	Three levels of severity with alpha ={0.25,0.5,0.75}
4	No Demosaic	Map RGB into RGGB pattern with Bayer colour-filled [48]	Three channel same size (h, w)
4	No Bayer filter	Apply ITU-R BT.601 with numPy array [13]	The scaling factor = {0.2989, 0.5870, 0.1140}
5	Gaussian noise	Apply imgaug.augmenters.imgcorruptlike.GaussianNoise() in the imgaug library [47]	Three pre-defined levels of severity ={1,3,5}
5	Uniform noise	Apply mathematical model, CV2 and numPY to generate random distributed noises following [47]	Three Severity levels with mean value = 0 and standard deviation = {25, 50, 75}
5	Impulse noise	Apply augmenters.arithmetic.ImpulseNoise(severity) [47]	Three pre-defined levels of severity ={1,3,5}
5	Poisson noises	Apply augmenters.arithmetic.AdditivePoissonNoise(lambda) [47]	Three levels of severity with lambda ={5,10,15}
6	Unfocus Blur	Apply augmenters.imgcorruptlike.DefocusBlur() [47]	Three pre-defined levels of severity ={1,3,5}
6	Motion blur	imgaug.augmenters.imgcorruptlike.apply_motion_blur() [47]	Three pre-defined levels of severity ={1,3,5}

a tiny percentage of images with lighting that is essentially almost unchanged from daylight. Therefore, extreme light images are generated by the compound of the above two models with the equation:

$$I^1_{extreme} = EZ_{dce}(GAN_{cyc}(I), \theta)$$

2. Novel falling snow model. Due to the varied shape of the snowflakes, most of the existing snow simulation methods simply use Photoshop to augment the snow layer into the clean image layer, such as the Snow Cityscapes dataset [27] and the Snow 100K dataset [51]. However, the existing models neglect the veiling effect (haze or mist-like effect) when considering Koschmieder's theory of image degradation caused by light scattering and absorption [52]. There are no simulated datasets for automotive cameras that consider the veiling effect, nor labelled datasets captured in the real world with falling snow. Therefore, in this research, we improve the snow model with the added veiling effect to synthesise the snow dataset $I^{2'}_{snow}$ with the following equation:

$$\begin{aligned} I^2_{snow} &= z(x) \ \odot A(x) + I(x) \ \odot (1 - z(x)) \\ I^{2'}_{snow} &= I^2_{snow} * e^{-}\beta d(x) + A(x)(1 - e^{-}\beta d(x)), \end{aligned} \tag{1}$$

where I^2_{snow} are the non-veiling effect snowy images. $\odot$ is the element-wise multiplication, $A(x)$ represents the chromatic aberration map, and $z(x)$ is the independent snow mask. $e^{-}\beta d(x)$ is the median transmission map. β, and $d(x)$ are the scattering coefficient and the distance of the object to the camera, respectively.

IV. PROPOSED PIPELINE

The proposed unifying degradation impact pipeline, shown in Fig. 3, consists of three main steps: synthesis of the degraded dataset (see Sec. III), panoptic segmentation, and impact evaluation. This chapter gives details about the dataset, selected panoptic models, and evaluation metrics.

A. Dataset and Experiment

1) The Cityscapes dataset [53] is used as the clean dataset for the synthetic pipeline. The Cityscapes dataset comprises high-resolution daytime images, meticulously annotated with 19 classes at the pixel level and 30 classes at the instance level. The dataset is chosen to generate the D-Cityscapes+ for: 1) one of the most commonly used driving datasets; 2) high quality during the day with less noise that may influence the synthetic process; 3) containing labels for both instance segmentation and semantic segmentation, along with existing panoptic models that are trained on this dataset. To maintain consistency and expedite comparisons, all 500 images in the dataset are employed for validation. The standard resolution of the images is 2048 $\times$1024, with a crop size of 1024 $\times$512. **2) ACDC dataset** is the only panoptic-labelled real-world driving dataset that contains four noise factors: fog, night, rain, and snow. Therefore, the validation dataset of ACDC is used in the pipeline for evaluation purposes. **3) Experimental setup:** all the experiments are conducted or trained on the NVIDIA RTX 4090 device.

IEEE Transactions on Automation Science and Engineering (T-ASE) paper, presented at ICRA 2026, Vienna, Austria.

B. Selected Panoptic Segmentation Models.

To comprehensively evaluate the robustness of panoptic segmentation models under realistic automotive degradation scenarios, we selected representative state-of-the-art architectures grouped into CNN-based and Transformer-based categories. The CNN-based models include Panoptic DeepLab [36], chosen for its effective multi-scale feature extraction via Atrous Spatial Pyramid Pooling (ASPP); EfficientPS [35], explicitly optimised for automotive applications with an EfficientNet backbone balancing accuracy and efficiency; YOSO [41], designed specifically for real-time automotive deployment; and Mask2Former with a ResNet backbone [43], a widely recognised baseline offering robust segmentation performance.

Transformer-based models, known for capturing long-range dependencies and superior segmentation performance, include multiple variants of Mask2Former (Swin-S and Swin-L backbones) [43] as well as OneFormer [39] (with Swin-L, ConvNeXt-L, ConvNeXt-XL, and DiNAT-L backbones). Additionally, we included OneFormer pretrained on Mapillary Vistas (ConvNeXt-L*/-XL*) to examine the effects with models trained with a larger number of driving datasets. All models utilise pretrained weights from large-scale driving datasets for consistency and fairness.

C. Evaluation Metrics

This research evaluates the degradation impact from three perspectives: synthetic image quality evaluation, image evaluation using panoptic segmentation-based perception, and the correlation between them.

1) Image Quality Analysis. For a comprehensive analysis of the image quality reduction, 8 image quality metrics are selected, encompassing full-reference metrics (i.e., PSNR [54], SSIM [54], FID [55], LPIPS [56], CW-SSIM [57], FSIM [58]) and no-reference metrics (i.e., BRISQUE [59], NIQE [60]). They are chosen due to their wide coverage of considering both the spatial domain (i.e., local mean, local contrast, edge gradients, and chrominance), such as SSIM, NIQE, and the frequency domain (i.e., Fourier- and wavelet-based), such as FSIM and CW-SSIM [61].

2) Panoptic Robustness Evaluation Metrics. To assess how panoptic models perform against the various noise models, the widely used panoptic quality (PQ) measure is adopted. The index PQ serves as the ideal indicator for this research since it contains both object-level information and fine-grained pixel-level information. Specifically, PQ is the product of segmentation quality (SQ) and recognition quality (RQ). RQ calculates the true positive (TP), false positive (FP), and false negative (FN) to get the precision and recall. The larger the PQ values, the better the quality. The PQ can be formulated as in Eq. 2.

$$PQ = \frac{\sum_s^g IOU(s,g)}{|TP|} \times \frac{|TP|}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|} \quad (2)$$

where s represents the segmentation results, g represents the ground truth, and $(s, g) \in TP$. For a set of N images, the average PQ is calculated as: $aPQ = \sum_{i=1}^{N} PQ_i/N$.

3) Correlation Metrics. The correlation between the 8 image quality index and the panoptic quality is calculated using Pearson's linear correlation coefficient (PLCC) and Spearman's rank correlation coefficient (SRCC) [61]. The main difference between PLCC and SRCC is that the former is mainly based on the value, while the latter is based on the rank of each value.

V. Experimental Results and Analysis

A. Degraded Datasets (D-Cityscapes+)

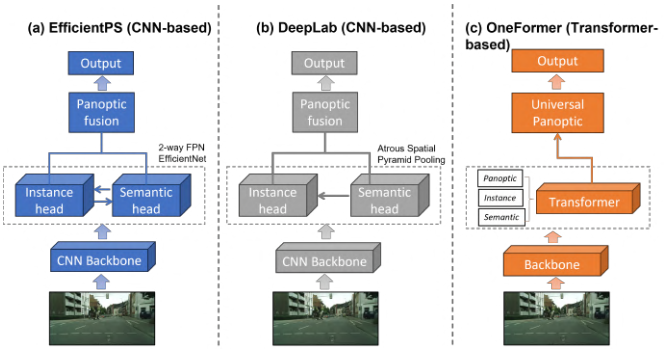


Fig. 4. The overview of the selected panoptic segmentation models: (a) EfficientPS [35] (b) DeepLab [36] (c) Oneformer [39].

1) Overall Visual quality. The visual results of the proposed D-Cityscapes+ are shown in Fig. 1, where 19 types of noise factors are considered. Fig. 8 displays the visual results of the multi-weather at different severity levels. Compared to real-world captured datasets, the newly created dataset shows a better coverage of extreme weather conditions, as they are rarely seen and difficult to capture in the real world (e.g., no raindrops and snowflakes in the ACDC rain and fog dataset). In this work, the snow veiling effect is modelled and added with different severity levels according to the different severity levels of the precipitation, Fig. 8. Overall, our synthetic dataset D-Cityscapes+ is more vivid, especially in representing extreme real-world weather conditions.

2) Novelty and Validation of the Proposed Model. The primary motivation behind introducing our synthetic degradation models for unfavourable lighting (darkness) and snowfall conditions is to realistically simulate complex visual effects encountered in real-world automotive scenarios. Specifically, we distinguish between multiple nuanced dark conditions—i.e., low light, night, and extreme darkness—as each significantly differs visually (see Fig. 1) and leads to substantially different outcomes in downstream perception tasks (Fig. 5 and Tab. III). For instance, while the sky region can be segmented under low-light conditions, it becomes largely indistinguishable in night and extreme-darkness scenarios, posing critical safety implications. Similarly, our snowfall degradation model explicitly incorporates a realistic veiling effect that is characteristic of real-world medium-to-heavy snowfall conditions. This phenomenon, caused by the clustering of snowflakes, significantly reduces visibility and induces a mist-like, blurred appearance of distant objects, closely matching real-world examples (Fig. 6). Moreover, our synthetic models are deliberately designed to be flexible and easily integrated ("plug-and-play"), allowing for diverse and

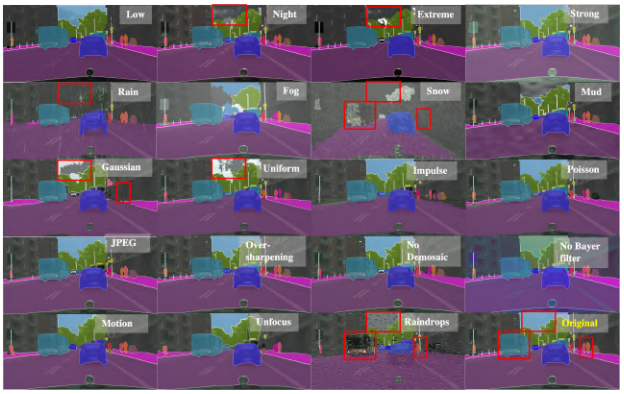


Fig. 5. The overview of the panoptic segmentation using Oneformer [39] *Errors are shown in the red bounding box.*

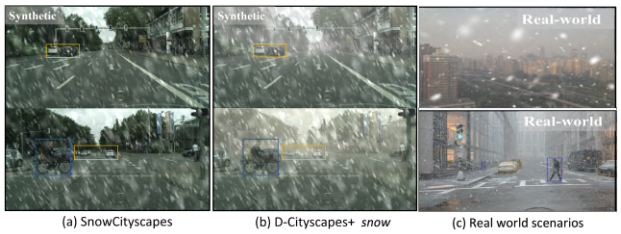


Fig. 6. The comparison of the extreme snow condition. (a) Extreme snow without veiling effect from snow cityscapes [27]. (b) Extreme snow with veiling effect from D-Cityscapes+ (ours). (c) Real-world extreme snow image from SRRS and nuScenes datasets [50], [62].

user-defined combinations of weather severity and illumination conditions. This versatility positions them well for future research into compound weather degradation scenarios.

3) Impact of the noise models and Image Quality. The quantitative results of eight image quality metrics applied to data under all noise factors are shown in Tab. I from the suppl. material. Most synthetic noise models show quality degradation with increasing noise severity, except for certain cases due to differing generation methods. Unfavourable lighting deviates from this trend: low light has lower quality than night light (except in FID), as night light retains more localised illumination. Mud obstructions also show an exception, with PSNR slightly lower at severity 1 than at severity 3, likely because the process reduces mud quantity with increasing kernel size, making PSNR more sensitive to the number rather than the size of obstructions. As for the quality indicated by NIQE and BRISQUE, they are sensitive to noise and artefacts. This may be because NIQE compares the feature distribution in the given image with a pre-computed natural image distribution, which sometimes differs much between the randomly generated noise. BRISQUE instead uses regression-based scoring, which may result in a different trend. Regarding frequency-based image quality, both CW-SSIM and FSIM generally follow the same trend as SSIM, as their frequency and spatial values effectively capture the structural information of the image. Overall, FID consistently aligns with the trend of increasing degradation severity across all factors, making it the most versatile metric to use. In addition, specific metrics, like BRISQUE and PSNR, can offer valuable insights and context for understanding the impact of particular noise factors on image quality, as they are sensitive to slight changes.

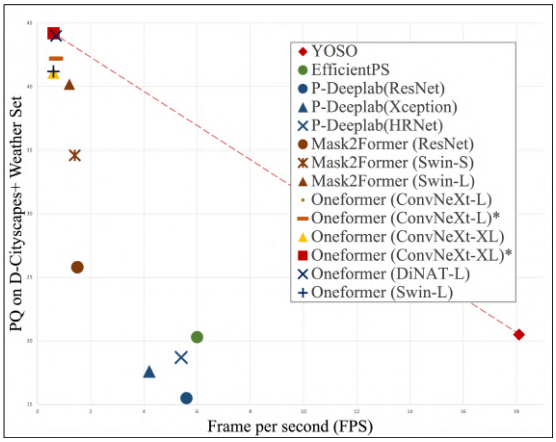


Fig. 7. The FPS with regard to PQ on the D-Cityscapes+ Weather Set.

B. Panoptic Segmentation Results

Impact of Noises on Panoptic Segmentation. The quantitative and visual panoptic segmentation results under 19 types of noise models are illustrated in Tab. III and Fig. 5. *It can be seen that the intensity, distortion, and scale of the noise factor artefacts will directly influence the panoptic segmentation performance.* Adverse weather conditions, characterised by their randomly distributed and complex noise patterns, further amplify these effects, posing unique challenges for segmentation accuracy. For example, Gaussian noise and raindrop occlusion are the most influential factors, while strong light, mud obstructions, and over-sharpening show the smallest impact. To make the comparison more meaningful, we analyse the different panoptic performances within the same categories. ***1)*** *Light level.* The low light and night light only show slight degradation (e.g., PQ drop around 4 and 7), while the combined extreme light with the dynamic exposure largely influences the panoptic segmentation performance (e.g., PQ drop $>$15). The results of strong light show that too bright illumination could also do harm to the panoptic quality. ***2)*** *Adverse weather.* With decreasing visibility, increasing intensity, and bigger particles occurring during adverse weather conditions, the performance decreases (i.e., Snow$<$Rain $<$Fog) under all severity levels. ***3)*** *Lens.* The mud obstructions show better panoptic segmentation performance compared with the image lens covered with droplets on the lens at the same severity levels, which might be caused by a distortion around the droplets on the lens. ***4)*** *ISP.* Compression shows the worst performance, while over-sharpening shows the slightest decrease in PQ since these processes are designed to maintain distinct texture information of the images. ***5)*** *Sensor noise.* Gaussian noise is one of the most impactful sensor noise factors in terms of the drop of PQ, due to the large scale of the noise distribution. ***6)*** *Blur.* Unfocused blur can result in slightly worse performance than motion blur at each severity level. This is potentially due to the different percentages of blur within the camera lens.

C. Robustness of Different Panoptic Segmentation Models

In the comparative study, three state-of-the-art panoptic segmentation models with 14 backbones have been evaluated (Fig. 4), to gauge their ability to handle noise factors.

IEEE Transactions on Automation Science and Engineering (T-ASE) paper, presented at ICRA 2026, Vienna, Austria.

TABLE III
THE PANOPTIC SEGMENTATION QUANTITATIVE RESULTS USING ONEFORMER ON D-CITYSCAPES+ WITH SEVERITY $s = 3$, NOTE THAT S=1,2,3 MEANS SEVERITY LEVELS FROM LIGHT, MEDIUM TO HEAVY.

Metrics	**Clean**	**light level**		**Weather**			**Lens**		**ISP**
		Dark	Strong	Rainy	Foggy	Snowy	mud obstructions	Rain Drops	Compression
Swin-L† (PQ_1)	**67.2**	63.4	66.1	58.5	64.0	48.8	65.4	**34.9**	58.2
ConvNeXt-L† (PQ_1)	68.5	64.2	**67.0**	60.9	65.0	50.1	66.7	**38.5**	59.8
Swin-L† (PQ_2)	**67.2**	60.2	64.5	50.2	61.6	37.5	63.0	47.7	54.3
ConvNeXt-L† (PQ_2)	**68.5**	61.2	66.4	54.4	63.1	38.4	65.5	42.0	53.6
Swin-L† (PQ_3)	**67.2**	52.5	62.6	36.1	56.9	21.5	60.0	36.6	46.6
ConvNeXt-L† (PQ_3)	**68.5**	53.5	63.9	36.6	59.3	20.5	64.7	31.6	43.8

Metrics	**Sensor Noises**				**Blur**		**ISP**		
	Gaussian	Uniform	Impulse	Poisson	Unfocus	Motion	Over-sharp	NoDemo	NoByaer
Swin-L† (PQ_1)	44.8	54.7	59.4	59.7	51.7	54.4	65.5	61.4	61.3
ConvNeXt-L† (PQ_1)	48.4	57.5	58.0	62.5	52.6	55.7	65.5	61.0	62.9
Swin-L† (PQ_2)	**28.2**	50.2	47.3	54.1	37.5	38.6	63.1	61.4	61.3
ConvNeXt-L† (PQ_2)	**32.4**	54.0	44.6	57.0	38.6	39.4	63.6	61.0	62.9
Swin-L† (PQ_3)	**7.4**	47.4	16.7	49.0	25.2	27.8	61.7	61.4	61.3
ConvNeXt-L† (PQ_3)	**11.5**	50.7	22.7	52.2	25.6	28.6	62.2	61.0	62.9

TABLE IV
COMPARISON OF PQ, FPS, PARAMS AND COMPUTATIONAL COST FOR PANOPTIC SEGMENTATION MODELS UNDER THE D-CITYSCAPES+ WEATHER SET

Methods	**Backbones**	**Day**	**Fog**	**Snow**	**Rain**	**Dark**	**#FPS**	**#Params.(M)**	**#FLOPs.(G)**
YOSO [41]	/	52.5	38.4	5.7	3.8	34.1	**18.1**	43.0	**67**
EfficientPS [35]	EfficientNet	62.8	35.5	4.4	17.2	24.0	6.0	41.0	108
P-Deeplab [36]	ResNet	60.3	34.2	2.7	7.5	17.7	5.6	**30.0**	91
P-Deeplab [36]	Xception	68.5	36.1	2.7	10.9	20.6	4.2	45.0	136
P-Deeplab [36]	HRNet	63.4	40.7	3.4	7.2	23.5	5.4	68.0	177
Mask2Former [43]	ResNet	62.4	45.3	5.7	10.6	41.5	1.5	63.0	171
Mask2Former [43]	Swin-S	64.8	50.9	13.3	29.7	44.4	1.4	69.0	185
Mask2Former [43]	Swin-L	66.6	55.5	**24.2**	30.7	50.3	1.2	215	514
Oneformer [39]	Swin-L	67.2	56.9	21.5	34.0	52.5	0.6	219	543
Oneformer [39]	DiNAT-L	67.6	59.1	22.8	**42.7**	51.2	0.7	223	450
Oneformer [39]	ConvNeXt-L	68.5	59.3	20.5	35.3	53.5	0.7	220	497
Oneformer [39]	ConvNeXt-XL	68.4	58.7	18.3	34.2	53.3	0.6	372	775
Oneformer [39]	ConvNeXt-L*	**70.1**	**61.5**	19.6	31.5	**56.2**	0.7	220	497
Oneformer [39]	ConvNeXt-XL*	69.7	**61.5**	23.1	36.5	55.8	0.6	372	775

1) Effect of Different Severity Levels. Both Tab. III and Fig. 4 from the appendix quantitatively and visually represent the comparison among these models in terms of panoptic quality under three different severity levels of the noise factors. More results can be seen in the supplementary material. The transformer-based OneFormer achieves superior performance, with an average PQ of 55.9, compared to Panoptic Deeplab (PQ=32.96), which relies on feature pyramid-based CNN networks. EfficientPS utilising EfficientNet showcases commendable performance in both PQ and speed, outperforming the Panoptic Deeplab, underscoring the potential for leveraging EfficientNet in lightweight, robust network architecture design.

2) Analysis on CNN VS Transformer. The potential for enhanced robustness and generality of transformers becomes evident, especially with respect to internal sensor noise factors. For instance, the average PQ value in 4 different internal sensor noise factors under different severity levels for the Oneformer is much better (55.7) compared with both P-Deeplab (14.6) and the EfficientPS (28.1). The analysis in Fig. 8 reveals distinct differences in how Panoptic-Deeplab (CNN-based) and Oneformer (transformer-based) architectures respond to increasing levels of snow corruption. As snow severity intensifies, both models exhibit degraded feature representations; however, the corruption manifests differently due to their architectural designs. Panoptic-Deeplab, emphasising local detail, captures individual snowflakes as pronounced distortions in the feature maps, reflecting its sensitivity to small-scale, spatially localised noise. In contrast, Oneformer's transformer-based architecture, with its global self-attention mechanism, demonstrates a smoother response, as individual snowflakes do not disrupt the feature maps as prominently. This comparison suggests that transformers are prone to capturing long-range dependencies and may possess a robustness advantage in adverse conditions where spatially sparse noise is prevalent, as their global feature extraction process can mitigate the impact of localised corruption.

3) Analysis on Backbone Designs and Performance. In the OneFormer model, the ConvNeXt backbone demonstrates a slight advantage over the Swin Transformer backbone. This advantage suggests that incorporating larger kernels and depth-wise convolutions in architectures with expanded receptive fields can improve spatial relationship learning. These CNN-enhanced, transformer-inspired strategies allow for more effective local and global feature capture to enhance performance and robustness in modern CNN designs. These findings highlight the transformer model's potential suitability for scenarios

requiring resilience to environmental noise, as it retains a more coherent scene representation under challenging conditions. Furthermore, as shown in Tab. IV, models pretrained on large-scale driving datasets exhibited only marginal improvements in robustness (approximately 0-+2 PQ compared to their counterparts without such pretraining). This suggests a saturation effect where additional data yields diminishing returns in performance improvement. Moreover, since datasets like Mapillary Vistas predominantly contain clean weather scenarios without significant degradations, it underscores the importance of addressing domain gaps for better robustness under adverse conditions.

4) Trades-off Between Performance and Cost. The practical deployment of panoptic segmentation models in automotive applications inherently involves balancing perception accuracy (PQ) with computational efficiency (e.g FPS). A detailed comparison of segmentation models PQ with regard to processing speed (FPS), parameters (Params), and computational costs (GFLOPs) is shown in Tab. IV. These values are further visualised in Fig. 7, where we observed a clear trade-off between these two critical metrics. Among the evaluated models, lightweight architectures, specifically optimised for automotive real-time scenarios, such as YOSO, achieve significantly higher inference speeds (18.1 FPS), enabling real-time operation on embedded platforms. However, these models typically exhibit notably lower segmentation accuracy, especially under severe degradation conditions, demonstrating substantial performance degradation (PQ below 40). Conversely, Transformer-based models like OneFormer (ConvNeXt-XL*) offer state-of-the-art panoptic segmentation performance (highest PQ of approximately 60 under adverse weather conditions), yet at a substantially increased computational cost (0.6 FPS), making real-time deployment challenging. The observed inverse relationship between PQ and FPS highlights a critical deployment challenge in automotive environments: achieving acceptable segmentation accuracy while meeting strict computational constraints and real-time inference requirements. Future research should thus prioritise designing models that intelligently balance accuracy and inference speed, possibly through domain-specific model compression and distillation and efficient architecture innovations.

D. Correlation between IQA, PQ and Error Distribution

For the analysed 19 noise factors, the average correlation indexes (PLCC and SRCC) between PQ and the selected IQA are reported in Tab. V. To enable a more comprehensive analysis of panoptic segmentation degradation across different severity levels, uncertainty and error maps are generated (see Fig. 9). In these maps, yellow and white areas indicate regions of high uncertainty and error, respectively. As shown in the figure, more intense rain streaks lead to a broader distribution of errors and increased model overconfidence.

As can be seen from Tab. V, the two panoptic models show a similar trend with the correlation index, which increases the reliability of the results. Specifically, the most correlated positive and negative indexes are CW-SSIM and LPIPS for PLCC, CW-SSIM and FID for SRCC, respectively. Specifically, the CW-SSIM metric shows the strongest alignment with PQ values, indicating its potential to reflect perceptual degradation. This may be due to CW-SSIM's ability to capture complex wavelet transform information, emphasising wavelet sub-bands associated with edges, textures, and fine details. These characteristics likely align with the high uncertainty and error distribution observed in panoptic segmentation (see Fig. 9).

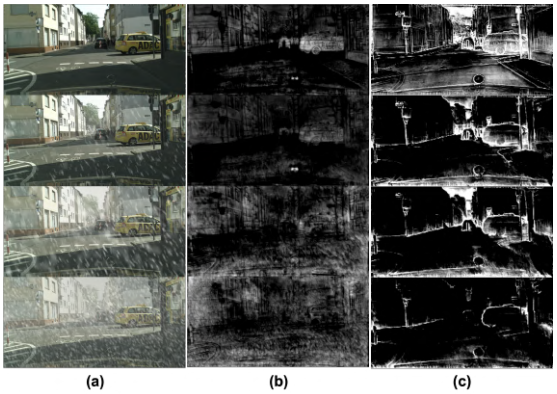


Fig. 8. Visual examples of the features extracted under different severity levels of the snowy conditions, left to right: (a) input (b) features from Panoptic-Deeplab (c) features from Oneformer.

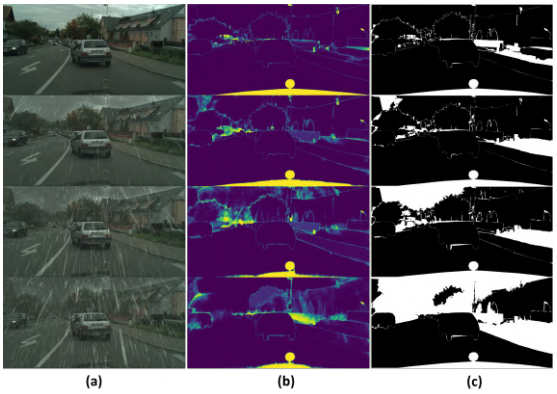


Fig. 9. Overview of clean and rainy images (S=1,2,3) with corresponding uncertainty and error maps generated by Oneformer. Notably, regions of high uncertainty and error are closely aligned along the edges, suggesting that noise primarily impacts the model's confidence in the high-frequency areas.

E. Analysis of Restoration Methods on the Degraded Dataset

Recent literature has introduced numerous open-source restoration methods for tasks such as rain removal, low-light image enhancement, and dehazing [63]–[67]. To assess the potential benefit of restoration techniques on perception performance under adverse conditions, we evaluated a recent state-of-the-art universal image restoration method—DA-CLIP [63], a vision-language approach controlling image restoration quality via text guidance. Specifically, we analysed the impact of applying DA-CLIP on both image quality (PSNR, SSIM, LPIPS) and downstream panoptic segmentation performance (PQ), see Tab. VI.

Interestingly, while DA-CLIP generally improved visual quality (higher PSNR and SSIM, lower LPIPS), these improvements did not consistently correlate with enhanced perception outcomes. For instance, in snow scenarios, restoration led

IEEE Transactions on Automation Science and Engineering (T-ASE) paper, presented at ICRA 2026, Vienna, Austria.

TABLE V
THE OVERALL CORRELATION BETWEEN THE 8 IQA AND THE PQ WITH THE AVERAGE PLCC AND SRCC.

Model	Index	PSNR↑	SSIM↑	FID↓	LPIPS↓	NIQE↓	CW-SSIM↑	FSIM↑	BRISQUE↓
EfficientPS	**PLCC**	0.807	0.948	-0.923	-0.948	-0.513	**0.954**	0.921	-0.519
EfficientPS	**SRCC**	0.800	0.933	**-0.967**	-0.867	-0.533	**0.967**	0.933	-0.600
Oneformer(ConNeXt-L)	**PLCC**	0.813	0.961	-0.947	**-0.979**	-0.569	**0.975**	0.966	-0.661
Oneformer(ConNeXt-L)	**SRCC**	0.800	0.933	-0.967	-0.933	-0.533	**0.967**	0.933	-0.600

TABLE VI
CHANGE IN IQA AND PQ UNDER SEVERITY LEVEL LIGHT (-L) AND SEVERE (-S) AFTER RESTORATION ON D-CITYSCAPES+ WEATHER SET.

Metric	Fog-L	Rain-L	Snow-L	Dark-L	Fog-S	Rain-S	Snow-S	Dark-S
PSNR↑	-0.65	-2.74	+1.06	+6.71	+2.81	-0.31	+3.75	+6.05
SSIM↑	-0.03	-0.07	-0.03	+0.39	+0.05	+0.01	+0.16	+0.44
LPIPS↓	+0.03	0.06	-0.19	-0.24	-0.04	-0.02	-0.29	-0.28
PQ↑	-0.72	-2.97	8.14	-7.79	+0.54	+4.83	+18.31	-2.77

TABLE VII
PQ METRIC COMPARISON ON DIFFERENT WEATHER CONDITIONS IN THE ACDC DATASET.

Method	Fog	Night	Rain	Snow
Oneformer (ConvNeXt-XL)	**69.95**	35.16	**57.56**	54.58
Oneformer (ConvNeXt-L)	67.89	34.98	56.90	**56.08**
Oneformer (Swin-L)	67.29	**35.34**	56.50	53.79
EfficientPS	31.20	6.30	26.70	23.10
P-Deeplab (HRNet)	34.73	0.90	25.26	20.85
P-Deeplab (Xception)	30.94	1.06	21.20	22.16
P-Deeplab (ResNet)	12.22	1.20	9.43	2.88

to substantial PQ increases (+8.14 and +18.31, respectively), aligned with visual quality improvements. However, in low-light conditions (Dark-L), panoptic segmentation quality significantly decreased (-7.79 in PQ) despite considerable improvements in image quality metrics (PSNR +6.71, SSIM +0.39, LPIPS -0.24). These findings emphasise that improved visual appearance alone does not ensure superior downstream perception results. Thus, future restoration in the wild methods should integrate task-aware objectives optimised specifically for perceptual performance rather than relying solely on visual quality metrics, potentially bridging this critical gap.

F. Robustness to the Real-World Adverse Weather

As shown in Table VII, the Oneformer model demonstrates strong robustness across various adverse weather conditions, both visually and quantitatively. In particular, nighttime conditions yield the poorest PQ scores, as the stark contrast between foreground and background elements and large dark areas in the sky significantly reduces model effectiveness. For the remaining weather conditions (i.e., fog, rain, and snow), PQ values for background elements consistently surpass those for foreground objects, highlighting the amplified impact of adverse weather on small objects and raising critical safety concerns. On average, robustness trends on the real-world ACDC dataset align well with those observed on our synthetic D-Cityscapes+ dataset, following the same model performance order (Oneformer >EfficientPS >Panoptic Deeplab) as illustrated in Table VII. Degradation trends across weather conditions are also consistent: snow impacts performance most significantly, followed by rain, then fog, for both real-world and synthetic datasets. This consistency underscores the reliability of D-Cityscapes+ as a benchmark for robustness research. Hence, they provide valuable insight with a comprehensive panoptic segmentation evaluation benchmark, both in synthetic and real-world degraded conditions.

G. Deployment challenges under extreme weather conditions

Deploying panoptic segmentation models in real-world automotive applications faces substantial challenges, especially under extreme weather conditions. While our experiments show that degradation patterns observed in real-world data typically align with light severity level in D-Cityscapes+, the representation for moderate-to-extreme conditions remains limited due to practical constraints. Specifically, our evaluations highlight severe performance drops under extreme conditions (e.g., panoptic quality (PQ) nearly halves for severe rain and snow), emphasising the vulnerability of current models to extreme challenging weather scenarios.

One fundamental challenge is capturing sufficient real-world data under extreme weather conditions due to inherent safety risks, logistical constraints, and the long-tail distribution of these events. Our synthetic dataset D-Cityscapes+, with carefully designed severe-level scenarios, provides essential resources to bridge this gap. However, the extreme weather conditions can result in compound degradations, such as simultaneous blur, noise for extreme dark scenarios. In the future, multi-factor synthesis pipelines and compounded noise models capable of accurately simulating realistic multiple noise factors can be further developed based on D-Cityscapes+. Another significant deployment issue is balancing accuracy with real-time computational constraints. For instance, lightweight architectures (e.g., YOSO) typically suffer significant performance degradation under extreme weather, whereas more robust models (e.g., OneFormer) incur high computational costs, challenging real-time applicability. Future research should thus prioritise developing domain-aware, computationally efficient models specifically optimised for extreme weather conditions, which can reliably operate within bandwidth-limited, resource-constrained automotive platforms. Moreover, interdisciplinary integration—such as leveraging advanced communication technologies, Internet of Things (IoT) frameworks, or emerging computational paradigms like quantum computing—holds promise for overcoming current limitations. Such holistic approaches could provide robust solutions for reliable

autonomous perception systems operating in extreme, safety-critical environments.

VI. Conclusion

This study introduces a comprehensive evaluation framework for assessing perception robustness in conjunction with sensor data quality. The framework includes (i) an augmented, balanced dataset, D-Cityscapes+, that features improved snow and extreme lighting noise models and (ii) an experimental evaluation of perception performance. (iii) examines the relationship between image quality, panoptic quality, feature corruptions, model uncertainty, and error distribution, offering predictive metrics for machine learning performance. Results reveal that high-frequency, large-scale corruptions, such as lens droplets and internal noises, significantly impact perception, with ViT-based architectures demonstrating greater global resilience due to their ability to model long-range dependencies. This framework provides critical insights into the effects of noise on perception systems, guiding the development of safer, more robust applications in automation science and engineering. **Limitations and Future Work:** As mentioned, this work has some limitations that will be further investigated in the future work. First, there is an inherent domain gap between models trained on synthetic data and real-world degraded data. This gap may partly explain why visually improved images sometimes paradoxically reduce downstream segmentation performance (see Tab. VI). Future research may therefore, focus on domain adaptation techniques and compound data augmentation to achieve better generalise segmentation results across diverse and extreme real-world degradation scenarios. Additionally, our experiments demonstrated that Transformer-based architectures, while superior in robustness, are significantly constrained by computational costs (see Tab. IV), limiting their real-world automotive deployment. Future work could address this practical limitation by developing lightweight Transformer architectures or hybrid models that balance accuracy and computational efficiency. Model compression, quantisation, and knowledge distillation strategies specifically tailored for degraded automotive environments might prove particularly beneficial. Furthermore, autonomous systems may leverage temporal and multimodal data for robustness research.

References

[1] Z. Xu, B. Liu, X. Xiao, A. Nair, and P. Stone, "Benchmarking reinforcement learning techniques for autonomous navigation," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 9224–9230.

[2] X. Yao, Y. Wang, L. Dai, S. Zhang, M. Dou, and Y. Deng, "Semi-supervised domain adaptation with dual-adversarial learning for lane detection," *IEEE Transactions on Automation Science and Engineering*, pp. 1–13, 2024.

[3] Y. Sun, W. Zuo, P. Yun, H. Wang, and M. Liu, "Fuseseg: Semantic segmentation of urban scenes based on rgb and thermal data fusion," *IEEE Transactions on Automation Science and Engineering*, vol. 18, no. 3, pp. 1000–1011, 2021.

[4] M. S. Ramanagopal, C. Anderson, R. Vasudevan, and M. Johnson-Roberson, "Failing to learn: Autonomously identifying perception failures for self-driving cars," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3860–3867, 2018.

[5] Y. Lee, Y. Kim, J. Yu, and M. Jeon, "Learning to remove bad weather: Towards robust visual perception for self-driving," *IEEE Robotics and Automation Letters*, pp. 1–1, 2022.

[6] H. Huang, M. Zhang, L. Li, J. Hu, and H. Wang, "Gtscalib: Generalized target segmentation for target-based extrinsic calibration of non-repetitive scanning lidar and camera," *IEEE Transactions on Automation Science and Engineering*, pp. 1–13, 2024.

[7] W. Tong, A. Gu, X. Wu, X. Deng, Y. Cai, Y. Duan, and Y. Hou, "Semi-supervised image domain adaption for aerial refueling drogue detection on embedded chip under foggy conditions," *IEEE Transactions on Automation Science and Engineering*, pp. 1–12, 2024.

[8] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollár, "Panoptic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9404–9413.

[9] W. Zhou, Y. Xiao, W. Yan, and L. Yu, "Cmpffnet: Cross-modal and progressive feature fusion network for rgb-d indoor scene semantic segmentation," *IEEE Transactions on Automation Science and Engineering*, vol. 21, no. 4, pp. 5523–5533, 2024.

[10] D. Hendrycks and T. Dietterich, "Benchmarking neural network robustness to common corruptions and perturbations," *arXiv preprint arXiv:1903.12261*, 2019.

[11] Y. Dong, C. Kang, J. Zhang, Z. Zhu, Y. Wang, X. Yang, H. Su, X. Wei, and J. Zhu, "Benchmarking robustness of 3d object detection to common corruptions in autonomous driving," *arXiv preprint arXiv:2303.11040*, 2023.

[12] C. Kamann and C. Rother, "Benchmarking the robustness of semantic segmentation models with respect to common corruptions," *International journal of computer vision*, vol. 129, pp. 462–483, 2021.

[13] A. Ceccarelli and F. Secci, "Rgb cameras failures and their effects in autonomous driving applications," *IEEE Transactions on Dependable and Secure Computing*, 2022.

[14] O. Zendel, M. Schörghuber, B. Rainer, M. Murschitz, and C. Beleznai, "Unifying panoptic segmentation for autonomous driving," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 21 351–21 360.

[15] X. Yan, C. Zheng, Y. Xue, Z. Li, S. Cui, and D. Dai, "Benchmarking the robustness of lidar semantic segmentation models," *International Journal of Computer Vision*, pp. 1–24, 2024.

[16] B. Li, P. H. Chan, G. Baris, M. D. Higgins, and V. Donzella, "Analysis of automotive camera sensor noise factors and impact on object detection," *IEEE Sensors Journal*, vol. 22, no. 22, pp. 22 210–22 219, 2022.

[17] K. Wang, T. Zhou, X. Li, and F. Ren, "Performance and challenges of 3d object detection methods in complex scenes for autonomous driving," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 2, pp. 1699–1716, 2022.

[18] Y. Dalva, H. Pehlivan, S. F. Altındiş, and A. Dundar, "Benchmarking the robustness of instance segmentation models," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[19] Y. Wang, P. H. Chan, and V. Donzella, "Semantic-aware video compression for automotive cameras," *IEEE Transactions on Intelligent Vehicles*, 2023.

[20] K. Xian, Z. Cao, C. Shen, and G. Lin, "Towards robust monocular depth estimation: A new baseline and benchmark," *International Journal of Computer Vision*, pp. 1–19, 2024.

[21] Y. Wang, H. Zhao, K. Debattista, and V. Donzella, "The effect of camera data degradation factors on panoptic segmentation for automated driving," in *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2023, pp. 2351–2356.

[22] 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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 2636–2645.

[23] X. Hu, C.-W. Fu, L. Zhu, and P.-A. Heng, "Depth-attentional features for single-image rain removal," in *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, 2019, pp. 8022–8031.

[24] K. Garg and S. K. Nayar, "Vision and rain," *International Journal of Computer Vision*, vol. 75, pp. 3–27, 2007.

[25] M. Tremblay, S. S. Halder, R. De Charette, and J.-F. Lalonde, "Rain rendering for evaluating and improving robustness to bad weather," *International Journal of Computer Vision*, vol. 129, pp. 341–360, 2021.

[26] C. Sakaridis, D. Dai, and L. Van Gool, "Semantic foggy scene understanding with synthetic data," *International Journal of Computer Vision*, vol. 126, pp. 973–992, 2018.

[27] K. Zhang, R. Li, Y. Yu, W. Luo, and C. Li, "Deep dense multi-scale network for snow removal using semantic and depth priors," *IEEE Transactions on Image Processing*, vol. 30, pp. 7419–7431, 2021.

[28] C. Sakaridis, D. Dai, and L. V. Gool, "Guided curriculum model adaptation and uncertainty-aware evaluation for semantic nighttime image segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 7374–7383.

[29] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "Carla: An open urban driving simulator," in *Conference on robot learning*. PMLR, 2017, pp. 1–16.

[30] T. Sun, M. Segu, J. Postels, Y. Wang, L. Van Gool, B. Schiele, F. Tombari, and F. Yu, "Shift: a synthetic driving dataset for continuous multi-task domain adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 21 371–21 382.

[31] A. Gaidon, Q. Wang, Y. Cabon, and E. Vig, "Virtual worlds as proxy for multi-object tracking analysis," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 4340–4349.

[32] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[33] A. Kirillov, R. Girshick, K. He, and P. Dollár, "Panoptic feature pyramid networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 6399–6408.

[34] Y. Xiong, R. Liao, H. Zhao, R. Hu, M. Bai, E. Yumer, and R. Urtasun, "Upsnet: A unified panoptic segmentation network," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8818–8826.

[35] R. Mohan and A. Valada, "Efficientps: Efficient panoptic segmentation," *International Journal of Computer Vision*, vol. 129, no. 5, pp. 1551–1579, 2021.

[36] B. Cheng, M. D. Collins, Y. Zhu, T. Liu, T. S. Huang, H. Adam, and L.-C. Chen, "Panoptic-deeplab: A simple, strong, and fast baseline for bottom-up panoptic segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 12 475–12 485.

[37] Y. Li, H. Zhao, X. Qi, L. Wang, Z. Li, J. Sun, and J. Jia, "Fully convolutional networks for panoptic segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 214–223.

[38] Z. Li, W. Wang, E. Xie, Z. Yu, A. Anandkumar, J. M. Alvarez, P. Luo, and T. Lu, "Panoptic segformer: Delving deeper into panoptic segmentation with transformers," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 1280–1289.

[39] J. Jain, J. Li, M. T. Chiu, A. Hassani, N. Orlov, and H. Shi, "Oneformer: One transformer to rule universal image segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 2989–2998.

[40] F. Li, H. Zhang, H. Xu, S. Liu, L. Zhang, L. M. Ni, and H.-Y. Shum, "Mask dino: Towards a unified transformer-based framework for object detection and segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 3041–3050.

[41] J. Hu, L. Huang, T. Ren, S. Zhang, R. Ji, and L. Cao, "You only segment once: Towards real-time panoptic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 17 819–17 829.

[42] B. Cheng, A. Schwing, and A. Kirillov, "Per-pixel classification is not all you need for semantic segmentation," *Advances in Neural Information Processing Systems*, vol. 34, pp. 17 864–17 875, 2021.

[43] B. Cheng, I. Misra, A. G. Schwing, A. Kirillov, and R. Girdhar, "Masked-attention mask transformer for universal image segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 1290–1299.

[44] H. Wang, Y. Zhu, H. Adam, A. Yuille, and L.-C. Chen, "Max-deeplab: End-to-end panoptic segmentation with mask transformers," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 5463–5474.

[45] G. Baris, L. Boda, C. P. Hung, C. A. Avizzano, D. Valentina *et al.*, "Analysis of faster r-cnn network prediction in the presence of lens occlusion and video compression," 2023.

[46] C. Michaelis, B. Mitzkus, R. Geirhos, E. Rusak, O. Bringmann, A. S. Ecker, M. Bethge, and W. Brendel, "Benchmarking robustness in object detection: Autonomous driving when winter is coming," *arXiv preprint arXiv:1907.07484*, 2019.

[47] A. B. Jung, K. Wada, J. Crall, S. Tanaka, J. Graving, C. Reinders, S. Yadav, J. Banerjee, G. Vecsei, A. Kraft, Z. Rui, J. Borovec, C. Vallentin, S. Zhydenko, K. Pfeiffer, B. Cook, I. Fernández, F.-M. De Rainville, C.-H. Weng, A. Ayala-Acevedo, R. Meudec, M. Laporte *et al.*, "imgaug," https://github.com/aleju/imgaug, 2020, online; accessed 01-Feb-2020.

[48] P. H. Chan, C. Wei, A. Huggett, and V. Donzella, "Raw camera data object detectors: an optimisation for automotive processing and transmission," *Authorea Preprints*, 2023.

[49] S. Zhou, C. Li, and C. C. Loy, "Lednet: Joint low-light enhancement and deblurring in the dark," in *ECCV*, 2022.

[50] W.-T. Chen, H.-Y. Fang, J.-J. Ding, C.-C. Tsai, and S.-Y. Kuo, "Jstasr: Joint size and transparency-aware snow removal algorithm based on modified partial convolution and veiling effect removal," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXI 16*. Springer, 2020, pp. 754–770.

[51] Y.-F. Liu, D.-W. Jaw, S.-C. Huang, and J.-N. Hwang, "Desnownet: Context-aware deep network for snow removal," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 3064–3073, 2018.

[52] H. Koschmieder, "Theorie der horizontalen sichtweite," *Beitrage zur Physik der freien Atmosphare*, pp. 33–53, 1924.

[53] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3213–3223.

[54] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[55] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," 2018.

[56] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 586–595.

[57] M. P. Sampat, Z. Wang, S. Gupta, A. C. Bovik, and M. K. Markey, "Complex wavelet structural similarity: A new image similarity index," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2385–2401, 2009.

[58] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "Fsim: A feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.

[59] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012.

[60] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2013.

[61] D. Gummadi, P. H. Chan, H. Wang, and V. Donzella, "Correlating traditional image quality metrics and dnn-based object detection: a case study with compressed camera data," *Authorea Preprints*, 2023.

[62] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuscenes: A multimodal dataset for autonomous driving," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11 621–11 631.

[63] Z. Luo, F. K. Gustafsson, Z. Zhao, J. Sjölund, and T. B. Schön, "Controlling vision-language models for universal image restoration," *arXiv preprint arXiv:2310.01018*, vol. 3, no. 8, 2023.

[64] T. Wang, K. Zhang, Z. Shao, W. Luo, B. Stenger, T. Lu, T.-K. Kim, W. Liu, and H. Li, "Gridformer: Residual dense transformer with grid structure for image restoration in adverse weather conditions," *International Journal of Computer Vision*, vol. 132, no. 10, pp. 4541–4563, 2024.

[65] K. Zhang, D. Li, W. Luo, W. Ren, and W. Liu, "Enhanced spatio-temporal interaction learning for video deraining: faster and better," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp. 1287–1293, 2022.

[66] K. Zhang, W. Luo, Y. Yu, W. Ren, F. Zhao, C. Li, L. Ma, W. Liu, and H. Li, "Beyond monocular deraining: Parallel stereo deraining network via semantic prior," *International Journal of Computer Vision*, vol. 130, no. 7, pp. 1754–1769, 2022.

[67] T. Wang, K. Zhang, Y. Zhang, W. Luo, B. Stenger, T. Lu, T.-K. Kim, and W. Liu, "Lldiffusion: Learning degradation representations in diffusion models for low-light image enhancement," *Pattern Recognition*, vol. 166, p. 111628, 2025.