PAIR : Perception Aided Image Restoration for Natural Driving Conditions

Pranjay Shyam
Faurecia IRYStec Inc.
Montreal, Canada
pranjay.shyam.psm@forvia.com

HyunJin Yoo
Faurecia IRYStec Inc.
Montreal, Canada
hyunjin.yoo@forvia.com

Abstract

We present a two-stage mechanism for generic image restoration in natural driving conditions, where multiple non-linear degradations simultaneously impact perception for humans and driving assistance systems. Our approach overcomes the limitations of utilizing a single neural network that incurs excessive computational overhead and yields sub-optimal recovery. The proposed first stage comprises computationally inexpensive image processing operations applied at a patch level using a lightweight convolutional neural network (CNN) that determines their intensity of operation. This patch size is guided by the receptive field of the CNN, allowing for dynamic restoration of non-linear and non-homogeneous degradation profiles. The second stage leverages a lightweight end-to-end neural network functioning as an inpainting network. It identifies inadequately restored regions and leverages global semantic and structural information to fill the affected areas. This approach enhances the restoration process by considering the entire image and addresses the remainder of localized deficiencies. In addition, we integrate dense perception tasks such as semantic and depth estimation during the optimization cycle to ensure restored images that are perceptually pleasing and conducive for downstream perception tasks. Since datasets covering diverse degradation scenarios for high- and low-level perception tasks are lacking, we utilize a synthetic data augmentation technique to generate non-homogeneous non-linear degradation profiles. Experiments on images captured in adverse weather conditions demonstrate the efficacy of our approach, yielding higher perceptual quality in restored images and improved performance in downstream perception tasks under adverse driving conditions. Importantly, our method offers computational efficiency compared to end-to-end image restoration algorithms, making it suitable for real-time applications.

1. Introduction

Enhancing the perceptual quality of images affected by natural weather degradation is paramount for human drivers and modern driving assistance systems. Consequently, extensive interest is shown in developing algorithms capable of recovering clear images from their degraded counterparts. The current body of literature on image restoration primarily focuses on degradation-specific approaches, wherein early methods employed traditional computer vision algorithms to construct simple filters [50, 59]. In contrast, state-of-the-art (SoTA) approaches rely on data-driven techniques [31, 53, 79], leveraging the feature extraction and representation capabilities of neural networks to restore degraded images. However, the degradation-specific nature of these algorithms necessitates a preliminary identification step, resulting in computationally expensive two-stage solutions (identity, then restore). Thus, despite the evident

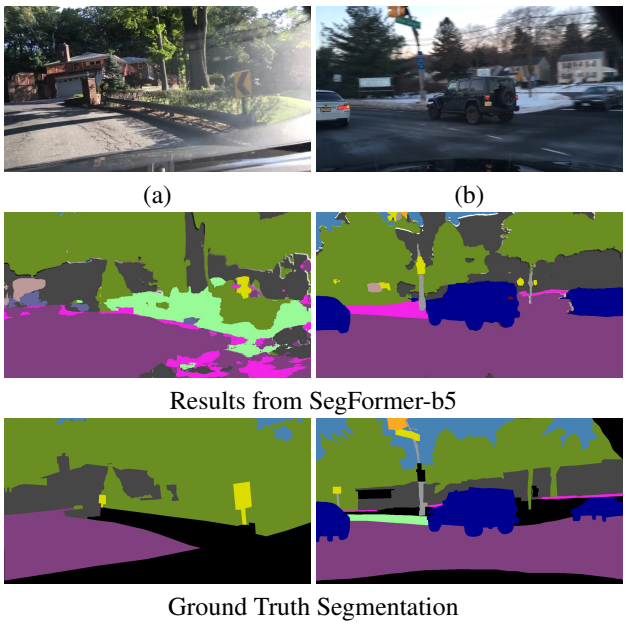

Figure 1. Demonstration of multiple weather degradations affecting the perceptual quality of an image at a local and global level. (a) Simultaneous solar glare and motion blur, (b) global illumination variation with localized motion blur. Utilizing these images directly for perception tasks results in poor prediction quality, as shown for semantic segmentation using BDD100K trained SegFormer-b5 [83] model.

need, practical deployment of such algorithms in the field has been limited.

Recently, existing works [67,74] adopt an end-to-end approach towards image restoration by training an underlying neural network to learn the mapping between images affected by single or unknown degradation combinations (motion blur, haze, rain, raindrop, and snow) and their clear counterparts. However, their problem definition either confines the input-output image pairs to static natural scenes or requires the generation of synthetic weather augmentations to produce noisy images from clean ones. While the former reduces the size of the training dataset and fails to capture diverse scene interactions, the latter introduces a performance gap [69,70] due to inaccuracies in degradation modeling. Furthermore, these restoration algorithms have demonstrated reduced performance when applied to images captured from different sources having distinct camera response functions compared to the training dataset. This highlights three significant challenges that hinder the practicality of learning-based generic image restoration algorithms: (1) the need to identify degradation profiles and corresponding restoration mechanisms, (2) sensitivity to different camera response functions and (3) inability to sufficiently restore corrupted regions of interest for consistent performance of performance algorithms. These challenges become more pronounced in scenarios with multiple global or local degradation profiles, as commonly encountered in driving scenes (see Fig. 1).

While perceptual aesthetics of restored images are desired from a human visual perception viewpoint, researchers have demonstrated the detrimental effects of poor image quality on downstream perception tasks such as feature matching [66], object detection [2, 60, 69, 72, 81, 82], semantic segmentation [19, 42, 57], depth estimation [62, 75, 77] and simultaneous localization and mapping [30]. Given that these tasks are crucial components in achieving a comprehensive scene understanding capability for modern driving assistance systems, it is desirable to have a general-purpose image restoration algorithm that generates perceptually appealing results and enhances the performance of downstream perception tasks. Thus, the ability to restore naturally degraded images benefits both drivers and driving assistance operations.

Considering the existence of degradations at global levels (e.g., illumination, rain, snow) and local levels (e.g., motion blur, glares), utilizing a general-purpose CNN for restoration would result in inefficient computations since these operations are applied globally. To address this, we propose a two-stage approach. In the first stage, we employ computationally inexpensive image processing operations that can be locally applied to remove mild to moderate local degradations focusing on the illumination of the image, such as brightness, contrast, saturation, and color variations. In the second stage, a lightweight CNN is deployed to leverage global semantics and texture information, enabling the recovery of occluded or missing details within an image. Furthermore, as we employ the CNN for the inpainting task, we identify mechanisms to reduce the computational footprint while maintaining consistent performance, resulting in a lightweight and practically viable framework. Since there is a lack of datasets capturing paired clean and degraded images, we utilize a synthetic degradation mechanism that utilizes instance maps combined with global illumination variations [47] and flare generation [21]. This approach ensures the complete framework is general-purpose and can simultaneously remove multiple degradations without excessive computational overhead.

We utilize perception and dense prediction metrics during optimization to ensure that the restored images are visually pleasing and do not adversely affect downstream perception tasks. This approach eliminates pixel variations that may lead to model sensitivities. Previous approaches [2, 60, 69, 81, 82] predominantly focused on coarse prediction tasks such as object detection, which exhibit robustness towards minor imperfections in the input. Consequently, images restored by these approaches improve object detector performance but may result in perceptually unpleasant outputs. In contrast, dense prediction tasks are sensitive to even minor imperfections, even within perceptually pleasing images. Therefore, incorporating dense prediction tasks during optimization ensures visually satisfying results and consistent performance in downstream dense prediction tasks.

Finally, we highlight that current image restoration tasks suffer performance limitations when evaluated on out-of-distribution datasets, as they are primarily optimized for images captured by similar devices. To overcome this limitation, we propose the utilization of multiple RGB-to-RAW CNNs, which convert input RGB images into camera-specific RAW format. Increasing the diversity of camera image-signal-processing (ISP) during training enhances the robustness of the proposed system to different image sources. We summarize our contributions as,

- We propose a two-stage local-global image restoration framework that disentangles image restoration into local image enhancement and global image restoration.
- To ensure local degradation diversity within training samples, we generate a combination of instance-aware local and global degradations.
- To ensure improved perceptual quality and prediction performance, we utilize dense prediction tasks during optimization.
- We highlight that current image restoration solutions are sensitive toward Camera ISP. Thus, we propose the utilization of multiple learnable inverse-ISP pipelines

to simulate the camera response functions of unique cameras.

2. Related Works

2.1. Enhanced Perceptual Image Restoration

A longstanding goal within the computer vision community is the ability to recover images affected by natural weather degradations such as fog, snow, rain, noise, and motion blur. This has fuelled widespread research wherein different degradation-specific architectures have been proposed. Specifically, tasks such as dehazing [15, 53, 73, 80, 88], deblurring [13,14,35,36,92,96], raindrop removal [52], deraining [3, 22, 27, 40, 54, 79, 86, 89], desnowing [5, 11, 12, 31,44,63], and noise removal [1,10,37,51,85,90,94] alongside mechanisms to improve their robustness [68, 87, 91] have been extensively studied. However, all leading approaches utilize an encoder-decoder architecture built using a CNN or, more recently, transformers. The objective behind such a framework design is to leverage mild/moderate affected regions to estimate the structural and textural properties of significantly degraded regions. While this approach restores images that are pleasant to human eyes and score well on perceptual metrics such as LPIPS [95] and NIQE [48]. We observe these algorithms as sensitive to camera response, limiting performance when deployed in the wild.

2.2. Perception Guided Image Restoration

Alternatively, different works have proposed to utilize the perception task within the training pipeline to ensure consistent performance of high-level perception tasks such as object detection [2, 60, 69, 81, 82]. Herein the perception model is fixed, and the underlying restoration algorithm is trained to enhance a degraded image to improve the detector performance. While prior approaches utilize object detection as the auxiliary optimization task, its coarse prediction nature makes object detectors less sensitive to minute pixel imperfections. In addition, utilizing an object detector reduces the scope of restoration as the underlying restoration algorithm is rewarded only to restore regions enclosing objects of interest while ignoring the remainder of the image. Hence we contend that utilizing object detector results in localized restoration determined by the labels associated with a given detector. Such a restoration mechanism is not unsuitable for scenarios requiring complete scene information.

2.3. Influence of Camera ISP

The purpose of a camera ISP is to convert raw digital signals of the scene to human-perceivable RGB images. The current camera ISP is designed manually and is primarily proprietary. Recent advances in image restoration by image restoration saw the development of a learning-based camera ISP to generate RGB images end-to-end. Despite the increasing attention, there is a lack of studies examining the influence of camera ISP on image restoration and perception. This issue is exacerbated by the restricted nature of the camera ISP, resulting in incorrect estimation of its influence on image restoration and perception as the building blocks within a given camera sensor still need to be discovered. Despite attempts to learn camera ISP in an end-to-end manner [20,23,28] or as a multi-part approach wherein each component is estimated such as tone-curve, white balance, exposure correction [16]. Such solutions are not generalizable as each distinct camera sensor has a unique ISP and response function. Hence we propose learning multiple inverse camera ISP pipelines corresponding to unique camera sensors. Given access to multiple camera ISP, we can easily estimate its influence on image restoration and perception.

2.4. General Purpose Image Restoration

Recently several works have highlighted the need for an end-to-end image restoration mechanism without requiring prior degradation information. Notably, GIQE [67] and TransWeather [74] are among the first works demonstrating the viability of such approaches. However, these approaches are computationally expensive and bound by the degradations that can be restored accurately. Specifically, TransWeather is designed for restoring Fog, Rain, and Snow, whereas GIQE can also restore low light and motion blur. Despite this, such approaches are computationally expensive and cannot account for multiple weather degradations that can co-exist within an image, thereby limiting the restoration quality. Furthermore, these approaches are unrestricted by the performance of prediction networks, thereby limiting the scope of image restoration.

3. Proposed Methodology

3.1. Non-Homogeneous and Non-Linear Degradation Generation

In the absence of real datasets capturing diverse driving conditions, there is a critical need for a mechanism to generate non-homogeneous and non-linear degradations. This mechanism is crucial in training image restoration algorithms to handle the complexities and variations encountered in real-world driving scenarios. By incorporating non-homogeneous and non-linear degradations, the degradation space is expanded, leading to a more holistic training approach and improved performance of the image restoration algorithm. Exposing the algorithm to a comprehensive set of non-homogeneous and non-linear degradations makes it more robust and capable of handling complex real-world scenarios. We construct the proposed mechanism by combining Instance Guided Degradation Generation, Illumination Change, and Flare Modeling augmentations. While the

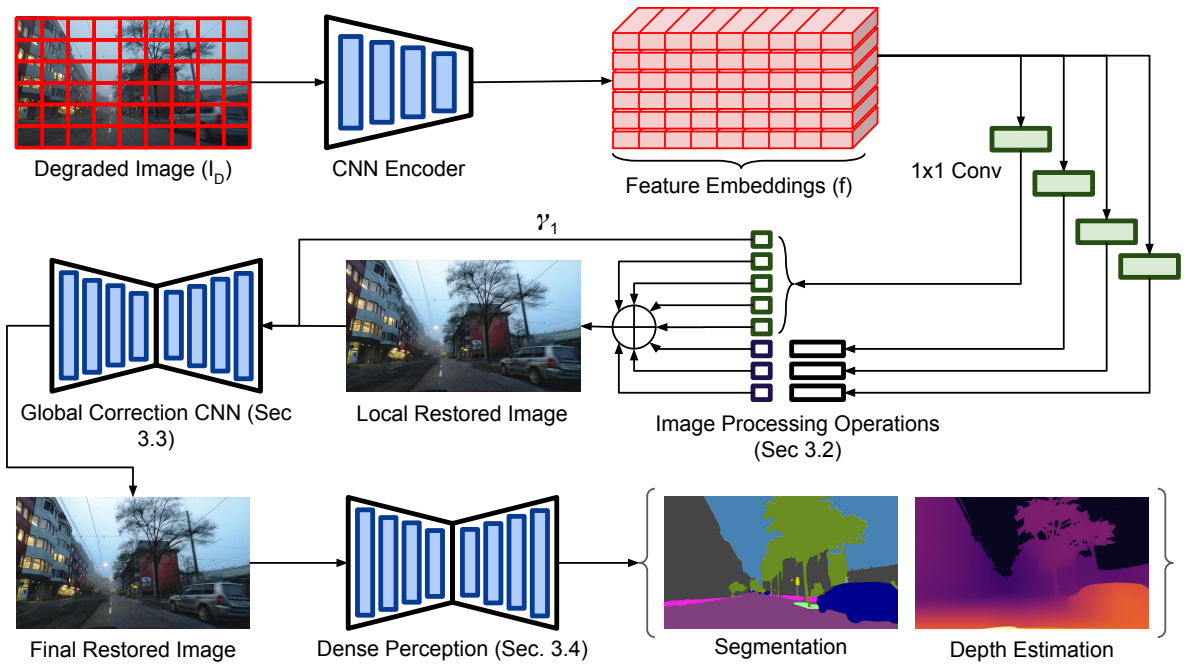

Figure 2. Overview of the proposed two-stage mechanism for restoring images captured in natural driving conditions. For clarity we omit different sub-operations and would redirect readers to corresponding subsections for further details.

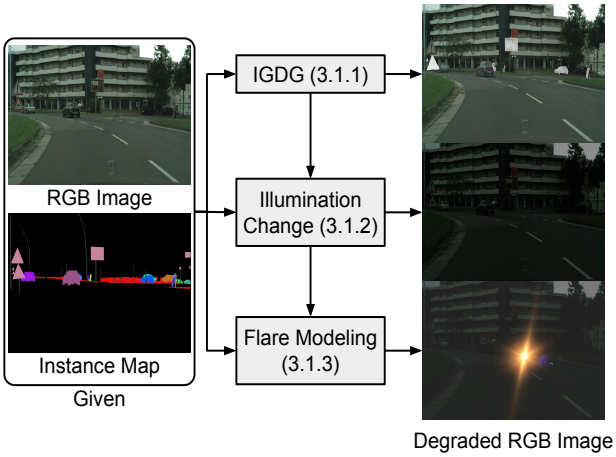

Figure 3. Illustration of the Non-Homogeneous and Non-Linear Degradation Generation mechanism. For simplicity we show direct application of different mechanism and not their multiple combinations.

proposed mechanism is similar to [24] in scope, i.e., data augmentation, the proposed mechanism focuses on generating realistic degradations to improve image restoration algorithm's performance. It considers the specific challenges and requirements of the image restoration domain, such as non-linearity and non-homogeneity in degradation distribution, to simulate realistic weather and environmental conditions. On the other hand, [24] is a method primarily aimed at augmenting datasets, for instance, segmentation tasks, by manipulating instance placements and backgrounds without explicitly modeling degradation effects. We illustrate the proposed mechanism in Fig. 3 and present the algorithm in Algo. 1.

3.1.1 Instance Guided Degradation Generation

We provide a summary of instance-guided degradation generation (IGDG), wherein a clear image (I_C) and an instance map (M) are the inputs. The global degradation space is defined by rain, snow, and fog, while the local degradation space is defined by motion blur, water droplets, and noise augmentations generated using [33]. Given these inputs, the algorithm iterates (i) over each instance and applies degradation if the probability value (p) is greater than 0.5 using a randomly sampled degradation type (d). This ensures the introduction of diverse and realistic degradation patterns uniquely affecting each instance. In order to introduce the non-linearity of the degradation, we use degradation order $(degradation\_order)$ that specifies the number of iterations the degradation type would be applied. As this is a data-preprocessing step, we limit the order of non-linearity to 3 to ensure a compute-bound. The algorithm iterates over the selected degradation order to introduce non-linearity of degradation. By allowing non-linear and non-homogeneous degradations, our algorithm significantly expands the degradation space, enabling a more holistic training of image restoration algorithms.

3.1.2 Global Illumination Change

For synthesizing low-light images, our approach draws inspiration from the methodology proposed by [47] wherein to emulate the characteristics of low-light conditions, a

combination of both linear and gamma transformations, involving three essential parameters: α, β, and γ, are utilized. The transformation is defined as follows:

$$I_{out} = \beta \times (\alpha \times I_D)^{\gamma} \tag{1}$$

The parameters α, β, and γ are dynamically determined during synthesis. We sample these parameters from a uniform distribution to introduce diversity and naturalness into the generated images. Specifically, α is uniformly sampled from the range $[0.5, 1]$, β from $[0.5, 1]$, and γ from $[1, 5]$. While this pipeline models the illumination changes, it does not account for noise distribution, which we include to obtain a faithful representation of low-light images via,

$$I_{out} = \beta \times (\alpha \times I_D)^{\gamma} + \eta \tag{2}$$

The term η represents image noise arising from photon shot noise, read noise, and quantization noise. Photon shot noise is modeled using a Poisson distribution [4, 25, 29], accounting for the randomness of photon detection. Read noise, which accounts for long-tailed distributions, is modeled with the Tukey lambda distribution [32]. Quantization noise, resulting from limited discrete levels, is modeled using a zero-mean Gaussian distribution. By incorporating these noise models into our image processing pipeline, we accurately capture noise sources such as photon shot noise, read noise, and quantization noise, ensuring the quality and fidelity of the final image.

3.1.3 Flare Generation

To generate realistic flares, we adopt the pipeline proposed by [21], which defines flares as a combination of scattering and reflective components. This approach enables the synthesis of authentic night-time flare effects. The paper generously provides access to a 5000 scattering flares and 2000 reflective flares dataset. These flares can be integrated into the training pipeline to introduce flare artifacts. By incorporating a wide range of flares in the training process, we ensure that our model can handle diverse degradation scenarios, resulting in a more robust and versatile flare synthesis system. The complete degradation pipeline is illustrated in Fig. 3.

Algorithm 1 Non-Linear and Non-Homogeneous Degradation Generation

Require: Clear Input Image (I_c)
Require: Instance map ($M \geq 1$)
Ensure: I_D: Generated degraded image

degradation_space = rain, snow, fog, motion blur, droplet, illumination, noise
$I_D \leftarrow I_C$
for i in M **do**
  $p \leftarrow random(0, 1)$
  $d \leftarrow random(degradation\_space)$
  $degradation\_order \leftarrow random(0, 3)$
  $r \leftarrow$ Extract_Instance(i)
  if $p > 0.5$ **then**
    for j in $degradation\_order$ **do**
      $I_D \leftarrow Apply\_Degradation(I_D, r, d)$
    end for
  end if
end for
if $p > 0.5$ **then**
  $I_D \leftarrow Apply\_Illumination\_Change(I_D, r, d)$
end if
if $p > 0.5$ **then**
  $I_D \leftarrow Apply\_Flare\_Modeling(I_D, r, d)$
end if
return I_D

3.2. Stage 1: Local Image Restoration

This section addresses the challenge of restoring corrupted images that undergo non-linear localized degradation, necessitating multiple image restoration operations. Although task-specific neural networks have shown promise in image restoration, their computational requirements make them impractical for real-time applications having multiple degradations. Moreover, these networks often perform global image restoration, including regions that do not require any restoration or enhancement. To overcome these limitations, we propose a novel approach for local image restoration and enhancement using traditional image processing operations, such as gamma correction, white balance adjustment, noise removal, sharpening, tone mapping, fog removal, and pixel corruption correction. With an emphasis on localized operations, our method efficiently restores and enhances specific regions of interest, improving image quality while avoiding unnecessary computational overhead.

We utilize a pre-trained Convolutional Neural Network (CNN) trained on ImageNet [56]. We adapt the architecture by removing the final pooling layer and adjusting the output channel dimension, defining the factor and weight for each image processing operation using 1x1 convolution. Here, factor refers to the intensity of the operation, and weight refers to the contribution of an operation toward final restoration. This modification allows us to perform patch-wise restoration, determined by the receptive field of the last convolutional layer. By removing the pooling layer, we preserve semantic information within each patch, enabling localized restoration capabilities for targeted enhancement and restoration of specific image regions. The patch-wise representation offers a comprehensive understanding of local image features, facilitating precise restoration operations. We visually illustrate our proposed mechanism in Fig. 3, showcasing the modified CNN architecture and the resulting patch-wise outputs from the last convolutional layer. This approach empowers us to perform local-level opera-

tions, preserving spatial information and achieving precise control over the restoration process, ultimately enhancing image quality. We would redirect the readers to Appendix-A of supplementary for summarization of different image processing operations.

3.3. Stage 2: Global Image Restoration

While localized operations can restore localized degradations, there are certain limitations, such as (1) boundary artifacts in the final restored image and (2) inadequate restoration when the degradations cover an area beyond the patch size, which is determined by the receptive field size of the underlying CNN, leading to inaccurate restoration results. To overcome these, we introduce a lightweight UNet [55] architecture in the second stage of our approach. This UNet is designed to recover and restore the partially restored image from the first stage. Moreover, we leverage the pixel corruption score generated in the first stage as an attention guidance mechanism, highlighting the areas needing restoration. Unlike complex encoder-decoder networks used in global image restoration approaches, which suffer from computational complexity and limited performance in low-light conditions, our two-stage network provides a generic image restoration algorithm with improved efficiency and restoration capabilities.

While a typical UNet architecture [55] can be utilized for image restoration tasks, the computational footprint associated with it may restrict its deployment on resource-constrained devices. To address this issue, we reexamine the construction blocks of UNet, such as the convolutional block and fusion layer, to develop a more compact solution. We introduce split convolutional (SC) blocks, which are constructed using point-wise and depth-wise convolutional layers, allowing us to capture diverse feature representations in a gated manner. This modification reduces the computational requirements and enables the network to maintain a high level of accuracy. Additionally, we adapt the SK (Selective Kernel) module, initially proposed in [39], by incorporating dimensional reduction and expansion. This modification reduces the parameter requirement while leveraging the GPU's parallelizability, reducing computation time. Finally, in our approach, we input both the pixel degradation map and the input image into this network, enabling us to obtain a final restored image that is globally coherent, with consistent object boundaries and reduced artifacts.

3.3.1 Split-Convolutional Block

Given an input feature ($f_{in} \in R^{N,C,H,W}$), we first downsample the feature space using pixel unshuffle [65] operation ($f_{out} \in R^{N,4C,H//2,W//2}$). The obtained feature map is then normalized using RescaleNorm [73] and split along channel dimension into two parts. One of the parts is processed using pointwise convolution (1×1) followed by sigmoid activation, while the other half is processed by pointwise and depthwise convolution (3×3). Subsequently, these parts are combined using an element-wise product. Thus, the first part acts as a gating mechanism. We then perform dimensional expansion followed by a pixel shuffling operation to upsample the aggregated feature map. We summarize the complete illustration of the process in Fig. 4.

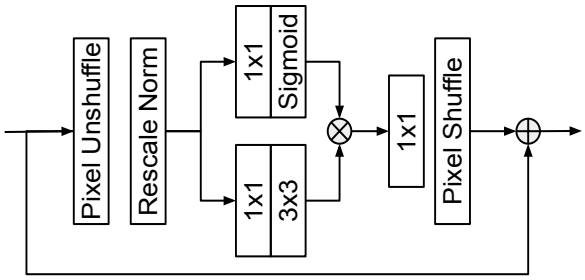

Figure 4. Illustration of the proposed Split-Convolutional Block as a Compute-Friendly alternative to standard convolutional layer.

3.3.2 Selective Kernel Fusion

Motivated by the Selective Kernel method proposed [39], we redesign the concatenation operation used to combine features from the encoder (f_1) and decoder (f_2) parts of the network. The encoder and decoder features are rich in spatial and semantic information, respectively. Thus, we propose a frequency selective fusion mechanism to effectively combine these different feature representations. This mechanism selectively fuses low and high-frequency information from the encoder and decoder. We employ the global average pooling operation to extract the low-frequency components from a feature map, following the approach described in Wang et al. [78]. However, our focus differs from theirs, as we aim to perform frequency-based feature concatenation. Given the input feature maps (f_1, f_2) to be fused, we first extract the low-frequency components using average pooling. Next, we subtract the low-frequency component from the original feature map to obtain the high-frequency component. In the frequency domain, we perform element-wise addition of the feature maps, followed by channel attention to identify relevant features within each channel. We employ a gating mechanism that utilizes fully connected layers to identify the relevant channels within the input feature maps. Subsequently, we fuse this channel attention with the input features to amplify the relevant features. Finally, the amplified features are aggregated using element-wise addition. We illustrate the modified Selective Kernel Fusion mechanism in Figure 5.

3.4. Dense Perception for Optimization

The proposed two-stage image restoration mechanism effectively recovers degraded images, producing visually

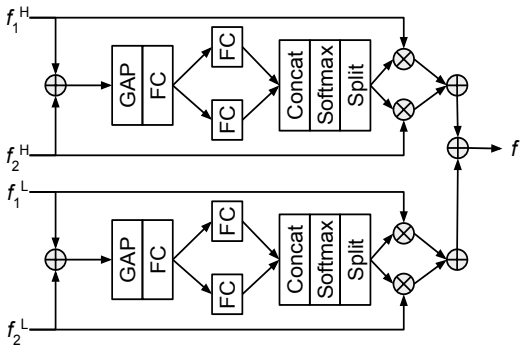

Figure 5. Overview of the proposed Selective Kernel Fusion block. Here we omit low and high frequency component extraction for simplicity.

pleasing results. However, the enhanced images may inadvertently introduce adversarial patterns that can degrade the performance of perception algorithms. To mitigate this issue, we delve into the analysis of error-causing areas, specifically focusing on the Image Signal Processing (ISP) pipeline [45] leading to Adversarial Restoration [69]. The ISP can influence the performance of the restoration algorithm as images captured from different camera sensors tend to alter the performance due to the bias of restoration networks to a particular camera response function. We present such examples in Appendix-B of the supplementary. Since this adversarial restoration affects the underlying perception tasks, we integrate such algorithms during the optimization process of the restoration network to ensure consistent performance. Furthermore, we introduce multiple camera response functions to expand the color space so that the underlying restoration network can restore images without any adversarial pattern.

We use RGB $\rightarrow$ RAW image translation networks [17, 84, 93] to obtain the camera response function of multiple sensors via a reverse ISP process. Such an approach allows us to capture camera-specific characteristics that might be inaccessible due to the proprietary nature of ISP or tolerances related to imaging sensors. Furthermore, we consider dense perception tasks, deviating from commonly used coarse perception tasks. We highlight that commonly used object detection algorithms are robust towards localized degradations, and therefore, such restoration algorithms cannot recover minute imperfections. On the contrary, dense perception tasks such as segmentation and depth estimation are sensitive toward minute localized imperfections, which may not be visible. Hence, utilization of such algorithms during optimization can provide benefits against generating adversarial patterns. With this motivation, we integrate segmentation and depth estimation as the dense prediction tasks. Since we utilize a synthetic degradation generation mechanism, we can use cityscapes pretrained segmentation and zero-shot depth estimation approach such as SegFormer [83] with Mix Transformer encoders (MiT) and ZoeDepth [7] for the generation of ground truth. Specifically, using MiT-b5 and BEiT [6] encoders, respectively, prioritizing accuracy over inference speed. Furthermore, to reduce computational overhead during the training stage, we use lightweight backbones such as MiT-b0 and LeViT [26] for segmentation and depth estimation, respectively.

3.5. Training Mechanism

The training process for our proposed framework follows a stage-wise approach. In the first stage, we train the model for 100 epochs using the Cityscapes dataset [18]. We initialize the learning rate to 2e-4 and employ the ADAM optimizer with parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$ [34]. To control the learning rate, we use cosine annealing [46], starting from 2e-4 and reducing it to 1e-6. The input image resolution is set to 512x512, and the batch size is set to 8. After completing the first stage's training, we jointly train both networks for an additional 300 epochs. The hyperparameter settings remain the same as in the first stage. To compute the optimization function during training, we combine several loss functions. These include L1 loss, Contrastive loss [80], and SSIM (Structural Similarity Index Measure) loss, which focuses on improving perceptual quality. Additionally, we incorporate Segmentation and Depth losses, which aim to enhance dense prediction quality. By combining these various loss functions, our training addresses both perceptual and dense prediction quality, resulting in a robust and comprehensive optimization approach.

$$L = \lambda_1 * L_1 + \lambda_2 * L_{Cons} + \lambda_3 * L_{SSIM} + \lambda_4 * L_{Seg} + \lambda_5 * L_{Depth}$$

Here L_{Seg} refers to the cross entropy loss and L_{Depth} refers to MSE loss. The weighting factors $(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$ are empirically set to 1.0, 0.1, 10, 1, 1.

4. Experimental Evaluation

4.1. Datasets and Evaluation Metrics

For training purposes, we utilize the Cityscapes [18] datasets and generate the synthetic degradations following the mechanisms proposed above. For comparison of restoration quality, we utilize pretrained weights of prior works GIQE [67], AirNet [38], Weather-Diffusion (WD) [49] and evaluate performance on datasets capturing adverse driving conditions such as ACDC [58], DENSE [8], NuScenes [9] and Radiate [64] datasets. While these datasets are meant for autonomous driving and provide accurate depth estimation results, along with their semantic extensions, they do not provide any reference clear ground truth image. Thus for quantitative evaluation of image restoration, we utilize no reference perceptual quality metrics such as NIQE [48] and CLIP-IQA [76]. In addition, we utilize standard performance metrics for segmentation and

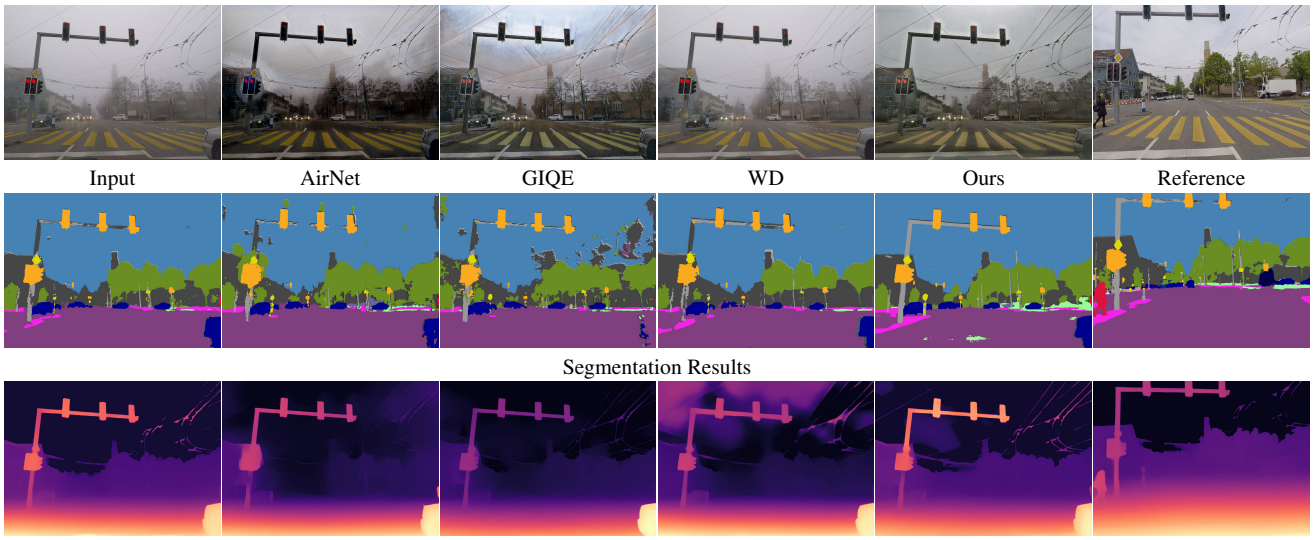

Figure 6. Impact of different general purpose image restoration algorithms on foggy image from ACDC dataset. In addition, performance of both segmentation and depth estimation algorithms using the restored images are included. Additional results and comparison with SoTA dehazing algorithms is provided in Appendix-F of Supplementary. It should be noted that since reference image is captured in degradation free conditions, there exists scene inconsistencies due to change in presence of dynamic objects.

depth estimation, i.e., mIOU (mean intersection over union) and RMSE (root mean square error), respectively.

4.2. Comparison with SoTA

We conducted a qualitative comparison of our proposed mechanism with several generic image restoration algorithms, including GIQE [67], AirNet [38], and WD [49]. The results are summarized in Fig. 6, showcasing the visual performance comparison using retrained variants of aforementioned algorithms using proposed mechanism to ensure fair comparison. To further evaluate the impact of the restored images on downstream perception tasks such as depth estimation and semantic segmentation, we present qualitative and quantitative results in Table 1.

Based on the qualitative and quantitative results, our observations are as follows: Firstly, poor weather conditions have a detrimental effect on the performance of downstream perception tasks, as the degraded images hinder the accuracy and reliability of these tasks. Secondly, while current generic image restoration algorithms successfully restore visually pleasing images, more than these restored images are needed to improve the performance of pretrained algorithms in downstream tasks s ignificantly. Merely achieving visual quality does not guarantee enhanced performance in perception tasks. Lastly, integrating perception tasks into the optimization process improves the perceptual quality of the restored images and the performance of downstream perception tasks. By considering the specific requirements and objectives of these tasks during the restoration process, we can enhance the overall quality and effectiveness of the restored images, leading to improved performance in subsequent perception tasks. We conduct further comparisons in Appendix-C of supplementary and ablation in Appendix-D, with additional qualitative results in Appendix-E.

Table 1. Quantitative performance of SoTA generic image restoration algorithm and its implication on downstream perception tasks.

Method	ACDC [58]	DENSE [8]
	NIQE / CLIP / mIOU	NIQE / CLIP-IQA / RMSE
Baseline	3.91 / 0.63 / 47.61	5.74 / 0.34 / 15.37
GridDehazeNet [43]	3.35 / 0.48 / 32.14	4.57 / 0.41 / 22.34
DeHamer [15]	3.68 / 0.46 / 34.98	3.67 / 0.76 / 15.13
TridentNet [41]	4.47 / 0.44 / 38.42	5.72 / 0.53 / 12.40
DA-Dehaze [61]	4.65 / 0.54 / 39.97	4.37 / 0.55 / 11.43
DIDH [71]	3.14 / 0.36 / 38.43	4.98 / 0.46 / 10.80
AECRNet [80]	2.46 / 0.25 / 39.15	2.65 / 0.54 / 11.03
GIQE [67]	2.85 / 0.59 / 53.42	3.49 / 0.51 / 10.67
AirNet [38]	2.99 / 0.77 / 47.65	4.59 / 0.76 / 14.39
WD [49]	3.58 / 0.79 / 42.42	4.21 / 0.68 / 12.96
Ours	2.99 / 0.61 / 55.97	3.54 / 0.51 / 10.95

5. Conclusion

Our two-stage mechanism for generic image restoration in natural driving conditions addresses the limitations of single neural network approaches. The first stage employs lightweight CNNs for dynamic restoration of nonlinear and non-homogeneous degradation profiles. The second stage uses an inpainting network to fill inadequately restored regions by considering global semantic and structural information. Integration of dense perception tasks enhances the perceptual quality of the restored images. We utilize synthetic data augmentation to overcome the lack of diverse degradation datasets. Experimental results on images captured in adverse weather conditions demonstrate improved perceptual quality and downstream task performance. Importantly, our approach offers computational efficiency, making it suitable for real-time applications.

References

[1] Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown. A high-quality denoising dataset for smartphone cameras. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1692–1700, 2018. 3

[2] Muhammad Ahmed, Khurram Azeem Hashmi, Alain Pagani, Marcus Liwicki, Didier Stricker, and Muhammad Zeshan Afzal. Survey and performance analysis of deep learning based object detection in challenging environments. *Sensors*, 21(15):5116, 2021. 2, 3

[3] Yunhao Ba, Howard Zhang, Ethan Yang, Akira Suzuki, Arnold Pfahnl, Chethan Chinder Chandrappa, Celso de Melo, Suya You, Stefano Soatto, Alex Wong, and Achuta Kadambi. Not just streaks: Towards ground truth for single image deraining. In *ECCV*, 2022. 3

[4] Richard L Baer. A model for dark current characterization and simulation. In *Sensors, Cameras, and Systems for Scientific/Industrial Applications VII*, volume 6068, pages 37–48. SPIE, 2006. 5

[5] Chris H Bahnsen and Thomas B Moeslund. Rain removal in traffic surveillance: Does it matter? *IEEE Transactions on Intelligent Transportation Systems*, 20(8):2802–2819, 2018. 3

[6] Hangbo Bao, Li Dong, Songhao Piao, and Furu Wei. Beit: Bert pre-training of image transformers. *arXiv preprint arXiv:2106.08254*, 2021. 7

[7] Shariq Farooq Bhat, Reiner Birkl, Diana Wofk, Peter Wonka, and Matthias Müller. Zoedepth: Zero-shot transfer by combining relative and metric depth. *arXiv preprint arXiv:2302.12288*, 2023. 7

[8] Mario Bijelic, Tobias Gruber, Fahim Mannan, Florian Kraus, Werner Ritter, Klaus Dietmayer, and Felix Heide. Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 7, 8

[9] Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. *arXiv preprint arXiv:1903.11027*, 2019. 7

[10] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3291–3300, 2018. 3

[11] Wei-Ting Chen, Hao-Yu Fang, Jian-Jiun Ding, Cheng-Che Tsai, and Sy-Yen 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*, pages 754–770. Springer, 2020. 3

[12] Wei-Ting Chen, Hao-Yu Fang, Cheng-Lin Hsieh, Cheng-Che Tsai, I Chen, Jian-Jiun Ding, Sy-Yen Kuo, et al. All snow removed: Single image desnowing algorithm using hierarchical dual-tree complex wavelet representation and contradict channel loss. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4196–4205, 2021. 3

[13] Sung-Jin Cho, Seo-Won Ji, Jun-Pyo Hong, Seung-Won Jung, and Sung-Jea Ko. Rethinking coarse-to-fine approach in single image deblurring. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4641–4650, 2021. 3

[14] Xiaojie Chu, Liangyu Chen, and Wenqing Yu. Nafssr: Stereo image super-resolution using nafnet. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 1239–1248, June 2022. 3

[15] Saeed Anwar Runmin Cong Wenqi Ren Chongyi Li Chun-Le Guo, Qixin Yan. Image dehazing transformer with transmission-aware 3d position embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 3, 8

[16] Marcos V Conde, Steven McDonagh, Matteo Maggioni, Ales Leonardis, and Eduardo Pérez-Pellitero. Model-based image signal processors via learnable dictionaries. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 481–489, 2022. 3

[17] Marcos V Conde, Radu Timofte, Yibin Huang, Jingyang Peng, Chang Chen, Cheng Li, Eduardo Pérez-Pellitero, Fenglong Song, Furui Bai, Shuai Liu, et al. Reversed image signal processing and raw reconstruction. aim 2022 challenge report. In *European Conference on Computer Vision*, pages 3–26. Springer, 2022. 7

[18] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016. 7

[19] Dengxin Dai, Christos Sakaridis, Simon Hecker, and Luc Van Gool. Curriculum model adaptation with synthetic and real data for semantic foggy scene understanding. *International Journal of Computer Vision*, 128:1182–1204, 2020. 2

[20] Linhui Dai, Xiaohong Liu, Chengqi Li, and Jun Chen. Awnet: Attentive wavelet network for image isp. In *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, pages 185–201. Springer, 2020. 3

[21] Yuekun Dai, Chongyi Li, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Flare7k: A phenomenological nighttime flare removal dataset. In *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2022. 2, 5

[22] Sen Deng, Mingqiang Wei, Jun Wang, Yidan Feng, Luming Liang, Haoran Xie, Fu Lee Wang, and Meng Wang. Detail-recovery image deraining via context aggregation networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14560–14569, 2020. 3

[23] Steven Diamond, Vincent Sitzmann, Frank Julca-Aguilar, Stephen Boyd, Gordon Wetzstein, and Felix Heide. Dirty pixels: Towards end-to-end image processing and perception. *ACM Transactions on Graphics (TOG)*, 40(3):1–15, 2021. 3

[24] Golnaz Ghiasi, Yin Cui, Aravind Srinivas, Rui Qian, Tsung-Yi Lin, Ekin D Cubuk, Quoc V Le, and Barret Zoph. Simple copy-paste is a strong data augmentation method for instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2918–2928, 2021. 4

[25] Ryan D Gow, David Renshaw, Keith Findlater, Lindsay Grant, Stuart J McLeod, John Hart, and Robert L Nicol. A comprehensive tool for modeling cmos image-sensor-noise performance. *IEEE Transactions on Electron Devices*, 54(6):1321–1329, 2007. 5

[26] Benjamin Graham, Alaaeldin El-Nouby, Hugo Touvron, Pierre Stock, Armand Joulin, Hervé Jégou, and Matthijs Douze. Levit: a vision transformer in convnet's clothing for faster inference. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 12259–12269, 2021. 7

[27] Qing Guo, Jingyang Sun, Felix Juefei-Xu, Lei Ma, Xiaofei Xie, Wei Feng, and Yang Liu. Efficientderain: Learning pixel-wise dilation filtering for high-efficiency single-image deraining. In *AAAI*, 2021. 3

[28] Saumya Gupta, Diplav Srivastava, Umang Chaturvedi, Anurag Jain, and Gaurav Khandelwal. Del-net: A single-stage network for mobile camera isp. *arXiv preprint arXiv:2108.01623*, 2021. 3

[29] Glenn E Healey and Raghava Kondepudy. Radiometric ccd camera calibration and noise estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(3):267–276, 1994. 5

[30] Sungchul Hong, Pranjay Shyam, Antyanta Bangunharcana, and Hyuseoung Shin. Robotic mapping approach under illumination-variant environments at planetary construction sites. *Remote Sensing*, 14(4):1027, 2022. 2

[31] Da-Wei Jaw, Shih-Chia Huang, and Sy-Yen Kuo. Desnowgan: An efficient single image snow removal framework using cross-resolution lateral connection and gans. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(4):1342–1350, 2020. 1, 3

[32] Brian L Joiner and Joan R Rosenblatt. Some properties of the range in samples from tukey's symmetric lambda distributions. *Journal of the American Statistical Association*, 66(334):394–399, 1971. 5

[33] Alexander B. Jung. imgaug. https://github.com/aleju/imgaug, 2018. [Online; accessed 30-Oct-2018]. 4

[34] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 7

[35] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiri Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. *ArXiv e-prints*, 2017. 3

[36] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2019. 3

[37] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. *arXiv preprint arXiv:1803.04189*, 2018. 3

[38] Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-in-one image restoration for unknown corruption. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17452–17462, 2022. 7, 8

[39] Xiang Li, Wenhai Wang, Xiaolin Hu, and Jian Yang. Selective kernel networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 510–519, 2019. 6

[40] Yuanchu Liang, Saeed Anwar, and Yang Liu. Drt: A lightweight single image deraining recursive transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 589–598, June 2022. 3

[41] Jing Liu, Haiyan Wu, Yuan Xie, Yanyun Qu, and Lizhuang Ma. Trident dehazing network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 430–431, 2020. 8

[42] Kunhua Liu, Zihao Ye, Hongyan Guo, Dongpu Cao, Long Chen, and Fei-Yue Wang. Fiss gan: A generative adversarial network for foggy image semantic segmentation. *IEEE/CAA Journal of Automatica Sinica*, 8(8):1428–1439, 2021. 2

[43] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Griddehazenet: Attention-based multi-scale network for image dehazing. In *ICCV*, pages 7314–7323, 2019. 8

[44] Yun-Fu Liu, Da-Wei Jaw, Shih-Chia Huang, and Jenq-Neng Hwang. Desnownet: Context-aware deep network for snow removal. *IEEE Transactions on Image Processing*, 27(6):3064–3073, 2018. 3

[45] William Ljungbergh, Joakim Johnander, Christoffer Petersson, and Michael Felsberg. Raw or cooked? object detection on raw images. In *Scandinavian Conference on Image Analysis*, pages 374–385. Springer, 2023. 7

[46] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016. 7

[47] Feifan Lv, Yu Li, and Feng Lu. Attention guided low-light image enhancement with a large scale low-light simulation dataset. *International Journal of Computer Vision*, 129(7):2175–2193, 2021. 2, 4

[48] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012. 3, 7

[49] Ozan Özdenizci and Robert Legenstein. Restoring vision in adverse weather conditions with patch-based denoising diffusion models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–12, 2023. 7, 8

[50] Cuong Cao Pham and Jae Wook Jeon. Efficient image sharpening and denoising using adaptive guided image filtering. *IET Image Processing*, 9(1):71–79, 2015. 1

[51] Tobias Plotz and Stefan Roth. Benchmarking denoising algorithms with real photographs. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1586–1595, 2017. 3

[52] Rui Qian, Robby T. Tan, Wenhan Yang, Jiajun Su, and Jiaying Liu. Attentive generative adversarial network for raindrop removal from a single image. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 3

[53] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 11908–11915, 2020. 1, 3

[54] Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. Progressive image deraining networks: A better and simpler baseline. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019. 3

[55] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015. 6

[56] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115:211–252, 2015. 5

[57] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Semantic foggy scene understanding with synthetic data. *International Journal of Computer Vision*, 126:973–992, 2018. 2

[58] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Acdc: The adverse conditions dataset with correspondences for semantic driving scene understanding. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10765–10775, 2021. 7, 8

[59] John GM Schavemaker, Marcel JT Reinders, Jan J Gerbrands, and Eric Backer. Image sharpening by morphological filtering. *Pattern Recognition*, 33(6):997–1012, 2000. 1

[60] Mark Schutera, Mostafa Hussein, Jochen Abhau, Ralf Mikut, and Markus Reischl. Night-to-day: Online image-to-image translation for object detection within autonomous driving by night. *IEEE Transactions on Intelligent Vehicles*, 6(3):480–489, 2020. 2, 3

[61] Yuanjie Shao, Lerenhan Li, Wenqi Ren, Changxin Gao, and Nong Sang. Domain adaptation for image dehazing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2808–2817, 2020. 8

[62] Aashish Sharma, Loong-Fah Cheong, Lionel Heng, and Robby T Tan. Nighttime stereo depth estimation using joint translation-stereo learning: Light effects and uninformative regions. In *2020 International Conference on 3D Vision (3DV)*, pages 23–31. IEEE, 2020. 2

[63] Neeraj Sharma, Vijay Kumar, and Sunil Kumar Singla. Single image defogging using deep learning techniques: past, present and future. *Archives of Computational Methods in Engineering*, 28:4449–4469, 2021. 3

[64] Marcel Sheeny, Emanuele De Pellegrin, Saptarshi Mukherjee, Alireza Ahrabian, Sen Wang, and Andrew Wallace. Radiate: A radar dataset for automotive perception in bad weather. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1–7. IEEE, 2021. 7

[65] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883, 2016. 6

[66] Pranjay Shyam, Antyanta Bangunharcana, and Kyung-Soo Kim. Retaining image feature matching performance under low light conditions. In *2020 20th International Conference on Control, Automation and Systems (ICCAS)*, pages 1079–1085. IEEE, 2020. 2

[67] Pranjay Shyam, Kyung-Soo Kim, and Kuk-Jin Yoon. Giqe: Generic image quality enhancement via nth order iterative degradation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2077–2087, 2022. 2, 3, 7, 8

[68] Pranjay Shyam, Sandeep Singh Sengar, Kuk-Jin Yoon, and Kyung-Soo Kim. Evaluating copy-blend augmentation for low level vision tasks. *arXiv preprint arXiv:2103.05889*, 2021. 3

[69] Pranjay Shyam, Sandeep Singh Sengar, Kuk-Jin Yoon, and Kyung-Soo Kim. Lightweight hdr camera isp for robust perception in dynamic illumination conditions via fourier adversarial networks. *arXiv preprint arXiv:2204.01795*, 2022. 2, 3, 7

[70] Pranjay Shyam and HyunJin Yoo. Data efficient single image dehazing via adversarial auto-augmentation and extended atmospheric scattering model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 227–237, 2023. 2

[71] Pranjay Shyam, Kuk-Jin Yoon, and Kyung-Soo Kim. Towards domain invariant single image dehazing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 9657–9665, 2021. 8

[72] Pranjay Shyam, Kuk-Jin Yoon, and Kyung-Soo Kim. Weakly supervised approach for joint object and lane marking detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2885–2895, 2021. 2

[73] Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision transformers for single image dehazing. *IEEE Transactions on Image Processing*, 32:1927–1941, 2023. 3, 6

[74] Jeya Maria Jose Valanarasu, Rajeev Yasarla, and Vishal M. Patel. Transweather: Transformer-based restoration of images degraded by adverse weather conditions, 2021. 2, 3

[75] Madhu Vankadari, Sourav Garg, Anima Majumder, Swagat Kumar, and Ardhendu Behera. Unsupervised monocular depth estimation for night-time images using adversarial domain feature adaptation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVIII 16*, pages 443–459. Springer, 2020. 2

[76] Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *AAAI*, 2023. 7

[77] Kun Wang, Zhenyu Zhang, Zhiqiang Yan, Xiang Li, Baobei Xu, Jun Li, and Jian Yang. Regularizing nighttime weirdness: Efficient self-supervised monocular depth estimation in the dark. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 16055–16064, 2021. 2

[78] Peihao Wang, Wenqing Zheng, Tianlong Chen, and Zhangyang Wang. Anti-oversmoothing in deep vision transformers via the fourier domain analysis: From theory to practice. *arXiv preprint arXiv:2203.05962*, 2022. 6

[79] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17683–17693, June 2022. 1, 3

[80] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive learning for compact single image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10551–10560, 2021. 3, 7, 8

[81] Yirui Wu, Haifeng Guo, Chinmay Chakraborty, Mohammad Khosravi, Stefano Berretti, and Shaohua Wan. Edge computing driven low-light image dynamic enhancement for object detection. *IEEE Transactions on Network Science and Engineering*, 2022. 2, 3

[82] Yuxuan Xiao, Aiwen Jiang, Jihua Ye, and Ming-Wen Wang. Making of night vision: Object detection under low-illumination. *IEEE Access*, 8:123075–123086, 2020. 2, 3

[83] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems*, 34:12077–12090, 2021. 1, 7

[84] Yazhou Xing, Zian Qian, and Qifeng Chen. Invertible image signal processing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6287–6296, 2021. 7

[85] Jun Xu, Yuan Huang, Ming-Ming Cheng, Li Liu, Fan Zhu, Zhou Xu, and Ling Shao. Noisy-as-clean: Learning self-supervised denoising from corrupted image. *IEEE Transactions on Image Processing*, 29:9316–9329, 2020. 3

[86] Yuntong Ye, Changfeng Yu, Yi Chang, Lin Zhu, Xi-Le Zhao, Luxin Yan, and Yonghong Tian. Unsupervised deraining: Where contrastive learning meets self-similarity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5821–5830, June 2022. 3

[87] Jaejun Yoo, Namhyuk Ahn, and Kyung-Ah Sohn. Rethinking data augmentation for image super-resolution: A comprehensive analysis and a new strategy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8375–8384, 2020.

[88] Yankun Yu, Huan Liu, Minghan Fu, Jun Chen, Xiyao Wang, and Keyan Wang. A two-branch neural network for non-homogeneous dehazing via ensemble learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 193–202, June 2021. 3

[89] Yi Yu, Wenhan Yang, Yap-Peng Tan, and Alex C Kot. Towards robust rain removal against adversarial attacks: A comprehensive benchmark analysis and beyond. *arXiv preprint arXiv:2203.16931*, 2022. 3

[90] Zongsheng Yue, Hongwei Yong, Qian Zhao, Deyu Meng, and Lei Zhang. Variational denoising network: Toward blind noise modeling and removal. *Advances in neural information processing systems*, 32, 2019. 3

[91] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. Cutmix: Regularization strategy to train strong classifiers with localizable features. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6023–6032, 2019. 3

[92] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*, 2022. 3

[93] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Cycleisp: Real image restoration via improved data synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2696–2705, 2020. 7

[94] Kai Zhang, Yawei Li, Jingyun Liang, Jiezhang Cao, Yulun Zhang, Hao Tang, Radu Timofte, and Luc Van Gool. Practical blind denoising via swin-conv-unet and data synthesis. *arXiv preprint*, 2022. 3

[95] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 3

[96] Wenbin Zou, Mingchao Jiang, Yunchen Zhang, Liang Chen, Zhiyong Lu, and Yi Wu. Sdwnet: A straight dilated network with wavelet transformation for image deblurring. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pages 1895–1904, October 2021. 3