



Article

g2D-Net: Efficient Dehazing with Second-Order Gated Units

Jia Jia ¹, Zhibo Wang ^{1,*} and Jeongik Min ^{2,*}¹ Graduate School of Artificial Intelligence, Artificial Intelligence Research Center, Jeonju University, Jeonju-si 55069, Republic of Korea² Department of Artificial Intelligence, Artificial Intelligence Research Center, Jeonju University, Jeonju-si 55069, Republic of Korea

* Correspondence: wang970128@jj.ac.kr (Z.W.); minji@jj.ac.kr (J.M.)

Abstract: Image dehazing aims to reconstruct potentially clear images from corresponding images corrupted by haze. With the rapid development of deep learning-related technologies, dehazing methods based on deep convolutional neural networks have gradually become mainstream. We note that existing dehazing methods often accompany an increase in computational overhead while improving the performance of dehazing. We propose a novel lightweight dehazing neural network to balance performance and efficiency: the g2D-Net. The g2D-Net borrows the design ideas of input-adaptive and long-range information interaction from Vision Transformers and introduces two kinds of convolutional blocks, i.e., the g2D Block and the FFT-g2D Block. Specifically, the g2D Block is a residual block with second-order gated units, which inherit the input-adaptive property of a gated unit and can realize the second-order interaction of spatial information. The FFT-g2D Block is a variant of the g2D Block, which efficiently extracts the global features of the feature maps through fast Fourier convolution and fuses them with local features. In addition, we employ the SK Fusion layer to improve the cascade fusion layer in a traditional U-Net, thus introducing the channel attention mechanism and dynamically fusing information from different paths. We conducted comparative experiments on five benchmark datasets, and the results demonstrate that the g2D-Net achieves impressive dehazing performance with relatively low complexity.

Keywords: image dehazing; CNN; U-Net

Citation: Jia, J.; Wang, Z.; Min, J. g2D-Net: Efficient Dehazing with Second-Order Gated Units. *Electronics* **2024**, *13*, 1900. <https://doi.org/10.3390/electronics13101900>

Academic Editor: Manohar Das

Received: 12 April 2024

Revised: 22 April 2024

Accepted: 9 May 2024

Published: 12 May 2024



Copyright: © 2024 by the authors. 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 dehazing is a crucial task in the field of computer vision. It aims to restore potentially clear images from haze-affected images, enhancing image quality and visual clarity. Given its critical role in various fields, such as remote sensing, autonomous driving, and security monitoring, image dehazing has sparked widespread interest in academia and industry.

With the rapid development of deep learning techniques over the past decade, deep convolutional neural networks (CNNs) [1,2] have been vital in driving significant progress in computer vision. Although the emergence of Vision Transformers (ViTs) [3] has posed a significant challenge to the dominance of CNNs, CNNs currently remain the preferred approach for image dehazing tasks. This preference stems from two main reasons:

- In image dehazing research, acquiring paired hazy and clear images is an arduous endeavor, resulting in the prevalent dehazing datasets being relatively small in scale.
- In applications such as autonomous driving and security surveillance, where image dehazing computations often occur on edge devices with limited computational resources, both computational efficiency and dehazing performance hold equal importance. CNNs generally offer higher computational efficiency in such scenarios.

However, this does not mean that a Vision Transformer cannot promote the development of the image dehazing field; many current studies have shown that ViTs and

CNNs can promote each other's development [4,5]. In order to achieve better results, this study borrows some of the key design concepts that have made ViTs successful in the field of computer vision and applies them to CNNs to improve the performance of dehazing while maintaining the efficient computation of the CNNs. The dehazing method proposed in this study is called the second-order Gate Dehaze U-Net (g2D-Net). To validate the effectiveness of the g2D-Net, we conducted experiments on mainstream dehazing datasets. As shown in Figure 1, the g2D-Net can achieve impressive performance using a small computational overhead.

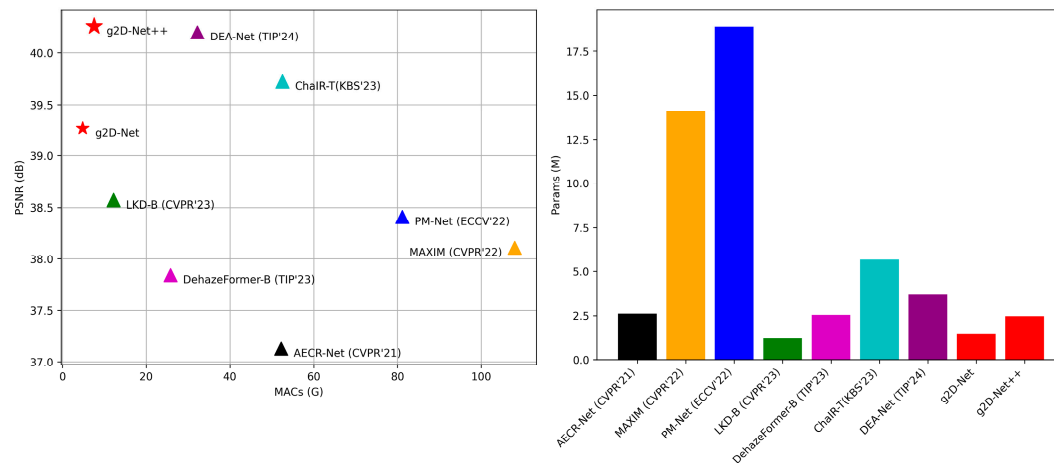


Figure 1. Comparisons between our methods and other state-of-the-art methods. (Left): PSNR vs. MACs on a SOTS-Indoor dataset. (Right): The number of parameters of the models.

Specifically, the study's aims were as follows:

- Inspired by the input-adaptive property of ViTs, we construct a residual block with the second-order spatial interaction mechanism based on gated convolution [6]: the g2D Block. We use it as a backbone to construct a lightweight dehazing U-Net [7] with seven stages.
- To provide the network with some long-range capability and keep the computation as efficient as possible, we propose the FFT-g2D Block. This residual block is a derivative of the g2D Block, which uses fast Fourier convolution [8] to extract global information about the feature map in the frequency domain.
- The g2D-Net replaces the cascading fusion layer in the traditional U-Net with an SK fusion layer modified from the SK-Attention mechanism [9], thus dynamically fusing information from different paths.
- In the g2D-Net's training, we input/output hazy/clear images with different sizes in multiple stages. This multi-input/output strategy can effectively reduce the training difficulty of the network and significantly improve the convergence speed of the network [10].

2. Related Work

With the continuous development of image processing technology, mainstream image dehazing methods are also changing. According to their different working principles, we categorize the dehazing methods into traditional and deep learning-based methods.

2.1. Traditional Methods

The traditional methods can be roughly divided into three types: image enhancement methods [11–14], image restoration methods [15–17], and fusion-based methods [18–21]. Traditional image dehazing methods typically rely on handcrafted priors based on statistical rules or Atmospheric Scattering Models (ASMs) [22–24]. They often possess high interpretability and low computational costs. However, these methods exhibit poor gener-

alization and robustness, usually performing well only in specific scenarios, and cannot handle hazy images in complex scenes. For instance, the well-known Dark Channel Prior (DCP) method [15] could perform better when dealing with hazy images without large sky areas. Although these traditional methods are now rarely used, they played a significant role in early dehazing research, laying the foundation for subsequent studies.

2.2. Deep Learning-Based Methods

With the popularity and rapid development of deep learning, deep learning-based methods have become the mainstream dehazing method. Deep learning-based methods are trained using pairs of images and learning the mapping between the hazy and clear images. Deep learning-based methods usually have better performance and robustness than traditional methods. Based on whether an ASM is introduced or not, deep learning-based methods can be categorized into two types.

Some deep learning-based methods rely on ASMs [25–27]. These methods usually use neural networks to estimate the medium transmission map or global atmospheric light and calculate potentially clear images using an ASM. As a priori knowledge, introducing an ASM can significantly reduce the complexity of neural networks. However, an ASM is not enough to explain the imaging process of haze images perfectly under complex conditions. Thus, it also limits the further improvement of the neural network's dehazing performance to some extent.

Nowadays, state-of-the-art (SOTA) deep learning-based methods no longer rely on ASMs. These end-to-end dehazing models utilize a data-driven approach to generate clear images directly from hazy inputs. Among the various methods, networks based on encoder–decoder architecture have shown promising performance [28–32]. An encoder–decoder network typically features a clear and concise framework structure, with its performance and model size mainly influenced by the residual blocks. As a lightweight network, the g2D-Net also adopts this architecture. Additionally, methods based on GANs [33,34] and knowledge transfer [35] have also achieved good results but come with higher complexity. Refs. [36,37] applied a Vision Transformer for the first time to an image dehazing task and achieved good results in several dehazing datasets. Compared with CNNs, ViTs have higher computational complexity and higher requirements for training data quality. These problems limit the further development of Transformer-based methods for dehazing.

These excellent works have achieved remarkable success in image dehazing tasks. However, they also face an acute problem: to achieve better dehazing results, the current research work continuously increases the depth and width of the network and introduces complex network structures, which increases the network's difficulty in overcoming training, inference, and deployment difficulties. This study aims to propose a lightweight convolutional neural network for dehazing that achieves a balance between complexity and performance.

2.3. The PSNR and the SSIM

Researchers commonly use the Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index (SSIM) [38] to evaluate the objective performance of a method in image dehazing. For image pair (x, y) :

$$PSNR(x, y) = 20 \log_{10} \left(\frac{MAX}{MSE(x, y)} \right) \quad (1)$$

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \leq 1 \quad (2)$$

MAX represents the maximum possible pixel value of the image, and MSE stands for the mean square error function. In Equation (2), μ_x and μ_y denote the means of x and y , respectively, while σ_x and σ_y represent the standard deviations. σ_{xy} signifies the covariance between x and y , and c_1 and c_2 are small constants introduced to prevent division by zero.

The PSNR is a metric for assessing the degree of degradation in the quality of reconstructed images, primarily reflecting the pixel-level similarity between image pairs. Higher PSNR values indicate better image quality. On the other hand, the SSIM considers differences in brightness, contrast, and structure between image pairs. Compared to the PSNR, the SSIM aligns better with the perceptual judgment of image quality made by the human visual system. A higher SSIM value indicates a higher similarity between the two images being compared.

3. Methods

The g2D-Net is a U-Net structure containing seven stages with local and global residuals, and its overall architecture and the design of some of its modules are shown in Figure 2. The g2D-Net contains two convolutional residual blocks: the g2D Block and the FFT-g2D Block. These two types of residual blocks allow the network to have capabilities such as being input-adaptive and having a long range. In addition, the g2D-Net uses the SK Fusion layer to improve performance further and uses multiple input/output strategies to reduce the difficulty of model training.

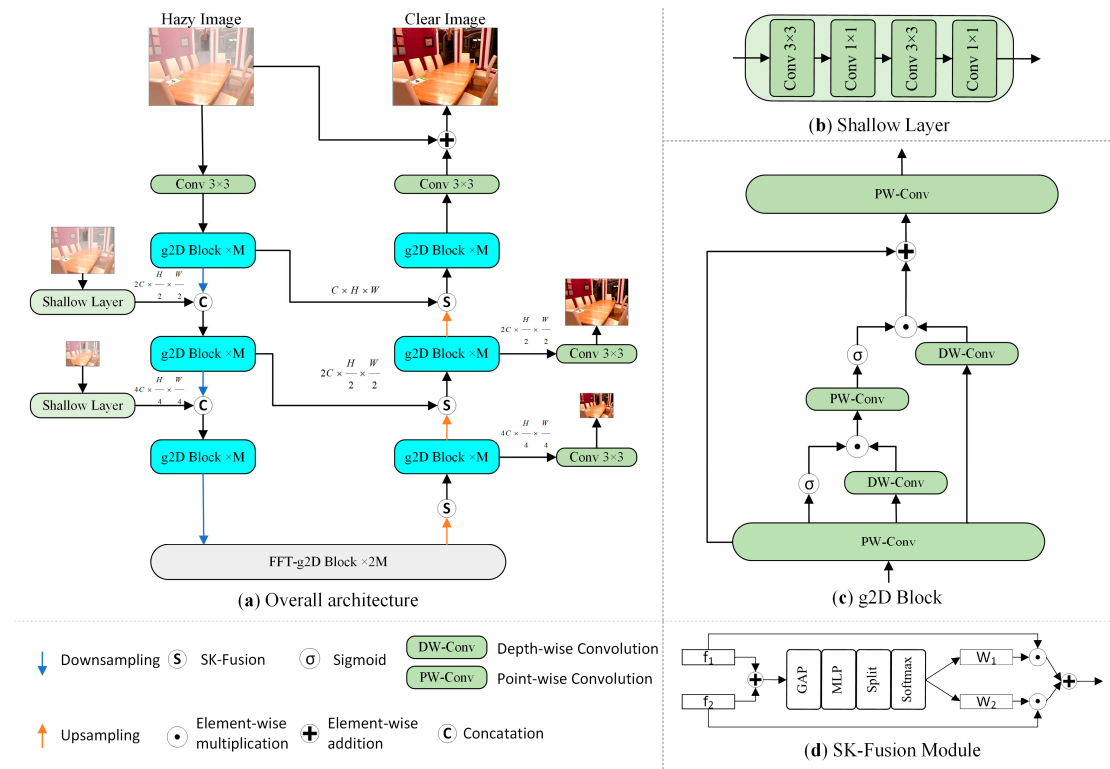


Figure 2. (a) The overall architecture of the second-order Gate Dehaze U-Net (g2D-Net). (b) The architecture of the shallow layer. (c) The architecture of the g2D Block. (d) The architecture of the SK Fusion layer. The architecture of the FFT-g2D Block is shown in Figure 3.

3.1. The g2D Block

Existing research indicates that gating units and their variants can effectively enhance the performance of models in various computer vision tasks [28,39,40]. However, the existing gating units cannot facilitate information being transferred across long-range and high-order interactions. On the other hand, input-adaptive, long-range, and high-order interactions may be critical factors in the success of ViTs. Inspired by this, we propose a convolutional residual block with input-adaptive and second-order interaction capabilities: the g2D Block. The g2D Block operates fundamentally based on gated convolution units [41]. Distinct from the typical gated convolution block, the g2D Block comprises two gated convolution units, thereby achieving a second-order spatial interaction. Gated convolution

learns, for each channel and each spatial location, a dynamic feature selection mechanism acting on the feature map, a property similar to the input adaptivity of self-attention, and is inherited by the g2D Block.

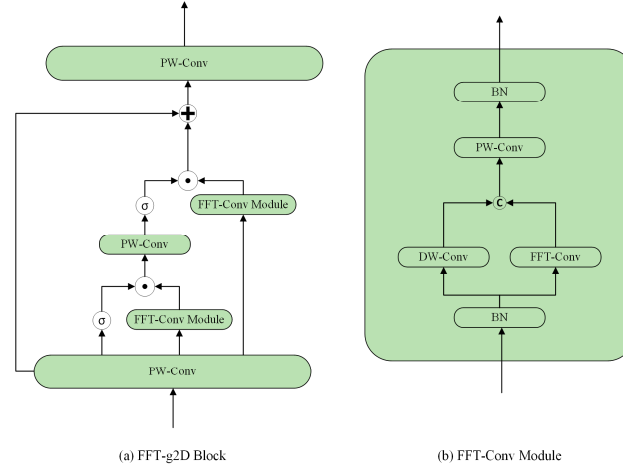


Figure 3. (a) The architecture of the FFT-g2D Block. (b) The architecture of the FFT-Conv module in the FFT-g2D Block. The meanings of the symbols in the figure are consistent with those listed in Figure 2.

Let $X \in R^{C \times H \times W}$ represent the input feature map, and ϕ_i represents the i -th point-wise convolution. The g2D Block initially employs ϕ_0 to project X to $\phi_0(X) \in R^{2C \times H \times W}$, subsequently dividing it into three feature vectors for processing based on channels:

$$\left[P_0^{\frac{C}{2} \times H \times W}, D_0^{\frac{C}{2} \times H \times W}, D_1^{C \times H \times W} \right] = \text{Split}(\phi_0(X)) \quad (3)$$

The g2D Block consists of two gated convolution operations. In these two operations, P_1 and P_2 represent their outputs, respectively. First, we compute P_1 using Equation (3), followed by mapping P_1 through ϕ_1 and participating in the computation of the second gated convolution along with D_1 to obtain P_2 :

$$P_1^{C/2 \times H \times W} = \sigma(P_0^{C/2 \times H \times W}) \odot DW(D_0^{C/2 \times H \times W}) \quad (4)$$

$$P_2^{C \times H \times W} = \sigma(\phi_1(P_1)^{C \times H \times W}) \odot DW(D_1^{C \times H \times W}) \quad (5)$$

Here, σ represents the Sigmoid function, which maps the output of the gated convolution operation to the interval (0, 1), thus helping to mitigate the risk of gradient explosion during the model's training process. DW represents depth-wise convolution. Although the Sigmoid function can be replaced with bounded functions like hard sigmoid, tanh, and hard tanh, we still strongly recommend using the sigmoid function for best performance.

3.2. The FFT-g2D Block

To give the neural network a certain degree of long-range capability, we propose the FFT-g2D Block based on the g2D Block. The FFT-g2D Block is a derivative version of the g2D Block used in some stages. Figure 3 illustrates its structure. The FFT-g2D Block employs both 3×3 depth-wise convolution (DW-Conv) [42] and FFT convolution (FFT-Conv) operators, allowing it to capture both local and global information from the feature maps simultaneously.

The FFT-Conv operator, utilizing channel-wise fast Fourier transform, effectively extracts global information from the feature maps in the frequency domain. Let $X \in R^{C \times H \times W}$

represent the input feature map. It is first subjected to discrete Fourier transformation, transitioning the feature map from the spatial domain to the frequency domain:

$$X_F = \mathcal{F}(X) \in \mathbb{C}^{C \times H \times W} \quad (6)$$

Performing convolution in the spatial domain is equivalent to point-wise multiplication in the frequency domain. Therefore, by applying a learnable frequency domain filter $K \in \mathbb{C}^{C \times H \times W}$ to X_F through point-wise multiplication, frequency domain information can be filtered. This filter in the frequency domain, referred to as the global filter, has the exact dimensions as X_F . Finally, the filtered features in the frequency domain are transformed back to the spatial domain using the inverse discrete Fourier transform:

$$X \leftarrow \mathcal{F}^{-1}(K \odot X_F) \in \mathbb{R}^{C \times H \times W} \quad (7)$$

FFT-Conv is equivalent to depth-wise global circular convolution, but it has a time complexity of only $O(CHW \log(HW))$. Due to the significantly higher complexity of the FFT-g2D Block compared to that of the g2D Block and its proportionality to the input feature map dimensions, we choose to incorporate FFT-g2D Blocks in only select stages.

3.3. SK Fusion

The SK Fusion layer is a simple improvement of the SK module, introducing channel attention mechanisms into the model while dynamically integrating feature maps from both the encoder and decoder stages. Several studies have demonstrated that the SK Fusion layer can significantly enhance model performance, mainly when dealing with low-level computer vision tasks [29,43]. Figure 2d illustrates the SK Fusion layer's structure. It takes feature maps f_1 from the encoder stage and f_2 from the main path. Initially, they are fused through element-wise addition. Subsequently, a sequence of operations, including Global Average Pooling (GAP), Multilayer Perceptron (MLP), and SoftMax, are performed to extract channel attention (the MLP in the SK Fusion layer consists of two fully connected layers. The first fully connected layer reduces the number of channels, c to $c/8$, while the second fully connected layer increases the number of channels to $2c$. The ReLU activation function is applied between the two fully connected layers). The attention vector is then separated in the channel dimension to obtain the weights w_1 and w_2 for f_1 and f_2 , respectively. Finally, f_1 and f_2 are weighted according to these weights and added to obtain the output of the SK Fusion layer. The process is as follows:

$$\{w_1, w_2\} = \text{Split}(\text{SoftMax}(\text{MLP}(\text{GAP}(f_1 + f_2)))) \quad (8)$$

$$\text{out} = w_1 f_1 + w_2 f_2 \quad (9)$$

3.4. The Loss Function

The loss function of the g2D-Net comprises two parts: the spatial loss function and the frequency loss function. Both of these components utilize L_1 loss functions. The ultimate loss is computed as the weighted sum of the spatial loss and the frequency loss:

$$L = \sum_{i=1}^3 (\|\hat{y}_i - y_i\|_1 + \lambda \|\mathcal{F}(\hat{y}_i) - \mathcal{F}(y_i)\|_1) \quad (10)$$

In the equation, i represents the index of different sizes of input/output images, where y and \hat{y} represent the output images and the label images, respectively. The hyperparameter λ is configured to be 0.1.

3.5. Architectural Details

The g2D-Net follows the design of [29,30], with the ratio of block quantities set as $[M : M : M : 2M : M : M : M]$ across different stages. As a lightweight model, in the g2D-Net, we set $M = 2$. We introduce a variant with increased depth called g2D-Net++ to

cater to different scenarios. g2D-Net++ is twice as deep as the g2D-Net, with $M = 4$. In the ablation study section, we also discussed the scheme of increasing the embedding dimension of the model. However, experimental results showed that increasing the number of blocks is a more effective way of expanding the g2D-Net. The architectural details can be found in Table 1.

Table 1. The detailed architecture specifications. Bold indicates that the FFT-g2D Block was used.

Model	M	Num. of Blocks	Embedding Dims
g2D-Net	2	[2,2,2, 4 ,2,2,2]	[24,48,96, 192 ,96,48,24]
g2D-Net++	4	[4,4,4, 8 ,4,4,4]	[24,48,96, 192 ,96,48,24]

4. Experiment

We conducted extensive experiments on synthetic datasets (RESIDE [44] and Haze-4K [45]) and real-world datasets (Dense-Haze [46] and NH-Haze [47]), evaluating the models' objective dehazing performance using the PSNR and the SSIM. We performed model training separately for indoor and outdoor scenes for the RESIDE dataset and evaluated them on the corresponding SOTS-Indoor and SOTS-Outdoor test sets. The Dense-Haze dataset contains densely and uniformly hazed images, and the NH-Haze dataset contains non-uniformly hazed images. Both datasets consist of 55 image pairs, with 50 pairs used for training and the remaining five used for testing. We utilized the PyTorch (Version: 1.9.0) framework and NVIDIA A100 GPU for model construction and training. The warmup strategy was employed during the initial 50 epochs to increase the learning rate to its initial value gradually. Subsequently, the learning rate was gradually reduced to 1/100 of the initial learning rate according to a cosine decay strategy [48]. We employed AdamW [49] as the optimizer for model training.

4.1. The Main Experimental Results

We first conducted experiments on synthetic datasets. For all variants of the g2D-Net, we randomly cropped training images to 256×256 , employed a Mini Batch Size of 32, and conducted training for 1000 epochs. The initial learning rates for the g2D-Net and g2D-Net++ were set to 8×10^{-4} and 4×10^{-4} , respectively. Figure 4 illustrates the training process of the g2D-Net across different datasets. To assess the effectiveness of our proposed approach, we performed quantitative performance comparisons between the g2D-Net and the SOTA methods. The results of the comparative experiments are elaborated in detail in Table 2. The experimental findings demonstrate that our method exhibits impressive performance across multiple datasets, striking a favorable balance between dehazing effectiveness and computational efficiency. Despite being a lightweight model, the g2D-Net achieves SOTA performance across various datasets. Specifically, compared to the classic FFA-Net [50], the g2D-Net utilizes only approximately 7.7% of the parameter count and about 1.7% of the MACs. However, it improves the PSNR by 2.88 dB and 2.46 dB on the RESIDE-IN and RESIDE-OUT datasets. In contrast to other more advanced methods, the g2D-Net achieves a notable reduction in parameter counts and MACs, yet it attains dehazing performance close to or surpassing their effects. For instance, compared to the advanced Transformer-based method DehazeFormer-B [37], g2D-Net++ exhibits a parameter reduction of approximately 70%. However, g2D-Net++ achieves an increase in the PSNR by 2.42 dB on the RESIDE-IN dataset. Figure 5 illustrates the dehazing effects of the g2D-Net in different scenarios. The g2D-Net handles various dehazing situations, effectively restoring details and textures affected by haze, suppressing artifacts, enhancing clarity, improving contrast, and recovering color in the images.

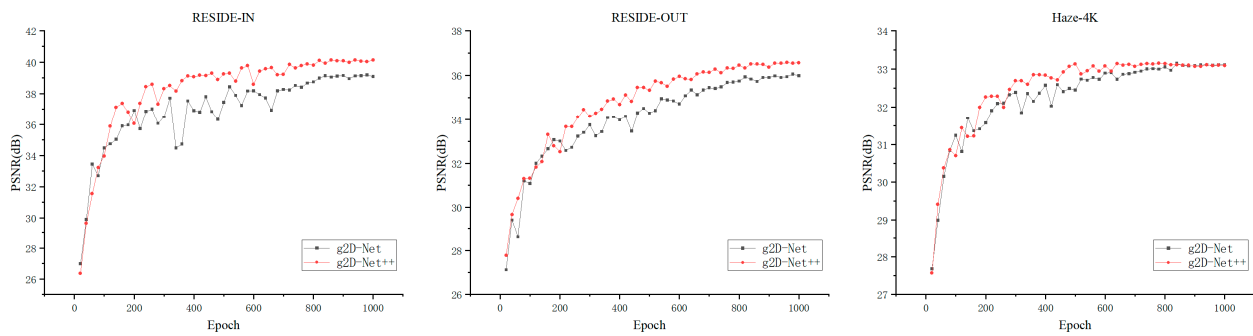


Figure 4. The training process of the g2D-Net and g2D-Net++. The vertical axis represents the PSNR of the models on the test set. The data were sampled at intervals with a sampling step of 20 for better visualization.

Table 2. The benchmarking dehazing methods on synthetic datasets. The data for the other methods in the table are taken from their respective papers. ‘-’ indicates that there are no such data in the original paper. The best performance will be displayed in bold; the second-best performance will be indicated using underlining.

Model	RESIDE-IN		RESIDE-OUT		Haze4K		Overhead	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	MACs(G)	Param(M)
(TPAMI’10) DCP [15]	16.62	0.818	19.13	0.815	14.01	0.760	-	-
(TIP’16) DehazeNet [25]	19.82	0.821	24.75	0.927	19.12	0.840	0.581	0.009
(ECCV’16) MSCNN [26]	19.84	0.833	22.06	0.908	14.01	0.510	0.525	0.008
(ICCV’17) AOD-Net [27]	20.51	0.816	24.14	0.920	17.15	0.830	0.115	0.002
(ICCV’19) GridDehazeNet [51]	32.16	0.984	30.86	0.982	-	-	21.49	0.956
(AAAI’20) FFA-Net [50]	36.39	0.989	33.57	0.984	26.96	0.950	287.8	4.456
(CVPR’20) MSBDN [52]	33.67	0.985	33.48	0.982	22.99	0.850	41.54	31.35
(ECCV’21) PMNet [53]	38.41	0.990	34.74	0.985	-	-	81.13	18.90
(CVPR’22) AECR-Net [54]	37.13	0.990	-	-	-	-	52.20	2.611
(CVPR’22) MAXIM-2S [28]	38.11	0.991	34.19	0.985	-	-	216	14.10
(CVPR’22) DeHamer [36]	36.63	0.998	35.18	0.986	-	-	48.93	132.45
(TIP’23) DehazeFormer-B [37]	37.84	0.994	34.95	0.984	-	-	25.79	2.541
(ICME’23) LKD-L [55]	39.44	0.994	34.82	0.983	-	-	23.93	2.38
(arXiv’23) MixDehazeNet [30]	39.47	0.995	35.09	0.985	-	-	22.06	3.16
(ICLR’23) SF-Net [32]	41.24	<u>0.996</u>	40.05	0.996			66.61	7.05
(KBS’23) ChaIR-T [56]	39.72	0.995	<u>38.01</u>	<u>0.995</u>			5.66	52.55
(TIP’24) DEA-Net [31]	40.20	0.993	36.03	0.989	<u>33.19</u>	0.99	32.23	3.653
(EAAI’24) HRA-Net [57]	38.83	0.985	36.15	0.988	-	-	-	7.29
g2D-Net	39.27	0.995	36.07	0.983	33.14	<u>0.986</u>	4.839	1.452
g2D-Net++	<u>40.26</u>	<u>0.996</u>	36.60	0.985	33.24	<u>0.986</u>	7.560	2.466

To better evaluate the g2D-Net, we conducted experiments on two more challenging real-world datasets. During training, input images were resized to 800×1200 , while full-size images were used during testing. Figure 6 illustrates the test results of the g2D-Net on the Dense-Haze and NH-Haze test sets. Table 3 presents the comparative experimental results between the g2D-Net and other methods. The experimental findings indicate that, compared to synthetic datasets, the model’s dehazing performance declines when faced with more challenging real-world datasets. This decline mainly manifests in suboptimal edge details and color reproduction when reconstructing clear images. This suggests that mainstream synthetic datasets still lack realism. However, comparative experimental results indicate that the objective performance metrics of the g2D-Net, particularly those of the SSIM metric, still outperform most existing methods. The excellent SSIM metric results suggest that the overall visual perception quality of images processed by the g2D-Net is higher, which may be attributed to the long-range interaction capability of the g2D-Net.

On the other hand, the lower PSNR compared to large-scale SOTA dehazing methods may imply that the performance of the g2D-Net in reconstructing pixel-level details needs improvement due to its smaller parameter size.

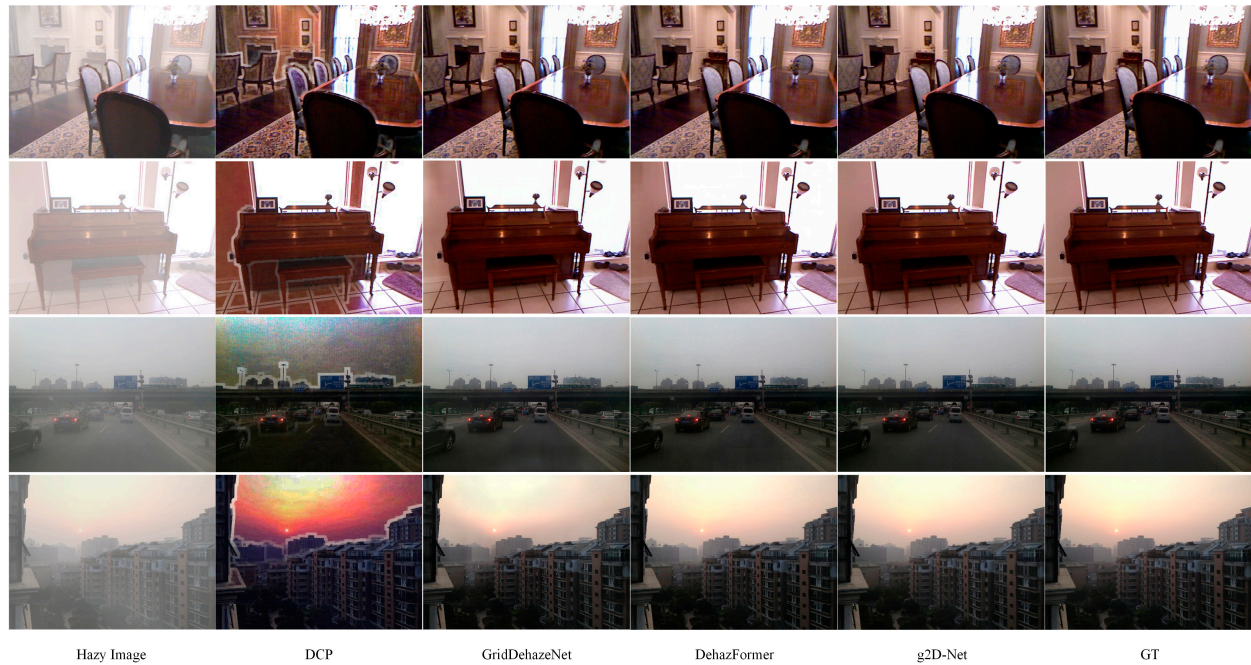


Figure 5. A comparison of the dehaze results of different methods in an indoor scene and an outdoor scene. The haze images are from the RESIDE dataset. GT denotes Ground Truth images.



Figure 6. The dehazing results on real-world datasets. The first row shows the hazy images, and the second row shows the Ground Truth images.

Table 3. The experiments on real-world datasets. The best performance will be displayed in bold; the second-best performance will be indicated using underlining.

Model	Dense-Haze		NH-Haze	
	PSNR	SSIM	PSNR	SSIM
(ICCV'19) GridDehazeNet [51]	13.31	0.368	13.80	0.537
(AAAI'20) FFA-Net [50]	14.39	0.407	19.87	0.692

Table 3. Cont.

Model	Dense-Haze		NH-Haze	
	PSNR	SSIM	PSNR	SSIM
(CVPR'21) AECR-Net [54]	15.80	0.466	19.88	0.717
(ECCV'21) PM-Net [53]	16.79	0.510	20.42	0.730
(CVPR'22) DeHamer [36]	16.62	0.560	20.66	0.684
(ICCV'23) FocalNet [58]	17.07	<u>0.630</u>	<u>20.43</u>	<u>0.79</u>
g2D-Net	<u>17.05</u>	0.649	20.22	0.796

4.2. Ablation Experiments

To analyze the critical designs in the g2D-Net, we conducted corresponding ablation experiments. In these experiments, we systematically examined the impact of modules such as the g2D Block, the FFT-g2D Block, the shallow layer, and the SK layer on the model's performance. The results of the ablation experiments are presented in Table 4. Ablation experiments will be performed on the RESIDE-OUT dataset if not otherwise specified.

Table 4. The ablation experiments on the RSEIDE-OUT dataset. Unless otherwise stated, the ablation experiments are performed on the g2D-Net. “-” indicates that the current metric is on par with the baseline. “↑” indicates a performance improvement compared to the baseline, while “↓” indicates a decrease in performance.

Method	RESIDE-OUT			Overhead	
	PSNR	SSIM	MACs (G)	Param (M)	Latency (ms)
g2D-Net (Baseline)	36.07 -	0.983 -	4.839 -	1.452 -	10.51 -
g2D Block (g2D-Net) → GC Block	35.26 ↓	0.983 -	4.240 ↓	1.169 ↓	8.78 ↓
g2D Block (g2D-Net) → GC Block + Conv	35.68 ↓	0.983 -	4.901 ↑	1.371 ↓	9.72 ↓
g2D Block (g2D-Net++) → GC Block	36.17 ↑	0.983 -	6.362 ↑	1.901 ↑	14.41 ↑
FFT-g2D Block → g2D Block	35.06 ↓	0.982 ↓	4.542 ↓	1.153 ↓	9.60 ↓
Multi input/output → Single input/output	35.32 ↓	0.983 -	3.333 ↓	1.217 ↓	8.97 ↓
SK layer → Cat layer	35.31 ↓	0.982 ↓	5.004 ↑	1.471 ↑	10.45 ↓
Depth × 2 (g2D-Net++)	36.60 ↑	0.985 ↑	7.560 ↑	2.446 ↑	18.13 ↑
Width × $\sqrt{2}$	36.32 ↑	0.983 -	7.818 ↑	2.019 ↑	11.47 ↓

We initially investigated the impact of the g2D Block on model performance. The g2D Block contains two gated convolutional units, enabling second-order spatial interactions between feature information. When the g2D Block includes only one gated convolutional unit, it degenerates into a gated convolutional (GC) Block (The architecture of the GC Block is illustrated in Figure 7). The experimental results demonstrate that the g2D Block can increase the PSNR by 0.81 dB compared to the GC Block. To validate that the performance improvement brought on by the g2D Block is not due to an increase in parameters, we conducted another set of experiments: we added a 3×3 depth-wise convolution operator and a 1×1 point-wise convolution in the GC Block to match the parameters and MACs of the g2D Block. However, compared to the g2D Block, adding the convolution operator to the GC Block still decreased the PSNR by 0.39 dB. The utilization of the FFT-g2D Block is aimed at efficiently extracting global and local features. When replacing the g2D Block in the fourth stage with the FFT-g2D Block, the PSNR increases by 1.01 dB.

In the g2D-Net, we employ a multi-input/output strategy to alleviate training difficulty. The role of the shallow layer is to input images of different sizes into the model. If abandoning the multi-input/output strategy, this results in a decrease of 0.75 dB in the PSNR.

The SK layer is incorporated to introduce channel attention to the model, dynamically combining feature map information from different branches. Compared to the commonly used cascaded fusion layers in a traditional U-Net, the SK layer, as a lightweight module,

introduces no additional computational overhead to the model while enhancing the PSNR by 0.76.

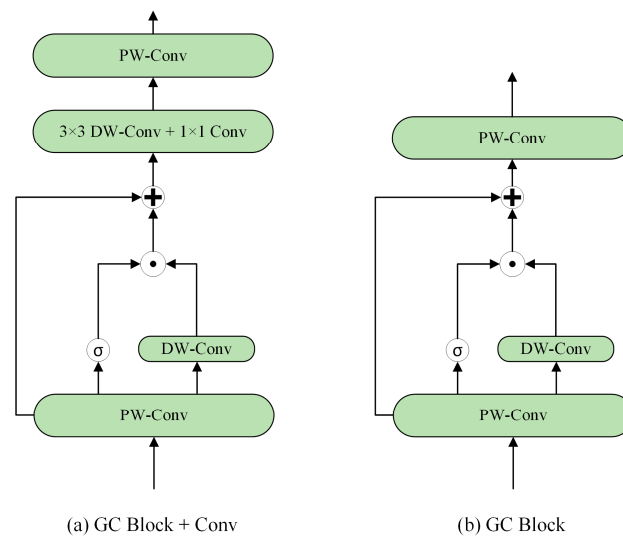


Figure 7. (a) The architecture of the GC Block with a convolution operator. (b) The architecture of the GC Block.

In ablation experiments, we also validated the scalability of the g2D-Net. The experiments show that regardless of expanding the depth or the width, the g2D-Net's performance significantly improves, with deepening the network depth being a more recommended choice.

5. Conclusions

In this study, we propose a lightweight convolutional neural network for image dehazing tasks: the g2D-Net. Inspired by a Vision Transformer, we propose the g2D Block and the FFT-g2D Block, two convolutional residual blocks with input-adaptive and long-range capabilities. In addition, we utilize the SK layer to improve the model performance further and adopt the multi-input/output strategy to reduce the model's training difficulty. Extensive experiments demonstrate that the g2D-Net achieves a balance between performance and computational complexity, delivering SOTA performance on multiple benchmark datasets with low amounts of computational overhead. This lightweight model with excellent performance effectively reduces the difficulty of model training and deployment, promoting the application and development of dehazing networks in real-world scenarios. Although the g2D-Net's performance is impressive, its performance on large-scale datasets, such as RESIDE-OUT, still cannot match that of the SOTA large-scale dehazing models due to network size limitations. Additionally, constrained by the quality and scale of the dataset, the g2D-Net's effectiveness in handling real-world haze still needs improvement. However, with a deeper understanding of neural networks and improved dataset quality, the g2D-Net's performance is poised to enhance further.

Author Contributions: Conceptualization, Z.W.; methodology, Z.W.; software, Z.W.; validation, J.J.; formal analysis, J.M.; investigation, J.J.; resources, J.M.; data curation, J.J.; writing—original draft preparation, J.J.; writing—review and editing, Z.W.; visualization, J.J.; supervision, J.M.; project administration, J.M.; funding acquisition, J.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available in this article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Comput.* **1989**, *1*, 541–551. [\[CrossRef\]](#)
2. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 30 June–1 July 2016; pp. 770–778.
3. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008.
4. Liu, Z.; Mao, H.; Wu, C.-Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A Convnet for the 2020s. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 11976–11986.
5. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Virtual, 11–17 October 2021; pp. 10012–10022.
6. Dauphin, Y.N.; Fan, A.; Auli, M.; Grangier, D. Language Modeling with Gated Convolutional Networks. In Proceedings of the International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; PMLR: New York, NY, USA, 2017; pp. 933–941.
7. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 6–12 June 2015; pp. 3431–3440.
8. Rao, Y.; Zhao, W.; Zhu, Z.; Lu, J.; Zhou, J. Global Filter Networks for Image Classification. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 980–993.
9. Li, X.; Wang, W.; Hu, X.; Yang, J. Selective Kernel Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 510–519.
10. Cho, S.-J.; Ji, S.-W.; Hong, J.-P.; Jung, S.-W.; Ko, S.-J. Rethinking Coarse-to-Fine Approach in Single Image Deblurring. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Virtual, 11–17 October, 2021; pp. 4641–4650.
11. Gonzalez, R.C. *Digital Image Processing*; Pearson Education India: Delhi, India, 2009; ISBN 81-317-2695-9.
12. Seow, M.-J.; Asari, V.K. Ratio Rule and Homomorphic Filter for Enhancement of Digital Colour Image. *Neurocomputing* **2006**, *69*, 954–958. [\[CrossRef\]](#)
13. Land, E.H.; McCann, J.J. Lightness and Retinex Theory. *Josa* **1971**, *61*, 1–11. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Dippel, S.; Stahl, M.; Wiemker, R.; Blaffert, T. Multiscale Contrast Enhancement for Radiographies: Laplacian Pyramid versus Fast Wavelet Transform. *IEEE Trans. Med. Imaging* **2002**, *21*, 343–353. [\[CrossRef\]](#) [\[PubMed\]](#)
15. He, K.; Sun, J.; Tang, X. Single Image Haze Removal Using Dark Channel Prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *33*, 2341–2353.
16. Zhu, Q.; Mai, J.; Shao, L. A Fast Single Image Haze Removal Algorithm Using Color Attenuation Prior. *IEEE Trans. Image Process.* **2015**, *24*, 3522–3533. [\[PubMed\]](#)
17. Li, Z.; Zheng, J. Edge-Preserving Decomposition-Based Single Image Haze Removal. *IEEE Trans. Image Process.* **2015**, *24*, 5432–5441. [\[CrossRef\]](#)
18. Zhu, Z.; Wei, H.; Hu, G.; Li, Y.; Qi, G.; Mazur, N. A Novel Fast Single Image Dehazing Algorithm Based on Artificial Multiexposure Image Fusion. *IEEE Trans. Instrum. Meas.* **2020**, *70*, 5001523. [\[CrossRef\]](#)
19. Ancuti, C.O.; Ancuti, C.; Bekaert, P. Effective Single Image Dehazing by Fusion. In Proceedings of the 2010 IEEE International Conference on Image Processing, Hong Kong, China, 26–29 September 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 3541–3544.
20. Zhao, D.; Xu, L.; Yan, Y.; Chen, J.; Duan, L.-Y. Multi-Scale Optimal Fusion Model for Single Image Dehazing. *Signal Process. Image Commun.* **2019**, *74*, 253–265. [\[CrossRef\]](#)
21. Galdran, A.; Vazquez-Corral, J.; Pardo, D.; Bertalmio, M. Fusion-Based Variational Image Dehazing. *IEEE Signal Process. Lett.* **2016**, *24*, 151–155. [\[CrossRef\]](#)
22. McCartney, E.J. *Optics of the Atmosphere: Scattering by Molecules and Particles*; John Wiley and Sons, Inc.: New York, NY, USA, 1976.
23. Narasimhan, S.G.; Nayar, S.K. Vision and the Atmosphere. *Int. J. Comput. Vis.* **2002**, *48*, 233–254. [\[CrossRef\]](#)
24. Nayar, S.K.; Narasimhan, S.G. Vision in Bad Weather. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999; IEEE: Piscataway, NJ, USA, 1999; Volume 2, pp. 820–827.
25. Cai, B.; Xu, X.; Jia, K.; Qing, C.; Tao, D. Dehazenet: An End-to-End System for Single Image Haze Removal. *IEEE Trans. Image Process.* **2016**, *25*, 5187–5198. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Ren, W.; Liu, S.; Zhang, H.; Pan, J.; Cao, X.; Yang, M.-H. Single Image Dehazing via Multi-Scale Convolutional Neural Networks. In Proceedings of the Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Proceedings, Part II 14; Springer: Berlin/Heidelberg, Germany, 2016; pp. 154–169.
27. Li, B.; Peng, X.; Wang, Z.; Xu, J.; Feng, D. Aod-Net: All-in-One Dehazing Network. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 4770–4778.
28. Tu, Z.; Talebi, H.; Zhang, H.; Yang, F.; Milanfar, P.; Bovik, A.; Li, Y. Maxim: Multi-Axis Mlp for Image Processing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5769–5780.
29. Song, Y.; Zhou, Y.; Qian, H.; Du, X. Rethinking Performance Gains in Image Dehazing Networks. *arXiv* **2022**, arXiv:2209.11448.
30. Lu, L.; Xiong, Q.; Chu, D.; Xu, B. MixDehazeNet: Mix Structure Block for Image Dehazing Network. *arXiv* **2023**, arXiv:2305.17654.

31. Chen, Z.; He, Z.; Lu, Z.-M. DEA-Net: Single Image Dehazing Based on Detail-Enhanced Convolution and Content-Guided Attention. *IEEE Trans. Image Process.* **2024**, *33*, 1002–1015. [[CrossRef](#)] [[PubMed](#)]
32. Cui, Y.; Tao, Y.; Bing, Z.; Ren, W.; Gao, X.; Cao, X.; Huang, K.; Knoll, A. Selective Frequency Network for Image Restoration. In Proceedings of the Eleventh International Conference on Learning Representations, Virtual, 25–29 April 2022.
33. Engin, D.; Genç, A.; Kemal Ekenel, H. Cycle-Dehaze: Enhanced Cyclegan for Single Image Dehazing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 825–833.
34. Singh, A.; Bhawe, A.; Prasad, D.K. Single Image Dehazing for a Variety of Haze Scenarios Using Back Projected Pyramid Network. In Proceedings of the Computer Vision—ECCV 2020 Workshops, Glasgow, UK, 23–28 August 2020; Proceedings, Part IV 16. Springer: Berlin/Heidelberg, Germany, 2020; pp. 166–181.
35. Wu, H.; Liu, J.; Xie, Y.; Qu, Y.; Ma, L. Knowledge Transfer Dehazing Network for Nonhomogeneous Dehazing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 478–479.
36. Guo, C.-L.; Yan, Q.; Anwar, S.; Cong, R.; Ren, W.; Li, C. Image Dehazing Transformer with Transmission-Aware 3d Position Embedding. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5812–5820.
37. Song, Y.; He, Z.; Qian, H.; Du, X. Vision Transformers for Single Image Dehazing. *IEEE Trans. Image Process.* **2023**, *32*, 1927–1941. [[CrossRef](#)] [[PubMed](#)]
38. Hore, A.; Ziou, D. Image Quality Metrics: PSNR vs. SSIM. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 2366–2369.
39. 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.
40. Chen, L.; Chu, X.; Zhang, X.; Sun, J. Simple Baselines for Image Restoration. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 17–33.
41. Yu, J.; Lin, Z.; Yang, J.; Shen, X.; Lu, X.; Huang, T.S. Free-Form Image Inpainting with Gated Convolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 4471–4480.
42. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861.
43. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.-H.; Shao, L. Learning Enriched Features for Fast Image Restoration and Enhancement. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 1934–1948. [[CrossRef](#)]
44. 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](#)]
45. Liu, Y.; Zhu, L.; Pei, S.; Fu, H.; Qin, J.; Zhang, Q.; Wan, L.; Feng, W. From Synthetic to Real: Image Dehazing Collaborating with Unlabeled Real Data. In Proceedings of the 29th ACM International Conference on Multimedia, Virtual, 20–24 October 2021; pp. 50–58.
46. Ancuti, C.O.; Ancuti, C.; Sbert, M.; Timofte, R. Dense-Haze: A Benchmark for Image Dehazing with Dense-Haze and Haze-Free Images. In Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 22–25 September 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1014–1018.
47. Ancuti, C.O.; Ancuti, C.; Timofte, R. NH-HAZE: An Image Dehazing Benchmark with Non-Homogeneous Hazy and Haze-Free Images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 444–445.
48. Loshchilov, I.; Hutter, F. Sgdr: Stochastic Gradient Descent with Warm Restarts. *arXiv* **2016**, arXiv:1608.03983.
49. Loshchilov, I.; Hutter, F. Decoupled Weight Decay Regularization. *arXiv* **2017**, arXiv:1711.05101.
50. Qin, X.; Wang, Z.; Bai, Y.; Xie, X.; Jia, H. FFA-Net: Feature Fusion Attention Network for Single Image Dehazing. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 11908–11915.
51. Liu, X.; Ma, Y.; Shi, Z.; Chen, J. Griddehazenet: Attention-Based Multi-Scale Network for Image Dehazing. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 7314–7323.
52. 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, Seattle, WA, USA, 13–19 June 2020; pp. 2157–2167.
53. Ye, T.; Jiang, M.; Zhang, Y.; Chen, L.; Chen, E.; Chen, P.; Lu, Z. Perceiving and Modeling Density Is All You Need for Image Dehazing. *arXiv* **2021**, arXiv:2111.09733.
54. 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, Nashville, TN, USA, 19–25 June 2021; pp. 10551–10560.
55. Luo, P.; Xiao, G.; Gao, X.; Wu, S. LKD-Net: Large Kernel Convolution Network for Single Image Dehazing. In Proceedings of the 2023 IEEE International Conference on Multimedia and Expo (ICME), Brisbane, Australia, 10–14 July 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1601–1606.

56. Cui, Y.; Knoll, A. Exploring the Potential of Channel Interactions for Image Restoration. *Knowl. Based Syst.* **2023**, *282*, 111156. [[CrossRef](#)]
57. Chao, Q.; Yan, J.; Sun, T.; Li, S.; Chi, J.; Yang, G.; Chen, C.; Yu, T. Instance-Aware Image Dehazing. *Eng. Appl. Artif. Intell.* **2024**, *133*, 108346. [[CrossRef](#)]
58. Cui, Y.; Ren, W.; Cao, X.; Knoll, A. Focal Network for Image Restoration. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2–3 October 2023; pp. 13001–13011.

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.