

Article

Efficient Image Restoration for Autonomous Vehicles and Traffic Systems: A Knowledge Distillation Approach to Enhancing Environmental Perception [†]

Yongheng Zhang

State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications (BUPT), No. 10 Xitucheng Road, Haidian District, Beijing 100876, China; zhangyongheng@bupt.edu.cn

[†] This article is a revised and expanded version of a conference paper entitled Soft Knowledge Distillation with Multi-Dimensional Cross-Net Attention for Image Restoration Models Compression, which was presented at ICASSP2025, Hyderabad, India, 6–11 April 2025.

Abstract

Image restoration tasks such as deraining, deblurring, and dehazing are crucial for enhancing the environmental perception of autonomous vehicles and traffic systems, particularly for tasks like vehicle detection, pedestrian detection and lane line identification. While transformer-based models excel in these tasks, their prohibitive computational complexity hinders real-world deployment on resource-constrained platforms. To bridge this gap, this paper introduces a novel Soft Knowledge Distillation (SKD) framework, designed specifically for creating highly efficient yet powerful image restoration models. Our core innovation is twofold: first, we propose a Multi-dimensional Cross-Net Attention (MCA) mechanism that allows a compact student model to learn comprehensive attention relationships from a large teacher model across both spatial and channel dimensions, capturing fine-grained details essential for high-quality restoration. Second, we pioneer the use of a contrastive learning loss at the reconstruction level, treating the teacher's outputs as positives and the degraded inputs as negatives, which significantly elevates the student's reconstruction quality. Extensive experiments demonstrate that our method achieves a superior trade-off between performance and efficiency, notably enhancing downstream tasks like object detection. The primary contributions of this work lie in delivering a practical and compelling solution for real-time perceptual enhancement in autonomous systems, pushing the boundaries of efficient model design.

Keywords: knowledge distillation; multi-dimensional cross-net attention; image restoration; object detection



Academic Editors: Selene Tomassini and M. Ali Akber Dewan

Received: 28 September 2025

Revised: 22 October 2025

Accepted: 23 October 2025

Published: 24 October 2025

Citation: Zhang, Y. Efficient Image Restoration for Autonomous Vehicles and Traffic Systems: A Knowledge Distillation Approach to Enhancing Environmental Perception. *Computers* **2025**, *14*, 459. <https://doi.org/10.3390/computers14110459>

Copyright: © 2025 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Image restoration—encompassing tasks such as deraining, deblurring, and dehazing—stands as a critical frontline defense for the safety and reliability of autonomous vehicles and traffic systems. These tasks are particularly important in challenging environmental conditions where image quality is often degraded by factors such as inclement weather, poor lighting, and low visibility, leading to perceptual results such as poor object detection, as shown in Figure 1a. Effective restoration of these images is essential for autonomous vehicles to perform critical tasks, including vehicle detection, pedestrian

detection, and lane line identification, which are foundational to the safety and reliability of autonomous driving systems [1].

While transformer-based encoder–decoder models [2–4] have demonstrated exceptional performance in image restoration tasks, their computational demands—particularly in terms of FLOPs (floating-point operations), memory, and energy usage—make them less suitable for resource-constrained environments such as autonomous vehicles and traffic systems. These platforms typically operate under stringent resource limitations, where processing power, memory, and energy are at a premium. This constraint makes it infeasible to deploy computationally expensive models, as illustrated in Figure 1b. This creates a critical deployment gap: the very models that excel at restoring clarity are often unsuitable for the resource-constrained hardware of autonomous vehicles where power, memory, and processing capabilities are severely limited.

In response to this challenge, model compression techniques have gained attention as a way to reduce computational complexity while retaining the high performance of state-of-the-art image restoration models. Knowledge distillation, first introduced by Hinton et al. [5], has emerged as an effective method for model compression. In this approach, a lightweight student model is trained to emulate the behavior of a larger, more powerful teacher model, with various distillation strategies targeting different learning objectives such as logits, feature maps, and reconstruction outputs. Over time, distillation techniques have evolved to incorporate advanced methods like intermediate features [6,7], attention matrices [8,9], and instance relationships [10,11].

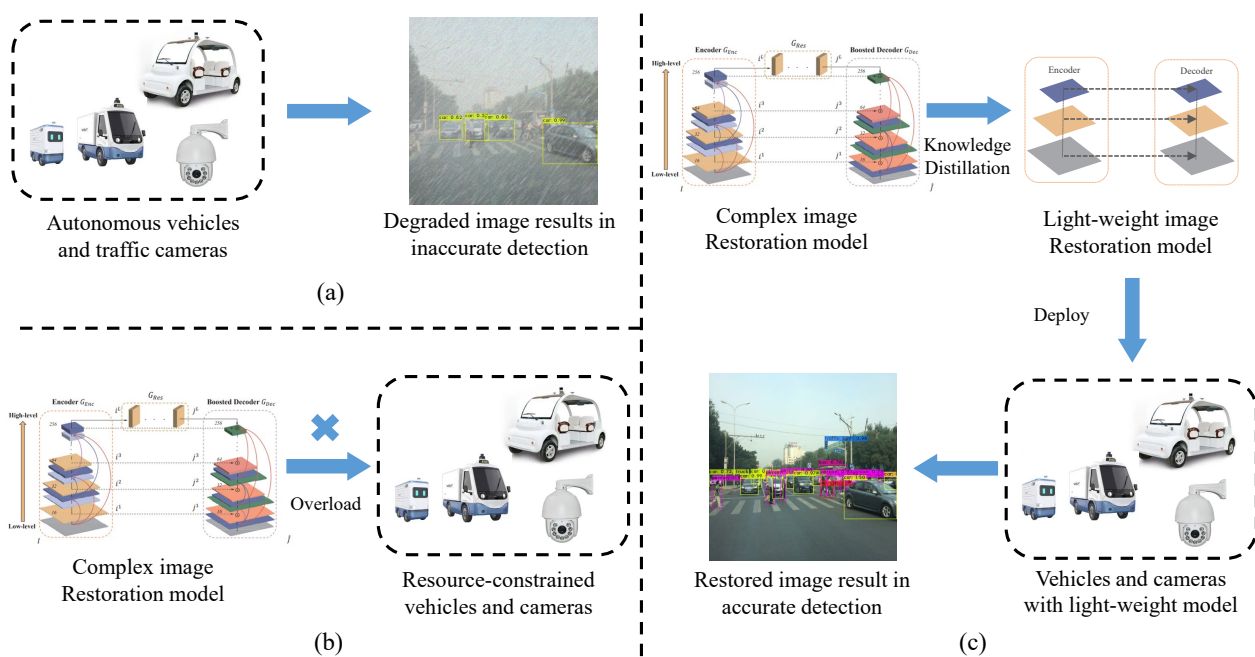


Figure 1. The pipeline for efficient image restoration model deployment on autonomous vehicles and traffic systems. (a) Degraded input and poor detection: Raw image under adverse weather conditions and its failed object detection result. (b) Computational bottleneck: A large, high-performance model cannot be deployed on the edge device due to resource constraints. (c) Efficient deployment via SKD: The deployed lightweight student model, distilled via our SKD framework, produces a clear image and enables accurate object detection.

While these techniques have been successful in domains like classification and detection, their application to image restoration has remained less explored. Existing distillation strategies for image restoration typically involve direct mimicking of teacher features [12–14] or reconstruction outputs [15,16]. However, these methods often overlook

the importance of implicit attention relationships between teacher and student networks. Such attention relationships are critical for preserving fine-grained details, such as edges and textures, which are vital for high-quality image restoration. Furthermore, traditional distillation approaches can suffer from instability due to the direct imitation of teacher features, limiting their scalability to resource-constrained platforms like autonomous vehicles.

To address these limitations, we propose a novel Soft Knowledge Distillation (SKD) strategy, synergistically integrated with a Multi-Dimensional Cross-Net Attention (MCA) mechanism. This work is specifically designed to bridge the critical performance–efficiency gap for image restoration in autonomous systems. The principal novelty and contributions of our work are fourfold:

1. **A Holistic distillation framework:** We introduce a comprehensive SKD framework that moves beyond conventional feature mimicking. Unlike prior arts that often focus on a single aspect, our method orchestrates knowledge transfer at multiple levels, deliberately designed to equip a compact student model with the robust capabilities of a large teacher model, while being inherently suitable for resource-constrained platforms.
2. **Multi-dimensional attention transfer:** A cornerstone of our innovation is the MCA mechanism. It enables the student to learn implicit, fine-grained attention relationships from the teacher across both spatial and channel dimensions. This is a significant departure from methods that directly mimic features, allowing the student to intrinsically understand what and where to attend for recovering critical details like edges and textures, which are paramount for high-quality restoration and subsequent detection tasks.
3. **Stable and efficient feature guidance:** To circumvent the instability of direct feature imitation, we introduce a Gaussian kernel function to guide the student’s learning in a transformed kernel space. This novel approach ensures a more stable and efficient optimization process, enhancing the reliability of the distillation under strict computational budgets.
4. **Contrastive reconstruction learning:** We pioneer a paradigm shift at the reconstruction level by replacing traditional pixel-wise loss functions with a contrastive learning loss. Here, the teacher’s high-quality reconstructions serve as positive anchors, explicitly pulling the student’s output towards them while pushing away from the degraded inputs. This strategy fundamentally boosts the perceptual quality of the student’s restorations, a critical factor for improving downstream task accuracy.

Our proposed SKD framework is not merely an incremental improvement but a cohesive solution that rethinks how knowledge is distilled for image restoration. Through extensive experiments on deraining, deblurring, and dehazing, as well as downstream detection tasks, we show that our approach significantly reduces computational complexity without sacrificing restoration quality, as shown in Figure 2. More importantly, we provide concrete evidence that our method leads to significant gains in critical downstream perception tasks, such as vehicle and pedestrian detection, thereby directly enhancing the safety and reliability of autonomous systems. These results underscore the potential of our SKD framework for real-time deployment, offering a practical and effective path forward for intelligent transportation systems.

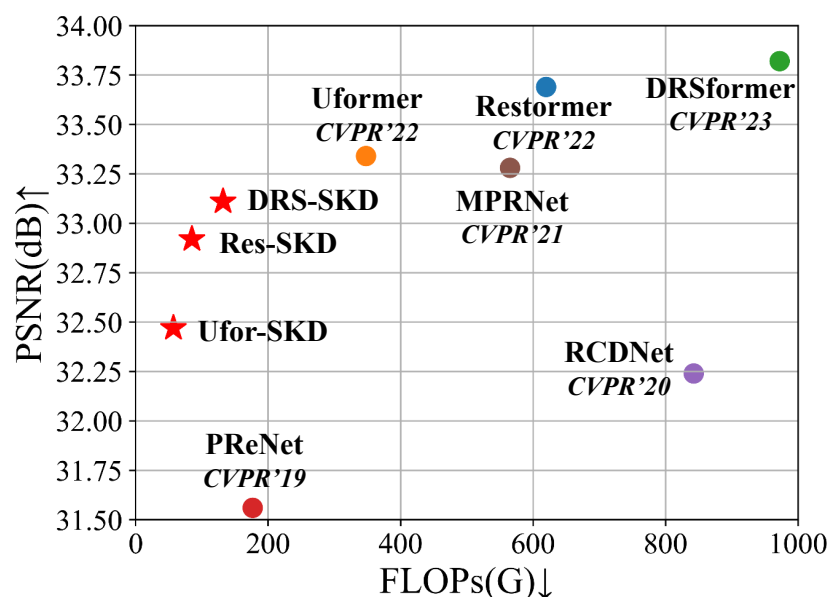


Figure 2. Performance–efficiency trade-off (PSNR ↑ vs. FLOPs ↓) of different methods on deraining dataset. Our distilled student model (marked in red) achieves restoration quality (PSNR) competitive with state-of-the-art (SOTA) models while requiring significantly lower computational complexity (FLOPs), demonstrating the effectiveness of our proposed distillation framework.

2. Related Work

In this section, we review relevant works related to the core components of our research, specifically image restoration, knowledge distillation, and contrastive learning.

2.1. Image Restoration Methods

2.1.1. Prior-Based Methods

Prior-based image restoration techniques rely on physical scattering models and statistical priors to estimate parameters such as transmission maps, atmospheric light intensity, and other image characteristics based on clear image statistics [17–23]. For example, He et al. [17] proposed a dehazing method based on the dark channel prior (DCP), which estimates transmission maps and global atmospheric light. Subsequent works [21,24–26] extended the DCP approach to other image restoration tasks. Additionally, Li et al. [23] developed a layer prior method to effectively remove rain streaks from images. Chen et al. [27] introduced a unified prior loss committee that combines several well-established physical priors to improve restoration performance.

While prior-based methods are effective in certain scenarios, they often face limitations when dealing with more complex or diverse degradation types.

2.1.2. CNN-Based Methods

Convolutional neural networks (CNNs) have revolutionized image restoration by learning complex features and restoration functions directly from data. These methods have achieved state-of-the-art performance across various restoration tasks [28–34]. For instance, Dong et al. [28] introduced a Multi-Scale Boosted Dehazing Network built on the U-Net architecture, which improves dehazing performance through dense feature fusion. Shao et al. [30] proposed an end-to-end image translation and dehazing network, which achieved improved results for both tasks. Similarly, Zamir et al. [34] developed a multi-stage architecture that progressively learns restoration functions tailored to degraded inputs, enhancing overall restoration performance.

CNN-based methods have greatly advanced image restoration, enabling the learning of complex features and producing high-quality restored images across a wide range of degradation types.

2.1.3. Transformer-Based Methods

The integration of Vision Transformers (ViTs) into image restoration tasks has gained considerable attention due to their ability to capture long-range dependencies, which are essential for understanding the global context and relationships within an image. Recent works have demonstrated the effectiveness of ViTs in various restoration tasks [2–4,35–37]. For example, Wang et al. [3] proposed Uformer, a hierarchical encoder–decoder network utilizing Vision Transformer blocks, which effectively captures both local and global dependencies. Zamir et al. [2] introduced Restormer, a Transformer-based model that incorporates key design innovations in multi-head attention and feed-forward networks, allowing it to capture long-range pixel interactions while remaining computationally efficient. Similarly, Chen et al. [4] proposed DRSformer, which incorporates a learnable top-k selection operator within the transformer block, enhancing the model’s focus on the most critical features during restoration.

Transformer-based methods, such as Restormer, Uformer, and DRSformer, have demonstrated strong performance in image restoration by leveraging self-attention mechanisms to capture long-range dependencies and focus on important features, making them particularly effective for handling a wide range of image degradations.

2.2. Knowledge Distillation

Knowledge distillation [5] is a technique that transfers knowledge from a large teacher model to a smaller student network. The process involves training the student on a transfer set, using the soft target distribution provided by the teacher model. Both the teacher’s outputs and intermediate representations can significantly enhance the training process, improving the final performance of the student network.

This technique has been widely applied across high-level computer vision tasks such as object detection [38], face recognition [39], and semantic segmentation [40]. More recently, knowledge distillation has been employed in image restoration tasks [41,42], where it helps guide smaller models in restoring images degraded by noise, blur, or other distortions. These studies demonstrate how distillation can enhance the performance of lightweight models in challenging restoration tasks.

Knowledge distillation is a powerful technique that transfers both outputs and intermediate representations from a teacher model to a smaller student model, boosting performance across various computer vision and image restoration tasks.

2.3. Contrastive Learning

Contrastive learning is a self-supervised technique that constructs positive and negative pairs around an anchor point, encouraging the model to pull positive samples closer and push negative ones farther apart in the representation space [43]. Traditionally used in high-level tasks such as image classification and natural language processing, contrastive learning has recently been extended to low-level vision tasks, including image-to-image translation [44,45], deraining [46,47], and dehazing [43,48,49].

For instance, Zheng et al. [49] proposed curricular contrastive regularization for image dehazing, which focuses on achieving consensus within the contrastive space, providing advantages over non-consensual strategies. Ye et al. [46] introduced a non-local contrastive learning mechanism specifically designed for image deraining. This approach exploits the intrinsic self-similarity of images, enhancing deraining performance by effectively applying contrastive learning principles.

Contrastive learning, by encouraging the differentiation of positive and negative samples in the representation space, has proven effective in low-level vision tasks, such as deraining and dehazing, where it enhances performance through self-supervised learning mechanisms.

3. Methods

3.1. Overall Pipeline

The core objective of our Soft Knowledge Distillation (SKD) framework is to enable a compact student network to acquire the powerful image restoration capabilities of a large teacher model, not through direct feature imitation, but by learning the teacher’s underlying “reasoning” process and perceptual standards. This is achieved via a dual-pathway strategy encompassing both feature-level and image-level guidance. It is important to note that the proposed SKD framework is architecture-agnostic. In this work, to fairly demonstrate the effectiveness of our distillation method, we construct the student models as compact counterparts of their teachers by reducing the number of layers and feature channels while retaining the same fundamental building blocks. This ensures that any performance gain can be primarily attributed to the distillation process itself rather than to architectural modifications. The overall process is illustrated in Figure 3, where a degraded image $I \in \mathcal{R}^{H \times W \times 3}$ is input into both networks. The teacher network, which is pre-trained and equipped with a sophisticated architecture, excels at restoring clean images from degraded inputs by effectively mitigating various degradation factors such as noise, blur, and weather-related distortions. In contrast, the student model is designed to be smaller and more efficient by reducing the number of transformer layers and feature channels in its residual blocks, making it suitable for resource-constrained environments like autonomous vehicles or other traffic systems. The distillation process is carried out at two levels: feature-level distillation and image-level distillation.

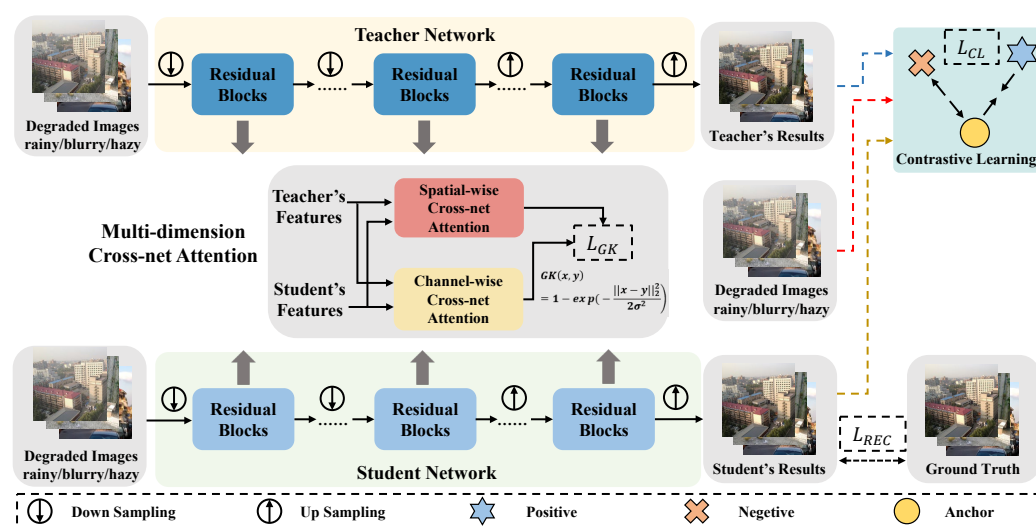


Figure 3. The overall architecture of the proposed Soft Knowledge Distillation (SKD) framework. The system comprises a pre-trained, computationally heavy teacher model and a lightweight student model. The core of our method is the Multi-Dimensional Cross-Net Attention (MCA) module, which facilitates knowledge transfer across both spatial and channel dimensions. Simultaneously, features are projected into a Gaussian kernel space to enable stable and efficient feature learning. At the output level, a contrastive reconstruction loss is employed, pulling the student’s restoration closer to the teacher’s high-quality output while pushing it away from the degraded input.

Feature-level learning is accomplished through our proposed Multi-Dimensional Cross-Network Attention (MCA) mechanism. The intermediate features of the student

network interact with features of corresponding blocks in the teacher network, allowing the student to implicitly absorb the attention knowledge embedded within the teacher. The resulting student and teacher features are then mapped to Gaussian kernel space, and the loss is computed based on their distance, enabling stable and efficient knowledge transfer.

At the image level, in addition to the reconstruction loss computed with ground truth, contrastive learning helps the student further refine its output. The student's reconstructed image uses the teacher's output as a positive example, aligning closely with it, while multiple original degraded images serve as negative examples, encouraging divergence from these degraded instances. What is more, we also utilize a feature extractor for both the student and teacher's reconstruction images and conduct contrastive learning on corresponding features.

These components of the SKD strategy work together to indirectly but significantly enhance the efficiency and stability of the student network during the distillation process, distinguishing our approach from direct imitation methods.

3.2. Multi-Dimension Cross-Net Attention

A central component of the SKD strategy is the Multi-Dimensional Cross-Network Attention (MCA) mechanism, which facilitates the transfer of attention knowledge from the teacher to the student by interacting across both channel-wise and spatial-wise dimensions. In traditional distillation methods, features from the teacher and student networks are directly compared. However, the MCA mechanism enhances this process by enabling a more sophisticated interaction between these features, allowing the student to benefit from richer attention knowledge, as shown in Figure 4.

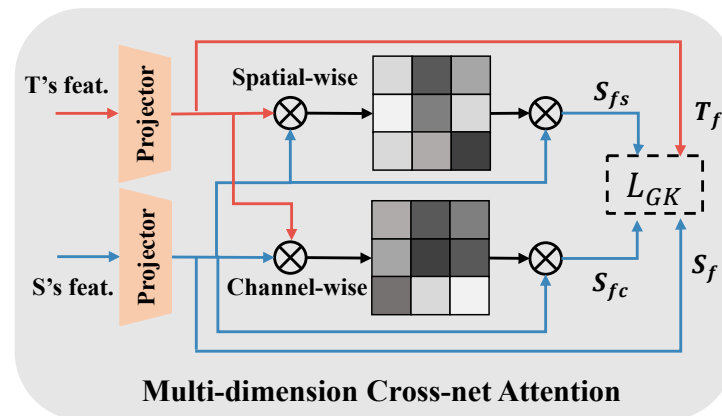


Figure 4. Architecture of the proposed Multi-Dimension Cross-Net Attention (MCA) module. The MCA enables cross-net interaction between the teacher and student models by independently operating on two dimensions: (1) the spatial-wise attention path, which exchanges and highlights important spatial locations (the “where”), and (2) the channel-wise attention path, which exchanges and recalibrates significant feature channels (the “what”). These two complementary attention streams are subsequently fused to achieve comprehensive knowledge transfer.

Given the features from corresponding blocks in the teacher and student networks, we first use projectors to map these features into a unified dimensional space, represented as T_f^i and S_f^i . The interaction process, which yields the updated student features S_{fc}^i (channel) and S_{fs}^i (spatial), can be expressed as:

$$S_{fc}^i = \text{softmax}(T_f^i \cdot (S_f^i)^T / \lambda) \cdot S_f^i, \quad (1)$$

$$S_{fs}^i = S_f^i \cdot \text{softmax}((T_f^i)^T \cdot S_f^i / \lambda), \quad (2)$$

where λ is an optional temperature factor defined by $\lambda = \sqrt{d}$.

To quantify the similarity between the teacher and student features, we map both feature sets to a Gaussian kernel space, where the distance between the features is computed. This approach is motivated by its superior stability and effectiveness compared to direct distance measures like Euclidean or Cosine similarity. The Euclidean distance $D_{eucl} = \|x - y\|_2$ is sensitive to the absolute magnitude of features and can lead to unstable optimization. In contrast, the Gaussian kernel measures relative proximity in a normalized, bounded space, providing a smoother gradient and greater robustness to outliers. While σ denotes the width of Gaussian kernel function, the Gaussian kernel distance and loss are defined as

$$GK(x, y) = 1 - \exp\left(-\frac{\|x - y\|_2^2}{2\sigma^2}\right), \quad (3)$$

$$L_{GK} = \sum_{i=1}^n (GK(S_f^i, T_f^i)) + \alpha 1(GK(S_{fc}^i, T_f^i) + GK(S_{fs}^i, T_f^i)). \quad (4)$$

3.3. Contrastive Learning for Knowledge Distillation

To further enhance the student's image restoration capabilities, we integrate contrastive learning into the SKD framework. Contrastive learning has been extensively used in representation learning [50] and image restoration [43,46] tasks, where it encourages the model to bring similar instances closer together while pushing dissimilar instances further apart. In our approach, this technique is applied to both the reconstructed images and the corresponding features during the distillation process (Figure 5).

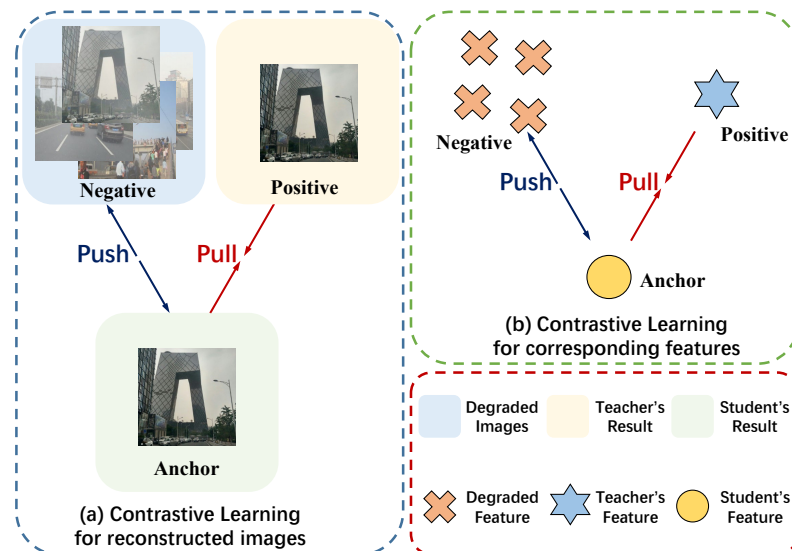


Figure 5. Schematic of the contrastive learning in SKD. The learning occurs at both the image and feature levels (using VGG-19 [51]), where the student's outputs are pulled toward the teacher's (positive) and pushed away from the degraded inputs (negative).

In the contrastive learning setup, the student's reconstructed images and their corresponding extracted features serve as anchors, while the teacher's output acts as the positive example. A batch of degraded images and their corresponding features serve as negative examples. By minimizing the distance between the anchor and positive examples, while maximizing the distance from the negative examples, the student network is guided to improve its ability to reconstruct clean images. The contrastive learning loss L_{CL} is composed of two components: the image contrastive learning loss L_{ICL} and feature contrastive learning loss L_{FCL} :

$$L_{ICL}(S_r, T_r, I) = -\log \frac{\text{sim}(S_r, T_r)}{\text{sim}(S_r, T_r) + \sum_{q=1}^b \text{sim}(S_r, I_q)} \quad (5)$$

$$L_{FCL}(S_r, T_r, I) = -\log \frac{\text{sim}(\phi(S_r), \phi(T_r))}{\text{sim}(\phi(S_r), \phi(T_r)) + \sum_{q=1}^b \text{sim}(\phi(S_r), \phi(I_q))}, \quad (6)$$

$$L_{CL}(S_r, T_r, I) = L_{ICL}(S_r, T_r, I) + L_{FCL}(S_r, T_r, I), \quad (7)$$

where S_r , T_r , I represent the student's output, the teacher's output (positive sample), and the degraded images (negative samples), respectively. The batch size is denoted by b , and $\text{sim}(u, v) = \exp\left(\frac{u^T v}{\|u\| \|v\| \tau}\right)$ measures the similarity between two feature vectors, with τ as the temperature parameter and $\phi(\cdot)$ representing a feature extraction operation using VGG-19 [51].

3.4. Overall Loss

The final loss function combines the three components discussed above: reconstruction loss, Gaussian kernel loss, and contrastive learning loss. The reconstruction loss measures the difference between the student's output S_r and ground truth G is formulated as

$$L_{REC} = \|G - S_r\|_1. \quad (8)$$

The overall loss function is then defined as

$$L = L_{REC} + \alpha_2 L_{GK} + \alpha_3 L_{CL}, \quad (9)$$

where α_2 and α_3 are trade-off weights that balance the contributions of the Gaussian kernel loss and contrastive learning loss. These weights are tuned to ensure that each component of the loss function contributes optimally to the distillation process.

4. Results

4.1. Implementation Details

We evaluate the proposed Soft Knowledge Distillation (SKD) strategy using a comprehensive set of eight datasets across three distinct image restoration tasks: deraining, deblurring, and dehazing. These tasks are critical for real-time traffic applications, where image quality must be restored under challenging environmental conditions.

For deraining, we use three datasets: the synthetic Rain1400 [52], Test1200 [53], and the real-world SPA [29]. For deblurring, we utilize the synthetic Gopro [54], HIDE [55], and the real-world BLUR-J [56]. Finally, for dehazing, we adopt the synthetic subset OTS and real-world subset RTTS from the RESIDE dataset [57].

To quantitatively assess the image restoration quality, we use a combination of full-reference and no-reference evaluation metrics. Specifically, for the synthetic datasets, we employ the Peak Signal-to-Noise Ratio (PSNR) [58] (in dB) and Structural Similarity Index (SSIM) [59], which measure pixel-level fidelity and structural similarity, respectively. For real-world datasets, where ground-truth references are unavailable, we use two no-reference metrics, the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [60] and Perception-based Image Quality Evaluator (PIQE) [61], which assess image quality based on perceptual and spatial features.

In addition to image quality, we evaluate the model complexity by measuring the Floating Point Operations (FLOPs) and inference time on each 512×512 image. These metrics are critical for evaluating the feasibility of our SKD strategy in resource-constrained

environments such as autonomous vehicles, where computational efficiency is paramount. We highlight the best results in bold and underline the sub-optimal results.

The entire SKD framework is implemented using PyTorch 1.10. We adopt Adam as the optimizer, with a temperature parameter set to $\tau = 1 \times 10^{-6}$. The trade-off weights used in the SKD loss function are $\alpha_1 = 0.5$, $\alpha_2 = 0.2$, and $\alpha_3 = 0.2$. The student models are trained for 100 epochs with a batch size of 8. The learning rate starts at 2×10^{-4} and is reduced gradually to 1×10^{-6} using cosine annealing [62] to ensure stable convergence. During training, all input images are randomly cropped into 128×128 patches, and pixel values are normalized to the range $[-1, 1]$. This augmentation strategy helps improve generalization and reduces the risk of overfitting, which is especially important when training on diverse and complex datasets like those used in this study.

For the teacher networks, we select three state-of-the-art transformer-based models known for their performance in image restoration tasks: Restormer [2], Uformer [3], and DRSformer [4]. These models were chosen due to their ability to capture long-range dependencies and complex features in the image restoration process. The architectures of these teacher models vary in terms of depth and dimensions:

- Restormer: 4, 6, 6, 8 layers per encoder–decoder level, with a feature dimension of 48.
- Uformer: 1, 2, 8, 8 layers, with a feature dimension of 32.
- DRSformer: 4, 4, 6, 6, 8 layers, with a feature dimension of 48.

The corresponding student models—Res-SKD, Ufor-SKD, and DRS-SKD—compress the hyper-parameters to more resource-efficient configurations:

- Res-SKD: 1, 2, 2, 4 layers, with a feature dimension of 32.
- Ufor-SKD: 1, 2, 4, 4 layers, with a feature dimension of 16.
- DRS-SKD: 2, 2, 2, 2, 4 layers, with a feature dimension of 32.

To ensure reproducibility, we clarify that the student models maintain the same fundamental architectural blueprint as their respective teachers (e.g., the same type and sequence of transformer blocks), with compression achieved solely by reducing the number of layers in each block and the dimensionality of the feature channels. It is worth noting that while our experiments utilize transformer-based models to construct a strong benchmark, the proposed SKD framework is not bound to any specific network family. The distillation mechanism operates on feature maps and output images, making it readily applicable to other architectures, such as CNNs. The exploration of its efficacy with CNN-based teachers is a valuable direction for future work. This compression results in significant reductions in both FLOPs and parameters, achieving an average reduction of 85.4% and 85.8%, respectively. Despite the reduced complexity, the student models retain much of the performance of their teacher counterparts, as demonstrated in the following experimental results.

4.2. Comparisons with State-of-the-Arts

Comparison with knowledge distillation methods. We begin by comparing our Soft Knowledge Distillation (SKD) strategy with two state-of-the-art (SOTA) image-to-image transfer knowledge distillation methods: Wave [16] and DCKD [13]. These methods are popular in the field of image restoration, particularly in distilling complex teacher models into simpler, more efficient student models. We present both qualitative and quantitative comparisons in Figure 6 and Table 1.

For the deraining task, results are averaged over the Rain1400 [52] and Test1200 [53] datasets, while for deblurring, results are averaged across the Gopro [54] and HIDE [55] datasets. In the dehazing task, we evaluate using the SOTS subset of the RESIDE dataset [57]. Our SKD-based model demonstrates a clear advantage over the other two SOTA methods in both visual quality and full-reference evaluation metrics (PSNR and SSIM). In particular,

the restoration quality produced by SKD exhibits fewer artifacts, sharper edges, and more realistic details in both synthetic and real-world degradation scenarios.



Figure 6. Qualitative results of knowledge distillation methods.

Table 1. Quantitative results of knowledge distillation methods across three tasks. The best results are highlighted in bold.

Tasks	Deraining		Deblurring		Dehazing	
Metrics	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Restormer [2]	33.69	0.935	32.07	0.952	36.51	0.991
Res-Wave [16]	32.27	0.918	30.36	0.933	34.77	0.981
Res-DCKD [13]	32.42	0.921	30.46	0.937	34.81	0.983
Res-SKD	32.92	0.927	31.07	0.946	35.06	0.985
Uformer [3]	33.34	0.931	31.98	0.960	36.28	0.989
Ufor-Wave [16]	32.10	0.917	30.50	0.938	34.62	0.971
Ufor-DCKD [13]	32.13	0.918	30.50	0.937	34.85	0.979
Ufor-SKD	32.47	0.921	31.14	0.940	34.97	0.983
DRSformer [4]	33.82	0.937	31.97	0.949	36.10	0.988
DRS-Wave [16]	32.51	0.922	30.30	0.931	34.67	0.975
DRS-DCKD [13]	32.58	0.922	30.43	0.936	34.81	0.980
DRS-SKD	33.11	0.927	31.09	0.945	34.94	0.983

Comparison with image restoration methods. In addition to comparing with knowledge distillation methods, we also benchmark our SKD strategy against seven image restoration models, including two for deraining (PReNet [63] and RCDNet [64]), two for deblurring (DMPHN [65] and MT-RNN [66]), two for dehazing (MSBDN [28] and PSD [27]), and one for generalized restoration (MPRNet [34]).

As shown in Figure 7 and Table 2, our SKD-based student models significantly outperform the baselines in terms of computational complexity while maintaining competitive image quality and performance metrics. Despite being more lightweight, the SKD models are able to restore images with a similar level of detail and clarity as the much more complex models such as MPRNet, which is known for its robust performance across various restoration tasks. This demonstrates the efficiency and potential of SKD in achieving high-quality image restoration with reduced resource consumption, making it particularly suitable for deployment in resource-constrained environments.



Figure 7. Qualitative comparison with lightweight methods.

Table 2. Quantitative comparison (PSNR $\uparrow$ /SSIM $\uparrow$) and efficiency metrics (FLOPs $\downarrow$ /inference time $\downarrow$) with lightweight methods across deraining, deblurring, and dehazing tasks. Note that some compared methods are specialized for a single task, hence the missing entries, whereas general-purpose models (including ours) are evaluated on all tasks. PSD [27] employs an identical network architecture to MSBDN [28] during inference. The best and second-best results are highlighted in bold and underline, respectively.

Tasks	Deraining	Deblurring	Dehazing	FLOPs	Infer Time
PReNet [63] Std.	31.56/0.914 (0.10/0.003)	-/-	-/-	192.1 G	0.0764 s
RCDNet [64] Std.	32.24/0.918 (0.12/0.003)	-/-	-/-	842.5 G	0.1919 s
DMPHN [65] Std.	-/-	30.14/0.932 (0.15/0.003)	-/-	780.4 G	1.694 s

Table 2. Cont.

Tasks	Deraining	Deblurring	Dehazing	FLOPs	Infer Time
MT-RNN [66] Std.	-/-	30.15/0.931 (0.14/0.004)	-/-	579.0 G	0.0387 s
MSBDN [28] Std.	-/-	-/-	33.79/0.984 (0.09/0.002)	617.4 G	0.1446 s
PSD [27] Std.	-/-	-/-	33.21/0.972 (0.11/0.003)	617.4 G	0.1446 s
MPRNet [34] Std.	33.28/0.927 (0.10/0.003)	31.81/0.949 (0.13/0.004)	31.31/0.971 (0.12/0.003)	565.0 G	0.0593 s
Res-SKD Std.	32.92/ 0.927 (0.11/0.003)	31.08/ <u>0.946</u> (0.14/0.003)	35.06/0.985 (0.08/0.002)	85.0 G	0.0356 s
Ufor-SKD Std.	32.47/0.921 (0.08/0.002)	31.14/0.940 (0.12/0.004)	<u>34.97/0.983</u> (0.09/0.002)	57.1 G	0.0540 s
DRS-SKD Std.	<u>33.11/0.927</u> (0.10/0.003)	31.09/0.945 (0.15/0.004)	34.94/0.983 (0.10/0.002)	132.0 G	0.0599 s

Comparison on real degraded images. In real-world applications, it is crucial to ensure that the model performs effectively on images that are not only synthetically degraded but also subjected to real-world conditions. To this end, we extend our evaluation to real degraded images, as shown in Figure 8 and Table 3.

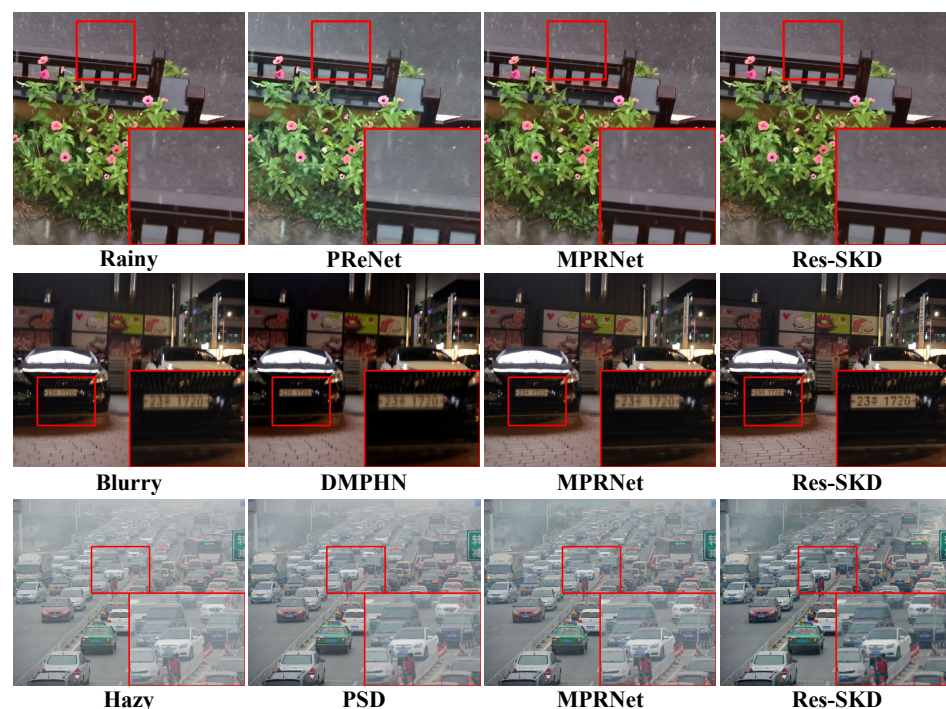


Figure 8. Restoration results on real-world datasets: SPA [29] (deraining), BLUR-J [56] (deblurring), and RTTS [57] (dehazing).

Although the models are primarily trained on synthetic datasets, our SKD-based Res-SKD student model exhibits strong performance in handling multiple degradations in real-world images. The results in Figure 8 clearly demonstrate that Res-SKD can effectively mitigate rain, blur, and haze artifacts, yielding visually appealing and restored images that retain important fine-grained details. Furthermore, the no-reference evaluation metrics presented in Table 3 (BRISQUE and PIQE) confirm that the SKD-based student models

achieve satisfactory image quality, reinforcing their applicability in practical scenarios where ground-truth references are unavailable.

Table 3. Quantitative comparison (BRISQUE ↓/PIQE ↓) and efficiency metrics (FLOPs ↓/inference time ↓) with lightweight methods across three real-world datasets. Note that some compared methods are specialized for a single task, hence the missing entries, whereas general-purpose models (including ours) are evaluated on all tasks. PSD [27] employs an identical network architecture to MSBDN [28] during inference. The best and second-best results are highlighted in bold and underline, respectively.

Dataset	SPA [29]	BLUR-J [56]	RTTS [57]	FLOPs	Infer Time
PReNet [63] Std.	59.64 / 53.18 (1.25/1.85)	-/-	-/-	192.1 G	0.0764 s
RCDNet [64] Std.	59.62 / 51.23 (1.30/1.72)	-/-	-/-	842.5 G	0.1919 s
DMPHN [65] Std.	-/-	70.42/50.33 (1.62/1.84)	-/-	780.4 G	1.694 s
MT-RNN [66] Std.	-/-	63.92/53.32 (1.45/1.88)	-/-	579.0 G	0.0387 s
MSBDN [28] Std.	-/-	-/-	28.74/50.66 (0.97/1.61)	617.4 G	0.1446 s
PSD [27] Std.	-/-	-/-	<u>25.24</u> /30.63 (0.86/1.33)	617.4 G	0.1446 s
MPRNet [34] Std.	76.63/46.76 (1.21/1.45)	64.643/ 36.165 (1.42/1.21)	36.15/38.70 (1.12/1.57)	565.0 G	0.0593 s
Res-SKD Std.	46.53/ <u>44.15</u> (1.07/1.36)	44.37 / <u>42.78</u> (0.93/1.18)	24.76 / <u>27.93</u> (0.88/1.23)	85.0 G	0.0356 s
Ufor-SKD Std.	<u>45.83</u> /46.27 (1.03/1.42)	<u>47.37</u> /45.83 (1.11/1.19)	26.37/ <u>29.52</u> (0.92/1.17)	57.1 G	0.0540 s
DRS-SKD Std.	43.12 / 42.62 (0.98/1.27)	47.61/43.92 (1.14/1.33)	26.15/32.72 (0.84/1.33)	132.0 G	0.0599 s

Our findings highlight the versatility and robustness of the proposed SKD strategy, which excels not only on synthetic data but also when applied to real-world challenges. This is a key strength for the deployment of image restoration models in real-time autonomous vehicle applications and other resource-limited environments.

4.3. Object Detection Results

Object detection is a critical high-level computer vision task for traffic systems. However, the accuracy of object detection is heavily influenced by the quality of captured images. In complex outdoor environments, various degradation factors, including weather conditions, significantly degrade image quality, leading to reduced object detection performance.

To evaluate the impact of image restoration on object detection, we use the lightweight object detection model Yolov4 [67] as a downstream task and assess performance using the mean Average Precision (mAP) metric. We test on the synthetic RMTD-test dataset [68] and the real-world RMTD-real dataset [68], both of which include images with multiple coexisting degradation factors.

The visualization results in Figure 9 clearly demonstrate the benefit of using restored images. Object detection performance on images restored by our SKD-based Res-SKD model exhibits significantly higher recall and precision compared to low-quality images. Additionally, the detection results on Res-SKD-restored images outperform those of images restored by MPRNet [34], and approach the accuracy levels achieved on high-quality images.

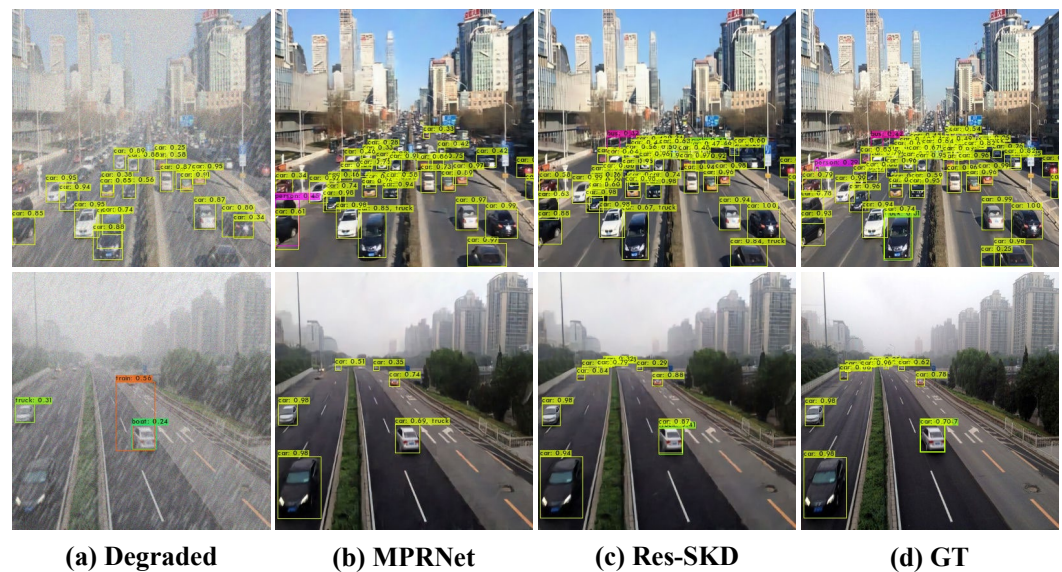


Figure 9. Visualization of object detection results on RMTD [68].

Quantitative evaluations in Table 4 corroborate these observations, showing a significant improvement in mAP scores for images restored using Res-SKD. These results highlight the practical utility of our SKD strategy in enhancing the effectiveness of downstream tasks like object detection, particularly in autonomous vehicle applications where degraded image quality is a common challenge.

Table 4. Object detection results in mAP $\uparrow$ on RMTD [68]. The best results are highlighted in bold.

Datasets	RMTD-Test [68]	RMTD-Real [68]
Degraded	0.1053	0.4772
RCDNet [64]	0.2648	0.4864
DMPHN [65]	0.2701	0.4892
MPRNet [34]	0.2836	0.5071
Res-SKD	0.2977	0.5146
Ground Truth	0.3461	-

4.4. Ablation Studies

Ablation studies were conducted on the Gopro [54] dataset for the deblurring task to validate the effectiveness of our proposed distillation strategy and to analyze the contribution of its individual components. A critical baseline is established by training the student architecture solely with the reconstruction loss (L_{REC}) and without employing any knowledge distillation, which serves to isolate the impact of our SKD framework. The results, summarized in Table 5 and Figure 10, reveal the following:

- Student with L_{REC} only: This setup reveals the native capacity of the compact student architecture. Its significantly lower performance, compared to all subsequent configurations that utilize the teacher network, clearly demonstrates that the performance gains are not merely a consequence of using a smaller model, but are primarily attributable to the proposed distillation method.
- Channel-wise attention: Adding this mechanism improves the student model's learning capacity, yielding a 0.41 dB gain in PSNR.
- Spatial-wise attention: This mechanism contributes an additional 0.51 dB gain in PSNR.

- Full Multi-Dimensional Cross-Net Attention (MCA): When both channel-wise and spatial-wise attention mechanisms are combined, the model achieves a 0.79 dB increase in PSNR and a 0.009 improvement in SSIM over the baseline model.
- Contrastive learning loss (L_{CL}): This loss further enhances the model's performance, adding a 0.25 dB gain in PSNR and a 0.004 improvement in SSIM.

The qualitative results in Figure 10 further validate these findings, illustrating that the inclusion of the MCA mechanism and contrastive learning significantly enhances the visual quality of restored images. These results demonstrate the effectiveness of our proposed components in improving the distilled model's performance.

Table 5. Quantitative ablation study results on Gopro [54]. The best results are highlighted in bold.

Sets	Teacher	Channel-Wise	Spatial-Wise	L_{CL}	PSNR/SSIM
(a)					30.25/0.907
(b)	✓				32.20/0.924
(c)	✓	✓			32.61/0.929
(d)	✓		✓		32.71/0.930
(e)	✓	✓	✓		32.99/0.933
Res-SKD	✓	✓	✓	✓	33.24/0.937



Figure 10. Qualitative ablation study results on Gopro [54]. (a–e) correspond to the experimental configurations (a–e) in Table 5.

To validate the choice of the Gaussian kernel for feature-level distillation, we conducted a controlled experiment comparing different similarity measures while keeping all other components of our SKD framework identical. As shown in Table 6, the Gaussian kernel achieves the best performance, outperforming both Euclidean distance and cosine similarity by 0.32 dB and 0.26 dB in PSNR, respectively.

Table 6. Ablation study on feature similarity measures for knowledge distillation. The experiment is conducted on the GoPro dataset for image deblurring, using the same student architecture and full SKD framework except for the feature-level distance metric. The best results are highlighted in bold.

Similarity Measure	Formula	PSNR ↑	SSIM ↑	Stability
Euclidean Distance	$L_{eucl} = \ x - y\ _2$	32.92	0.935	Medium
Cosine Similarity	$L_{cos} = 1 - \frac{x \cdot y}{\ x\ _2 \ y\ _2}$	32.98	0.935	Medium
Gaussian Kernel (Ours)	$GK(x, y) = 1 - \exp(-\frac{\ x - y\ _2^2}{2\sigma^2})$	33.24	0.937	High

The superior performance can be attributed to the Gaussian kernel's unique properties: (1) Unlike Euclidean distance, which is sensitive to absolute feature magnitudes and can cause gradient instability, the Gaussian kernel operates in a normalized similarity space that is more robust to feature-scale variations. (2) Compared to cosine similarity, which only considers angular alignment and ignores feature magnitude information, the Gaussian kernel incorporates both directional and magnitude relationships through the Euclidean distance in its exponent. (3) The exponential decay characteristic of the Gaussian kernel provides a soft-thresholding effect, focusing the distillation on semantically meaningful feature relationships while being tolerant to minor variations.

Empirically, we also observed that training with the Gaussian kernel exhibited smoother convergence and lower loss variance, confirming its stabilization effect on the distillation process.

4.5. Model Complexity and Hyper-Parameter Analysis

We evaluate the computational complexity of our model by comparing the FLOPs (floating-point operations) and inference time of our distilled student model with those of the teacher model and other SOTA image restoration models. The results are shown in Table 7.

Table 7. Comparison of model complexity.

Method	FLOPs	Infer Time
MPRNet [34]	565.0 G	0.0593 s
Restormer [2]	619.5 G	0.1103 s
Uformer [3]	347.6 G	0.1737 s
DRSformer [4]	972.0 G	0.3074 s
WGWS [69]	996.2 G	0.1919 s
Res-SKD	85.0 G	0.0356 s
Ufor-SKD	57.1 G	0.0540 s
DRS-SKD	132.0 G	0.0599 s

Our experiments demonstrate that the student model after distillation using the SKD strategy exhibits a substantial reduction in computational cost compared to the teacher models, with 85.4% reduction in FLOPs and 85.8% reduction in parameters. Moreover, our SKD-based model outperforms other SOTA models, including the relatively lightweight MPRNet [34], in terms of both computational efficiency and restoration quality. This confirms the efficacy of our SKD strategy in achieving high-performance image restoration with significantly reduced complexity, making it suitable for deployment on resource-constrained platforms such as autonomous vehicles.

Additionally, we conduct an analysis of key hyper-parameters to further optimize the SKD framework. The selection process for all hyper-parameters is directly illustrated by comparing the different experimental settings (a) to (e) within the table. Specifically, the results of these experiments are shown in Table 8, where we evaluate the impact of the trade-off weights (α_1 , α_2 , and α_3) used in the loss functions. We compare different settings for these hyper-parameters:

- The value of α_1 is determined by comparing setting (a) ($\alpha_1 = 0.2$) with setting (e) ($\alpha_1 = 0.5$), while keeping α_2 and α_3 constant at 0.2. The superior performance of (e) demonstrates that a higher weight of $\alpha_1 = 0.5$ for the L_{GK} loss strikes the best balance, providing more effective guidance from the teacher model.

- The values of α_2 and α_3 are analyzed by comparing settings (b), (c), (d), and (e). The optimal values are found to be $\alpha_2 = 0.2$ and $\alpha_3 = 0.2$, ensuring that the model prioritizes both the MCA mechanism and contrastive learning loss effectively.

These experiments highlight the importance of hyper-parameter tuning in optimizing the performance of the SKD strategy, ensuring that it achieves the best trade-off between image restoration quality, model complexity, and computational efficiency.

Table 8. Comparison of hyper-parameters on SOTS [57]. The best results are highlighted in bold.

Sets	α_1	α_2	α_3	PSNR/SSIM
(a)	0.2	0.2	0.2	34.95/0.981
(b)	0.5	0.1	0.1	34.74/0.974
(c)	0.5	0.1	0.2	34.89/0.978
(d)	0.5	0.2	0.1	34.91/0.979
(e)	0.5	0.2	0.2	35.06/0.985

5. Conclusions

We introduced a Soft Knowledge Distillation (SKD) strategy with Multi-Dimensional Cross-Net Attention (MCA) for compressing transformer-based image restoration models, tailored for autonomous vehicles and traffic systems. Our method reduces computational complexity while maintaining high restoration quality, making it ideal for real-time deployment in resource-constrained environments. The MCA mechanism improves feature learning by capturing critical details across both spatial and channel dimensions, and contrastive learning loss further enhances image quality.

Experiments show that our SKD-based models significantly improve deraining, deblurring, and dehazing performance, confirming their suitability for autonomous vehicle applications. Moreover, the models enhance object detection performance, boosting recall and precision on degraded images. This demonstrates the effectiveness of our approach in improving environmental perception in autonomous driving and traffic systems. Future work will focus on optimizing the method for even more resource-limited devices.

While this work demonstrates the effectiveness of our SKD framework with state-of-the-art transformer architectures, a natural extension is to validate its generalizability across other network families, such as modern CNN-based models. We anticipate that the architecture-agnostic design of our method would yield similarly compelling results, and we identify this as an important direction for future research.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Zhang, Y.; Yan, D. Soft Knowledge Distillation with Multi-Dimensional Cross-Net Attention for Image Restoration Models Compression. In Proceedings of the ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Hyderabad, India, 6–11 April 2025; pp. 1–5.
2. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.H. Restormer: Efficient transformer for high-resolution image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5728–5739.
3. Wang, Z.; Cun, X.; Bao, J.; Zhou, W.; Liu, J.; Li, H. Uformer: A general u-shaped transformer for image restoration. In Proceedings of the IEEE/CVF Conference on CVPR, New Orleans, LA, USA, 18–24 June 2022; pp. 17683–17693.

4. Chen, X.; Li, H.; Li, M.; Pan, J. Learning a sparse transformer network for effective image deraining. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 5896–5905.
5. Hinton, G.; Vinyals, O.; Dean, J. Distilling the knowledge in a neural network. *arXiv* **2015**, arXiv:1503.02531. [[CrossRef](#)]
6. Romero, A.; Ballas, N.; Kahou, S.E.; Chassang, A.; Gatta, C.; Bengio, Y. Fitnets: Hints for thin deep nets. *arXiv* **2014**, arXiv:1412.6550.
7. Chen, D.; Mei, J.P.; Zhang, Y.; Wang, C.; Wang, Z.; Feng, Y.; Chen, C. Cross-layer distillation with semantic calibration. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 2–9 February 2021; Volume 35, pp. 7028–7036.
8. Zagoruyko, S.; Komodakis, N. Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer. *arXiv* **2016**, arXiv:1612.03928.
9. Passban, P.; Wu, Y.; Rezagholizadeh, M.; Liu, Q. Alp-kd: Attention-based layer projection for knowledge distillation. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 2–9 February 2021; Volume 35, pp. 13657–13665.
10. Chen, H.; Wang, Y.; Xu, C.; Xu, C.; Tao, D. Learning student networks via feature embedding. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 25–35. [[CrossRef](#)] [[PubMed](#)]
11. Yu, L.; Yazici, V.O.; Liu, X.; van de Weijer, J.; Cheng, Y.; Ramisa, A. Learning metrics from teachers: Compact networks for image embedding. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2907–2916.
12. Gao, Q.; Zhao, Y.; Li, G.; Tong, T. Image super-resolution using knowledge distillation. In *Computer Vision—ACCV 2018, Proceedings of the Asian Conference on Computer Vision, Perth, Australia, 2–6 December 2018*; Springer: Cham, Switzerland, 2018; pp. 527–541.
13. Fang, H.; Long, Y.; Hu, X.; Ou, Y.; Huang, Y.; Hu, H. Dual cross knowledge distillation for image super-resolution. *J. Vis. Commun. Image Represent.* **2023**, *95*, 103858. [[CrossRef](#)]
14. Jin, Q.; Ren, J.; Woodford, O.J.; Wang, J.; Yuan, G.; Wang, Y.; Tulyakov, S. Teachers do more than teach: Compressing image-to-image models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13600–13611.
15. Chen, H.; Wang, Y.; Shu, H.; Wen, C.; Xu, C.; Shi, B.; Xu, C.; Xu, C. Distilling portable generative adversarial networks for image translation. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 3585–3592.
16. Zhang, L.; Chen, X.; Tu, X.; Wan, P.; Xu, N.; Ma, K. Wavelet knowledge distillation: Towards efficient image-to-image translation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 12464–12474.
17. He, K.; Sun, J.; Tang, X. Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *33*, 2341–2353. [[CrossRef](#)] [[PubMed](#)]
18. Fattal, R. Dehazing using color-lines. *ACM Trans. Graph. (TOG)* **2014**, *34*, 1–14. [[CrossRef](#)]
19. Chan, T.F.; Wong, C.K. Total variation blind deconvolution. *IEEE Trans. Image Process.* **1998**, *7*, 370–375. [[CrossRef](#)]
20. Berman, D.; Treibitz, T.; Avidan, S. Non-local image dehazing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 1674–1682.
21. Pan, J.; Sun, D.; Pfister, H.; Yang, M.H. Blind image deblurring using dark channel prior. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 1628–1636.
22. Hu, Z.; Cho, S.; Wang, J.; Yang, M.H. Deblurring low-light images with light streaks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 3382–3389.
23. Li, Y.; Tan, R.T.; Guo, X.; Lu, J.; Brown, M.S. Rain streak removal using layer priors. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2736–2744.
24. Meng, G.; Wang, Y.; Duan, J.; Xiang, S.; Pan, C. Efficient image dehazing with boundary constraint and contextual regularization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Portland, OR, USA, 23–28 June 2013; pp. 617–624.
25. Li, Y.; Tan, R.T.; Brown, M.S. Nighttime haze removal with glow and multiple light colors. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 226–234.
26. Chen, B.H.; Huang, S.C. An advanced visibility restoration algorithm for single hazy images. *ACM Trans. Multimed. Comput. Commun. Appl. (TOMM)* **2015**, *11*, 1–21. [[CrossRef](#)]
27. Chen, Z.; Wang, Y.; Yang, Y.; Liu, D. PSD: Principled synthetic-to-real dehazing guided by physical priors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 7180–7189.
28. Dong, H.; Pan, J.; Xiang, L.; Hu, Z.; Zhang, X.; Wang, F.; Yang, M.H. Multi-scale boosted dehazing network with dense feature fusion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 2157–2167.

29. Wang, T.; Yang, X.; Xu, K.; Chen, S.; Zhang, Q.; Lau, R.W. Spatial attentive single-image deraining with a high quality real rain dataset. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 12270–12279.
30. Shao, Y.; Li, L.; Ren, W.; Gao, C.; Sang, N. Domain adaptation for image dehazing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 2808–2817.
31. Liu, Y.F.; Jaw, D.W.; Huang, S.C.; Hwang, J.N. DesnowNet: Context-aware deep network for snow removal. *IEEE Trans. Image Process.* **2018**, *27*, 3064–3073. [[CrossRef](#)]
32. Chen, W.T.; Fang, H.Y.; Hsieh, C.L.; Tsai, C.C.; Chen, I.; Ding, J.J.; Kuo, S.Y. 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, Montreal, QC, Canada, 10–17 October 2021; pp. 4196–4205.
33. Cheng, B.; Li, J.; Chen, Y.; Zeng, T. Snow mask guided adaptive residual network for image snow removal. *Comput. Vis. Image Underst.* **2023**, *236*, 103819. [[CrossRef](#)]
34. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.H.; Shao, L. Multi-stage progressive image restoration. In Proceedings of the IEEE/CVF Conference on CVPR, Nashville, TN, USA, 20–25 June 2021; pp. 14821–14831.
35. Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L.; Timofte, R. Swinir: Image restoration using swin transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 1833–1844.
36. Lee, H.; Choi, H.; Sohn, K.; Min, D. Knn local attention for image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 2139–2149.
37. Song, Y.; He, Z.; Qian, H.; Du, X. Vision transformers for single image dehazing. *IEEE Trans. Image Process.* **2023**, *32*, 1927–1941. [[CrossRef](#)]
38. Wang, T.; Yuan, L.; Zhang, X.; Feng, J. Distilling object detectors with fine-grained feature imitation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 4933–4942.
39. Ge, S.; Zhao, S.; Li, C.; Li, J. Low-resolution face recognition in the wild via selective knowledge distillation. *IEEE Trans. Image Process.* **2018**, *28*, 2051–2062. [[CrossRef](#)]
40. Liu, Y.; Chen, K.; Liu, C.; Qin, Z.; Luo, Z.; Wang, J. Structured knowledge distillation for semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2604–2613.
41. Hong, M.; Xie, Y.; Li, C.; Qu, Y. Distilling image dehazing with heterogeneous task imitation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 3462–3471.
42. Chen, W.T.; Huang, Z.K.; Tsai, C.C.; Yang, H.H.; Ding, J.J.; Kuo, S.Y. Learning multiple adverse weather removal via two-stage knowledge learning and multi-contrastive regularization: Toward a unified model. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 17653–17662.
43. Wu, H.; Qu, Y.; Lin, S.; Zhou, J.; Qiao, R.; Zhang, Z.; Xie, Y.; Ma, L. Contrastive learning for compact single image dehazing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 10551–10560.
44. Chen, Y.; Pan, Y.; Wang, Y.; Yao, T.; Tian, X.; Mei, T. Transferrable contrastive learning for visual domain adaptation. In Proceedings of the ACM Multimedia Conference, Virtual, 20–24 October 2021; pp. 3399–3408.
45. Shen, K.; Jones, R.M.; Kumar, A.; Xie, S.M.; HaoChen, J.Z.; Ma, T.; Liang, P. Connect, not collapse: Explaining contrastive learning for unsupervised domain adaptation. In Proceedings of the 39th International Conference on Machine Learning, PMLR, Baltimore, MD, USA, 17–23 July 2022; pp. 19847–19878.
46. Ye, Y.; Yu, C.; Chang, Y.; Zhu, L.; Zhao, X.L.; Yan, L.; Tian, Y. Unsupervised Deraining: Where Contrastive Learning Meets Self-Similarity. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 5821–5830.
47. Chen, X.; Pan, J.; Jiang, K.; Li, Y.; Huang, Y.; Kong, C.; Dai, L.; Fan, Z. Unpaired Deep Image Deraining Using Dual Contrastive Learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 2017–2026.
48. Chen, X.; Fan, Z.C.; Zheng, Z.; Li, Y.; Huang, Y.; Dai, L.; Kong, C.; Li, P. Unpaired Deep Image Dehazing Using Contrastive Disentanglement Learning. In *Computer Vision—ECCV 2022, Proceedings of the 17th European Conference, Tel Aviv, Israel, 23–27 October 2022*; Springer: Cham, Switzerland, 2022; pp. 632–648.
49. Zheng, Y.; Zhan, J.; He, S.; Dong, J.; Du, Y. Curricular contrastive regularization for physics-aware single image dehazing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 5785–5794.
50. Sohn, K. Improved deep metric learning with multi-class n-pair loss objective. *NeurIPS* **2016**, *29*, 1849–1857.
51. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.

52. Fu, X.; Huang, J.; Zeng, D.; Huang, Y.; Ding, X.; Paisley, J. Removing rain from single images via a deep detail network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 3855–3863.
53. Zhang, H.; Patel, V.M. Density-aware single image de-raining using a multi-stream dense network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 695–704.
54. Nah, S.; Hyun Kim, T.; Mu Lee, K. Deep multi-scale convolutional neural network for dynamic scene deblurring. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3883–3891.
55. Shen, Z.; Wang, W.; Lu, X.; Shen, J.; Ling, H.; Xu, T.; Shao, L. Human-aware motion deblurring. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 5572–5581.
56. Rim, J.; Lee, H.; Won, J.; Cho, S. Real-world blur dataset for learning and benchmarking deblurring algorithms. In *Computer Vision—ECCV 2020, Proceedings of the 16th European Conference, Glasgow, UK, 23–28 August 2020*; Springer: Cham, Switzerland, 2020; pp. 184–201.
57. Li, B.; Ren, W.; Fu, D.; Tao, D.; Feng, D.; Zeng, W.; Wang, Z. Benchmarking single-image dehazing and beyond. *IEEE Trans. Image Process.* **2018**, *28*, 492–505. [[CrossRef](#)] [[PubMed](#)]
58. Huynh-Thu, Q.; Ghanbari, M. Scope of validity of PSNR in image/video quality assessment. *Electron. Lett.* **2008**, *44*, 800–801. [[CrossRef](#)]
59. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)] [[PubMed](#)]
60. Mittal, A.; Moorthy, A.K.; Bovik, A.C. No-reference image quality assessment in the spatial domain. *IEEE Trans. Image Process.* **2012**, *21*, 4695–4708. [[CrossRef](#)] [[PubMed](#)]
61. Venkatanath, N.; Praneeth, D.; Bh, M.C.; Channappayya, S.S.; Medasani, S.S. Blind image quality evaluation using perception based features. In Proceedings of the 2015 Twenty First National Conference on Communications (NCC), Mumbai, India, 27 February–1 March 2015; pp. 1–6.
62. Loshchilov, I.; Hutter, F. Sgdr: Stochastic gradient descent with warm restarts. *arXiv* **2016**, arXiv:1608.03983.
63. Ren, D.; Zuo, W.; Hu, Q.; Zhu, P.; Meng, D. Progressive image deraining networks: A better and simpler baseline. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3937–3946.
64. Wang, H.; Xie, Q.; Zhao, Q.; Meng, D. A model-driven deep neural network for single image rain removal. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 3103–3112.
65. Zhang, H.; Dai, Y.; Li, H.; Koniusz, P. Deep stacked hierarchical multi-patch network for image deblurring. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 5978–5986.
66. Park, D.; Kang, D.U.; Kim, J.; Chun, S.Y. Multi-temporal recurrent neural networks for progressive non-uniform single image deblurring with incremental temporal training. In *Computer Vision—ECCV 2020, Proceedings of the 16th European Conference, Glasgow, UK, 23–28 August 2020*; Springer: Cham, Switzerland, 2020; pp. 327–343.
67. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934. [[CrossRef](#)]
68. Zhang, Y.; Yan, D. RMTD: A Robust Multi-Type Degradation Dataset for Image Enhancement [EB/OL]. Available online: <https://github.com/ICME25/RMTD> (accessed on 20 December 2024).
69. Zhu, Y.; Wang, T.; Fu, X.; Yang, X.; Guo, X.; Dai, J.; Qiao, Y.; Hu, X. Learning Weather-General and Weather-Specific Features for Image Restoration Under Multiple Adverse Weather Conditions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 21747–21758.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.