electronics

MDPI

Article

Semi-Supervised Image-Dehazing Network Based on a Trusted Library

Wan Li [1,2,*] and Chenyang Chang [1,2]

[1] Shaanxi Joint Laboratory of Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an 710021, China
[2] School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an 710021, China
* Correspondence: liwan0_0@163.com

Abstract

In the field of image dehazing, many deep learning-based methods have demonstrated promising results. However, these methods often neglect crucial frequency-domain information and rely heavily on labeled datasets, which limits their applicability to real-world hazy images. To address these issues, we propose a semi-supervised image-dehazing network based on a trusted library (WTS-Net). We construct a dual-branch wavelet transform network (DBWT-Net). It fuses high- and low-frequency features via a frequency-mixing module and enhances global context through attention mechanisms. Building on DBWT-Net, we embed this backbone in a teacher–student model to reduce reliance on labeled data. To enhance the reliability of the teacher network, we introduce a trusted library guided by NR-IQA. In addition, we employ a two-stage training strategy for the network. Experiments show that WTS-Net achieves superior generalization and robustness in both synthetic and real-world dehazing scenarios.

Keywords: wavelet transforms; student-teacher model; semi-supervised learning; image dehazing

check for updates

Academic Editor: M-Tahar Kechadi

Received: 25 May 2025
Revised: 12 July 2025
Accepted: 22 July 2025
Published: 24 July 2025

Citation: Li, W.; Chang, C. Semi-Supervised Image-Dehazing Network Based on a Trusted Library. *Electronics* **2025**, *14*, 2956. https://doi.org/10.3390/electronics14152956

Copyright: © 2025 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

Hazy weather often occurs in daily life. It also reduces the visual information available in captured images due to global or local blurring, color degradation, and reduced depth of field. This affects the accuracy of subsequent tasks and becomes a major challenge in applications such as remote semantic segmentation, sensing-image analysis, and object detection. Accordingly, image-dehazing methods have drawn growing interest in recent years.

Early research mainly relied on dehazing methods based on handcrafted priors and physical models. Representative methods include the Dark Channel Prior (DCP) [1], non-local prior [2], and color-attenuation prior [3]. However, these traditional methods struggled to handle the variability and complexity of real-world haze. Researchers shifted their focus toward deep learning-based methods , with resulting performance improvements. CNNs that have been employed to predict these model parameters include AOD-Net [4] and DehazeNet [5]. Subsequently, end-to-end networks such as FFA-Net [6], AECR-Net [7], and DeHamer [8] were developed to map hazy inputs directly to dehazed outputs. Although these methods [4–10] have achieved promising results, they frequently overlook frequency-domain information. To address this limitation, many methods such as DWSR [11], MWCNN [12], and DW-GAN [13] have incorporated frequency-domain

information into image reconstruction, yielding promising outcomes. Meanwhile, semi-supervised and unsupervised learning have garnered attention for their ability to narrow the distribution gap between synthetic and real-world data. Researchers have proposed a variety of unsupervised loss functions [14] and explored mechanisms such as the teacher–student model [15] and knowledge distillation [16]. In the image-reconstruction domain, some methods employing the teacher–student model have achieved promising results; examples include MCMT [17] and Semi-UIR [18]. However, effectively enhancing model robustness and generalization in real-world hazy conditions remains a critical open challenge.

Therefore, image dehazing faces two major challenges. First, haze causes degradation of image texture and edge information, leading to loss of crucial visual details. Second, existing networks heavily rely on labeled datasets, which limits their ability to handle images in real hazy scenarios. To address these challenges, we introduce a semi-supervised image-dehazing network based on a trusted library (WTS-Net). We design the dual-branch wavelet transform network (DBWT-Net) to integrate frequency-domain information. A frequency-mixing module based on discrete wavelet transform is employed to preserve frequency information and restore fine details, while an attention mechanism is adopted to capture global image context and adapt to different haze densities. Then, we embed DBWT-Net within a teacher–student model to leverage unlabeled real-world images. To stabilize pseudo-label quality, we introduce a trusted library guided by non-reference IQA to select only high-quality labels. During the network-training process, a two-stage training strategy is applied to enhance performance on real-world hazy scenes. This design effectively facilitates adaptation from synthetic to real hazy images, significantly enhancing generalization capability and robustness.

Our contributions can be summarized as follows:

- The dual-branch wavelet transform network: A wavelet-based dual-branch architecture that preserves high-frequency details and enhances global feature extraction for better dehazing.
- Two-stage semi-supervised training: Stabilizes the teacher network using EMA in the first stage and refines pseudo-labels via a trusted library in the second stage.
- Real-World-Feature Adaptation: Our method that enables effective feature transfer from synthetic to real hazy images, improving generalization and robustness.

2. Related Work

2.1. Single Image Dehazing

Image dehazing has become a critical challenge in computer vision, attracting widespread research interest. Many effective dehazing methods can be broadly categorized into two classes: prior-based methods and deep learning methods. Prior-based dehazing methods, exemplified by the Dark Channel Prior (DCP), usually rely on a physical model and handcrafted priors. However, inaccurate inference of atmospheric and transmission parameters can compromise dehazing results. In the early stages of deep learning, researchers attempted to combine deep learning with prior-based approaches to optimize image dehazing. Typical examples include DehazeNet, AOD-Net, and DCPDN [19]. In recent years, supervised learning methods driven by labeled image datasets have also achieved remarkable performance. For example, FFA-Net improved image dehazing by introducing channel and pixel attention. AECR-Net introduced a new contrastive loss between its dehazed results and the true reference images. DeHamer utilized depth information and an attention mechanism to enhance dehazing performance.

2.2. Image Frequency-Domain Learning

Traditional image processing analyzes images in two domains, the spatial domain, which deals with pixel-level information, and the frequency domain, which captures structural features like edges and textures. Common frequency-domain analysis methods include fourier transform and wavelet transform. Recent studies have shown that integrating frequency-domain information into deep learning can significantly improve network performance. For example, DWSR predicts missing details for restoration in high-resolution images through wavelet transform. MWCNN combines multi-layer discrete wavelet transforms and CNN layers to effectively preserve detailed information in the super-resolution process. DW-GAN incorporates an up-sampling and a down-sampling module using wavelet transform to enhance high-frequency preservation. SFSNiD [20] introduces a spatial-frequency exchange module based on Fourier transform to address nighttime dehazing.

2.3. Semi-Supervised Learning

Although supervised methods achieve high performance on synthetic datasets, their generalization to real-world hazy scenarios remains limited. Therefore, semi-supervised and unsupervised learning have gained increasing research attention [17,18,20,21]. Examples include virtual adversarial training, knowledge distillation, and, in particular, the teacher–student model. In the teacher-student model, a student network $f_s(\bullet)$ is optimized by standard gradient descent, whereas the teacher network $f_t(\bullet)$ is updated via an EMA of its previous weights and the student's weights. The update method is as follows:

$$\theta_t = \mu\theta_t + (1-\mu)\theta_s \tag{1}$$

where $\mu \in (0,1)$ is the momentum coefficient and θ_t, θ_s are the weights of the teacher and student networks. Representative applications of this approach include DMT-Net [22], introduced as a disentangled image-dehazing network based on a teacher-student model. MCMT, another low-light enhancement network based on a teacher–student model, introduces a multi-consistency regularization loss. Semi-UIR is also a semi-supervised method for processing underwater images.

3. Method

In this section, Section 3.1 introduces DBWT-Net, a dual-branch wavelet transform network. Section 3.2 presents WTS-Net, which embeds DBWT-Net within a teacher–student model with a trusted library guided by NR-IQA.

3.1. Dual-Branch Wavelet Transform Network

The dual-branch network structure extracts distinct features for feature fusion, enhancing network performance. Therefore, we designed DBWT-Net with two branches. Its structure is shown in Figure 1. The Hybrid Information-Learning Branch $\mathcal{W}(\bullet)$ employs wavelet transforms to capture frequency-domain information from the input image, and the Feature Knowledge Adaptive Branch $\mathcal{A}(\bullet)$ uses the attention mechanism to obtain global contextual information. DBWT-Net can be formally described by the following equations:

$$\begin{aligned} F_{\mathrm{HI}} = \mathcal{W}(x), \quad F_{\mathrm{FA}} = \mathcal{A}(x), \\ F_{\mathrm{fuse}} = Concat(F_{\mathrm{HI}}, F_{\mathrm{FA}}), \\ \hat{x} = \mathcal{D}(F_{\mathrm{fuse}}). \end{aligned} \tag{2}$$

where x is the hazy image, $\mathcal{D}(\bullet)$ is the fusion convolutional block, and $\hat{x}$ is the dehazed image.

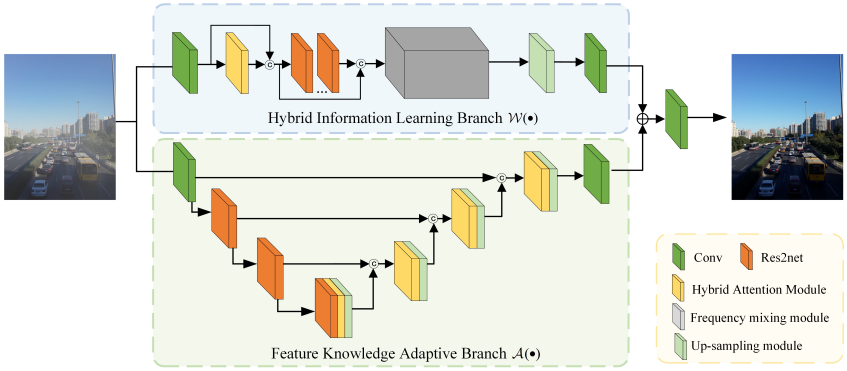

Figure 1. The structure of the DBWT-Net.

3.1.1. Hybrid Information-Learning Branch

The hybrid information-learning branch analyzes the frequency-domain information of images. It employs discrete wavelet transform to extract and fuse frequency components, thereby preserving and enhancing fine details.

To more accurately extract frequency-domain information, the hazy image x is processed to obtain the intermediate feature x_m:

$$x_m = Res(Att(Conv(x))) \tag{3}$$

where $Conv(\bullet)$ is the convolution module, $Res(\bullet)$ is the Res2Net module, and $Att(\bullet)$ is the hybrid attention module.

$Res(\bullet)$ represents multi-scale features in a hierarchical structure with almost no increase in parameter count. The intermediate feature x_m is then passed through a frequency-mixing module. $Att(\bullet)$ includes pixel attention and channel attention. After the feature have been passed through $Att(\bullet)$, the network captures both pixel-level and channel-level features, enhancing its ability to perceive critical details. The mechanism preserves fine textures and structural information in the image. The structures of $Res(\bullet)$ and $Att(\bullet)$ are illustrated in Figure 2.

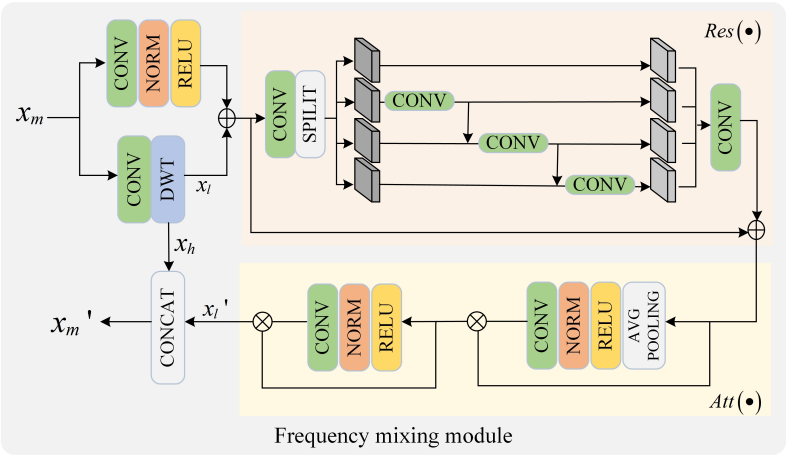

Figure 2. The structure of the Frequency-Mixing Module.

This frequency-mixing module applies the wavelet transform to split the hazy image into high-frequency information x_h and low-frequency information x_l. The high-frequency information x_h is retained. It is then fused with the fully learned low-frequency information

x_l', refining fine details while preserving the image's structural information. Figure 2 describes the module's architecture, and the detailed process is as follows:

$$\begin{aligned} x_l, x_h &= DWT(Conv(x_m)) \\ x_s &= (ReLU(Norm(Conv(x_m))) + x_l) \\ x_l' &= Att(Res(x_s) + x_s) \\ x_m' &= Concat(x_l', x_h) \end{aligned} \tag{4}$$

where $DWT(\bullet)$ is the discrete wavelet transform. The module outputs x_m', which is then passed through an upsampling module $Up(\bullet)$ followed by a convolution $Conv(\bullet)$ to produce the branch output F_{HI}.

3.1.2. Feature-Knowledge Adaptive Branch

This branch employs attention mechanisms to capture global contextual dependencies across the entire image, enabling the network to adapt its feature representations to diverse haze patterns and scene structures. Thus, we refer to the second branch as the feature-knowledge adaptive branch.

A U-Net architecture is adopted in this branch. The down-sampling encoder uses $Res(\bullet)$ as the primary backbone. The up-sampling module in the decoding layer replaces traditional deconvolution with the PixelShuffle operation, which is combined with a mixed attention module $Att(\bullet)$. The hazy image is processed through the $Res(\bullet)$ layer by layer to extract intermediate features with a fixed spatial resolution. These intermediate features are passed through skip connections to the corresponding layers of the decoder and progressively upsampled to produce the final output of the branch F_{FA}.

Finally, the outputs learned from both branches F_{HI}, F_{FA} are fused through a convolution $Conv(\bullet)$ to generate dehazed image $\hat{x}$.

In both branches of DBWT-Net, the $Res(\bullet)$ is implemented using pretrained Res2Net modules. This transfer-learning strategy injects additional knowledge into the network, fostering adaptive feature learning and markedly improving the dehazing model's generalization and adaptability.

3.2. Semi-Supervised Image-Dehazing Network Based on a Trusted Library

We present the WTS-Net architecture in this section. WTS-Net builds upon DBWT-Net as the backbone of a teacher–student model. It integrates a trusted library and adopts a two-stage training strategy to facilitate effective domain adaptation. Its architecture shown in Figure 3.

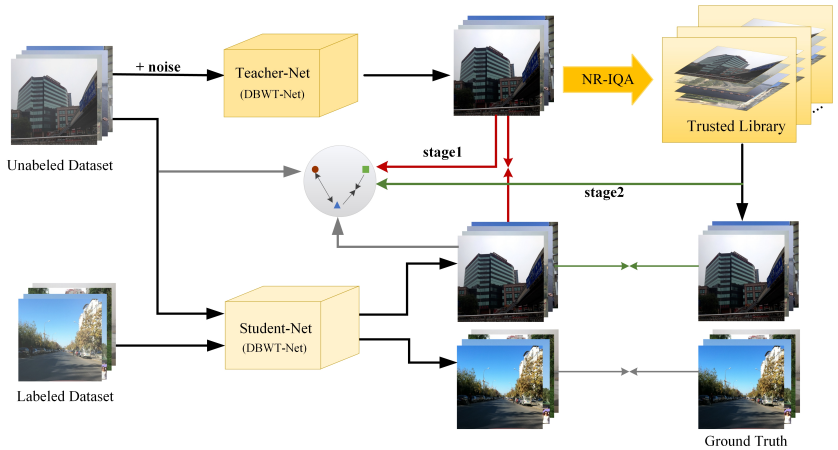

Figure 3. The structure of the WST-Net.

3.2.1. Teacher–Student Model

To reduce reliance on labeled data, WTS-Net adopts the teacher–student model that includes labeled and unlabeled datasets. We define the datasets as $D = (D_L, D_U)$. The datasets D include the labeled dataset $D_L = \{(X_i^l, Y_i^l)\}_{i=1}^M$, which consists of M pairs of hazy and clear images, and the unlabeled dataset $D_U = \{X_i^u\}_{i=1}^N$, which consists of N hazy images.

During training, both labeled hazy image x_i^l and unlabeled hazy image x_i^u are processed by the student network $f_s(\bullet)$, while the teacher network $f_t(\bullet)$ processes only the noisy versions of the unlabeled images x_i^u. For each labeled pair $(x_i^l, y_i^l) \in D_L$, the student output $f_s(x_i^l)$ is compared against the corresponding ground truth y_i^l. Thus, the supervised loss is presented as $L_{sup} = \ell(f_s(x_i^l), y_i^l))$, where $\ell(\bullet, \bullet)$ is the chosen per-sample loss function. For each unlabeled image $x_i^u \in D_U$, we add noise and pass it through the teacher network $f_t(\bullet)$ to obtain a pseudo-label $y_i^u = f_t(x_i^u + noise)$. The student's prediction is then aligned with this pseudo-label via the unsupervised loss $L_{unsup} = \ell(f_s(x_u), y_u)$.

Therefore, WTS-Net defines its total loss as combining the supervised and unsupervised losses, as follows:

$$L = L_{sup} + \alpha L_{unsup} \tag{5}$$

where α balances the supervised and unsupervised losses. The student network's parameters are optimized by backpropagating the combined loss L, while the teacher network's weights are updated via Equation (1), then learned through gradients. In theory, the dehazing effect of y_u is better than that of $f_s(x_u)$, which guides the student network. However, the quality of pseudo-label y_u, as generated by teacher networks, may be unstable, which in turn can adversely affect the student network's performance.

3.2.2. Trusted Library

Semi-UIR [18] addresses pseudo-label instability in the restoration of underwater images. Inspired by this work, we introduce a trusted library for image dehazing. The trusted library stores high-quality pseudo-labels y_u for unlabeled data, making the guidance of the teacher network more reliable.

Since unlabeled data lack ground truth for reference-based evaluation, we employed NR-IQA methods to assess pseudo-label quality. To select the most accurate NR-IQA metric for hazy images, we utilized the RESIDE-OTS dataset [23]. The RESIDE-OTS dataset contains synthesized images at varying haze levels using a global atmospheric-scattering model. We randomly sampled 500 image sets spanning different haze densities. Each set consisted of seven photographs arranged in order of quality, from high to low, as shown in Figure 4. An NR-IQA metric is identified as reliable if its score can accurately track the trend in image-quality decline across a set of images. We conducted a comparative assessment of several leading NR-IQA methods [24–31]. According to the results shown in Figure 5, the PAQ2PIQ [24] evaluation method has the highest accuracy; therefore, this article uses PAQ2PIQ as the method for judging image quality.



Figure 4. A Set of Images Demonstrating Image-Quality Degradation.

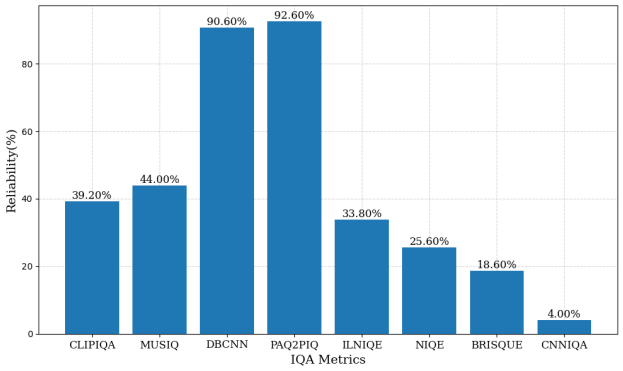

Figure 5. The NR-IQA reliability results.

In our WTS-Net training, we introduce a two-stage training strategy based on a trusted library. In Stage 1, y_u is the dehazed output $f_t(x_u)$ produced by the teacher network for each unlabeled image. During this stage, the trusted library is also dynamically updated. In Stage 2, the pseudo-label y_u is the high-quality image y_r in the trusted library. Our method reduces instability in the teacher network and leverages only the most effective dehazing outputs to guide the student network.

For every unlabeled image $x_u \in D_U$, we initialize its corresponding image y_r in the trusted library. Based on comparative experiments between black- and white-image initializations (results shown in Figure 6), we selected an all-black image as the default initial pseudo-label. During training, the teacher's output $f_t(x_u)$ is compared against both the student's output $f_s(x_u)$ and the current y_r using NR-IQA scores. Only if $f_t(x_u)$ achieves the higher score is y_r updated. This method ensures that the high-quality dehazed outputs generated by the teacher network are stored in the trusted library, thereby providing more effective guidance to the student network.



Figure 6. The trusted library initialization comparison results

3.3. Loss Function

According to Section 3.2.1, the loss function is presented as follows:

$$L = L_{sup} + \alpha L_{unsup} \tag{6}$$

For DBWT-Net, we set $\alpha = 0$ so that only the supervised loss is used. For WTS-Net, $\alpha \in (0, 1)$ gradually increases the proportion of unsupervised loss as training progresses, using an exponential decay method. The supervised loss L_{sup} and unsupervised loss L_{unsup} are presented as follows.

3.3.1. Supervised Loss

For each labeled pair (x_i^l, y_i^l), the student-network output $f_s(x_i^l)$ is produced by the student network $f_s(\bullet)$. Thus, the supervised loss is presented as follows:

$$L_{\text{sup}} = \lambda_{l1} L_{l1} + \lambda_{per} L_{per} + \lambda_{ms} L_{ms} + \lambda_{dcp} L_{dcp} \tag{7}$$

where $\lambda_{l1}, \lambda_{per}, \lambda_{ms}$, and λ_{dcp} are the weighting factors and the weighting factor is set to $\lambda_{l1} = 1.0, \lambda_{per} = 0.2, \lambda_{ms} = 0.2$, and $\lambda_{dcp} = 0.0001$. Each loss is defined as follows:

(1) L1 Loss L_{l1}. The L1 loss enforces pixel-wise fidelity, ensuring the student's dehazed output closely matches the corresponding ground-truth image. The L1 loss is presented as follows:

$$L_{l1} = \sum_{i=1}^{M} |f_s(x_i^l) - y_i^l| \tag{8}$$

(2) Perceptual Loss L_{per}. Apart from beyond pixel-level fidelity, the perceptual similarity is measured using the pre-trained VGG-16 as a loss network. The perceptual loss is presented as follows:

$$L_{per} = \sum_{j=1}^{3} \frac{1}{C_j H_j W_j} \left\| \phi_j(y_i^l) - \phi_j(f_s(x_i^l)) \right\|_2^2 \tag{9}$$

where C_j, H_j, and W_j denote the channel, height, and width information corresponding to the output features map from the j-th extraction layer and ϕ_j denotes the j-th feature-extraction layer.

(3) Multi-Scale Structural-Similarity Loss L_{ms}. This loss quantifies the structural consistency between two images by computing SSIM over multiple scales. It measures how closely the generated image's structural patterns align with those of the reference. The structural similarity is presented as follows:

$$\begin{aligned} \text{SSIM}(i) &= \frac{2\mu_H \mu_G + C_1}{\mu_H^2 + \mu_G^2 + C_1} \cdot \frac{2\sigma_{HG} + C_2}{\sigma_H^2 + \sigma_G^2 + C_2} \\ &= l(i) \cdot \text{c}(i) \cdot \text{s}(i) \end{aligned} \tag{10}$$

where H and G are two same-sized local windows centered at the target pixel in the dehazed and haze-free images; $l(i), c(i)$, and $s(i)$ denote the luminance, contrast, and structural measures, respectively; and C_1 and C_2 are introduced to avoid division by 0. Meanwhile, the MS_SSIM is presented as follows:

$$\text{MS_SSIM}(i) = l_M^{\alpha}(i) \cdot \prod_{m=1}^{M} c^{\beta}(i) s^{\gamma}(i) \tag{11}$$

where α, β, and γ are default parameters; thus, the multiscale structural-similarity loss is constructed and expressed as follows:

$$L_{ms} = 1 - \text{MS_SSIM} \tag{12}$$

(4) Dark-Channel Prior Loss L_{dcp}. The dark-channel prior is based on the observation that, in a haze-free image, every local patch contains at least one pixel whose intensity in at least one color channel is nearly zero. Thus, the $D(\bullet)$ of an image is presented as follows:

$$D(a) = \min_{c \in \{r,g,b\}} \left(\min_{b \in \Omega(a)} (J^c(b)) \right) \tag{13}$$

where $\Omega(a)$ is a local region centered on pixel point a and b is a pixel point in that local region. $J^c(\bullet)$ is the color channel corresponding to the pixel point, so the dark-channel prior loss is presented as follows:

$$L_{dcp} = \frac{1}{M}\sum_{i=1}^{M}\left| D\left(f_s(x_i^l)\right) - D(y_i^l)\right| \tag{14}$$

3.3.2. Unsupervised Loss

Unlabeled haze images x_i^u are processed by the student network to produce $f_s(x_i^u)$. During training, the pseudo-label y_i^u is obtained from $f_t(x_i^u)$ in Stage 1 and then replaced by the stored library image y_i^r in Stage 2. Thus, the unsupervised loss is presented as follows:

$$L_{unsup} = L_{l1} + L_{cr} \tag{15}$$

Each of these losses is defined separately, as follows:

(1) L1 Loss L_{l1}. The L1 loss enforces pixel-wise fidelity, ensuring the student's dehazed output closely matches its corresponding pseudo-label. The L1 loss presented as follows:

$$L_{l1} = \sum_{i=1}^{N} |f_s(x_i^u) - y_i^u\ | \tag{16}$$

(2) Contrastive Loss L_{cr}. Contrastive learning is mainly based on positive and negative samples, which usually correspond to labeled and degraded images in paired data. For unpaired datasets, the pseudo-label y_i^u serves as the positive member of the pair, while the corresponding hazy image acts as the negative member of the pair. Therefore, the contrast loss presented as follows:

$$L_{cr} = \sum_{j=1}^{K}\sum_{i=1}^{N} \omega_j \frac{|\ \varphi_j(f_s(x_i^u)), \varphi_j(y_i^u)\ |}{|\ \varphi_j(f_s(x_i^u)), \varphi_j(x_i^u))\ |} \tag{17}$$

where φ_j is the feature map from the j-th hidden layer of VGG-19 and ω_j is the weight assigned to that layer.

4. Experiments

In this section, the experimental design is presented and the results are reported. Section 4.1 details the implementation; Section 4.2 reports performance on supervised datasets; Section 4.3 evaluates results on unsupervised data; and Section 4.4 concludes with ablation studies.

4.1. Implementation Details

All experiments were performed on a Linux workstation with an NVIDIA RTX 2080 GPU and implemented in PyTorch 1.8 and Python 3.8. The AdamP optimizer was adopted with an initial learning rate of 0.0002, which was halved at predetermined epochs.

During the network-training process, unlabeled data accounted for 50% of the datasets. All training images were randomly cropped to 256 × 256 pixels and normalized to the [0, 1] range.

In this paper, we use multiple datasets. RESIDE-6K contains 6000 synthetic image pairs; NH-Haze [32] contains 55 real image pairs; RESIDE-OTS includes 2061 synthetic pairs; and RESIDE-URHI comprises 2061 real hazy images. For testing, we employ SOTS-Outdoor, five reserved NH-Haze pairs, 500 reserved RESIDE-URHI images, and RESIDE-HTST, which contains ten real hazy images and ten synthetic hazy images.

To evaluate dehazing quality on labeled datasets, we employ the PSNR and SSIM metrics. For unlabeled real hazy datasets, we employ NR-IQA metrics, including CLIPIQA [25], MUSIQ [26], and DBCNN [27], which aim for higher values, and NIQE [28], which aims for smaller values.

4.2. Supervised Dataset Evaluation

To validate dehazing performance, we compare DBWT-Net and WTS-Net qualitatively and quantitatively with DCP, AOD-Net, FFA-Net, DEA-Net [33], FCB-Net [34], CASM [35], and USID-Net [36] across various datasets. AOD-Net, FFA-Net, and DEA-Net use supervised learning, CASM uses semi-supervised learning, and USID-Net uses unsupervised learning.

We first train DBWT-Net on the full labeled datasets and then train WTS-Net using half of these labels as the unlabeled dataset. For RESIDE-6K, we use SOTS-Outdoor as the test set to evaluate dehazing performance under reduced label reliance. DBWT-Net is trained for 400 epochs, while WTS-Net is trained for 400 epochs in Stage 1 and 50 epochs in Stage 2. For NH-Haze, with five randomly chosen images withheld for testing, DBWT-Net was trained for 90 epochs, while WTS-Net was trained for 30 epochs in Stage 1 and 25 epochs in Stage 2.

Table 1 presents quantitative comparisons of our model against competing methods on SOTS-outdoor and NH-Haze datasets. Figures 7 and 8 visualize the dehazing results produced by each algorithm on these datasets.

Table 1. Results for the SOTS-outdoor dataset and NH-HAZE dataset.

Method	SOTS-Outdoor		NH-Haze	
	SSIM	**PSNR**	**SSIM**	**PSNR**
DCP	0.815	19.13	0.520	10.57
AOD-Net	0.927	20.08	0.569	15.40
FFA-Net	0.984	33.57	0.692	19.87
DEA-Net	0.980	31.68	-	-
FCB-Net	0.958	28.19	0.622	14.16
CASM	0.873	19.87	0.532	13.33
USID-Net	0.919	23.89	0.556	13.21
DBWT-Net (Ours)	0.974	30.59	0.741	20.34
WTS-Net (Ours)	0.961	28.47	0.677	19.64

In the SOTS-Outdoor dataset, we observe that DCP, AOD-Net, CASM, and USID-Net achieve lower PSNR and SSIM values, indicating their limited dehazing capabilities. In contrast, FFA-Net, DEA-Net, and our DBWT-Net demonstrate superior dehazing performance. Notably, our semi-supervised approach, WTS-Net, achieves the highest PSNR (28.47) and SSIM (0.961) among similar algorithms, highlighting its strong dehazing ability. Visual comparisons further confirm that our method preserves finer details, particularly in sky regions, as seen in the red-boxed areas.

On the NH-Haze dataset, DBWT-Net reaches the highest scores, with a PSNR of 20.34 and an SSIM of 0.741. Likewise, among semi-supervised methods, WTS-Net attains the best qualitative and quantitative dehazing results. Although DCP removes haze to a certain extent, it introduces a significant color shift. While AOD-Net, FFA-Net, and DEA-Net leave behind perceptible haze, FCB-Net, DBWT-Net, and WTS-Net achieve superior haze removal. FCB-Net's outputs sometimes have a pale yellow cast. To achieve a more intuitive visual contrast effect, We select the red-boxed areas. Overall, DBWT-Net and WTS-Net yield the most realistic dehazing effects.

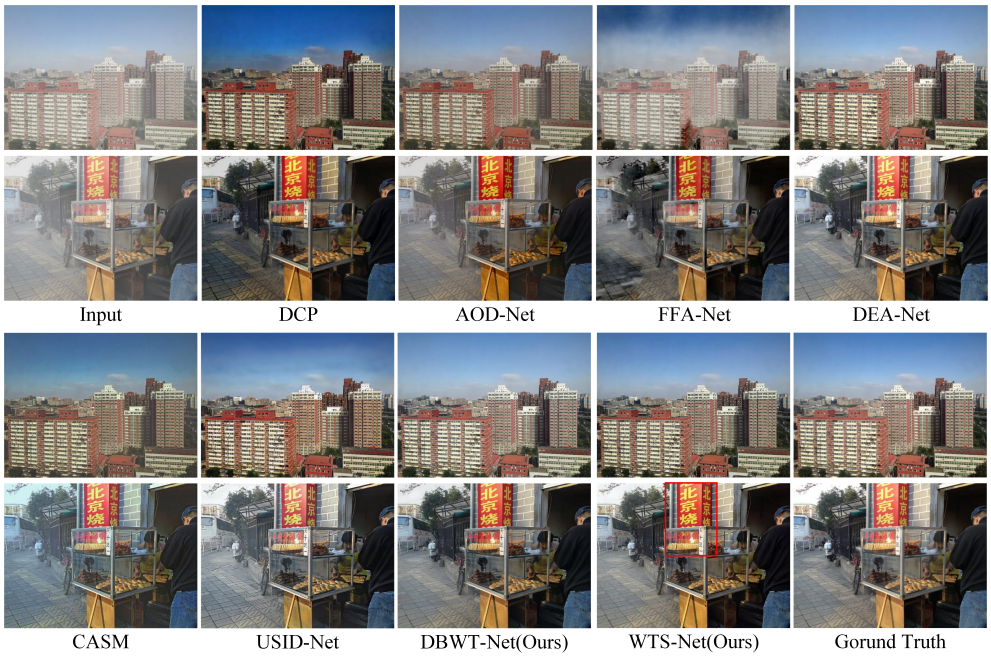

Figure 7. Visual comparison using the SOTS-indoor dataset.

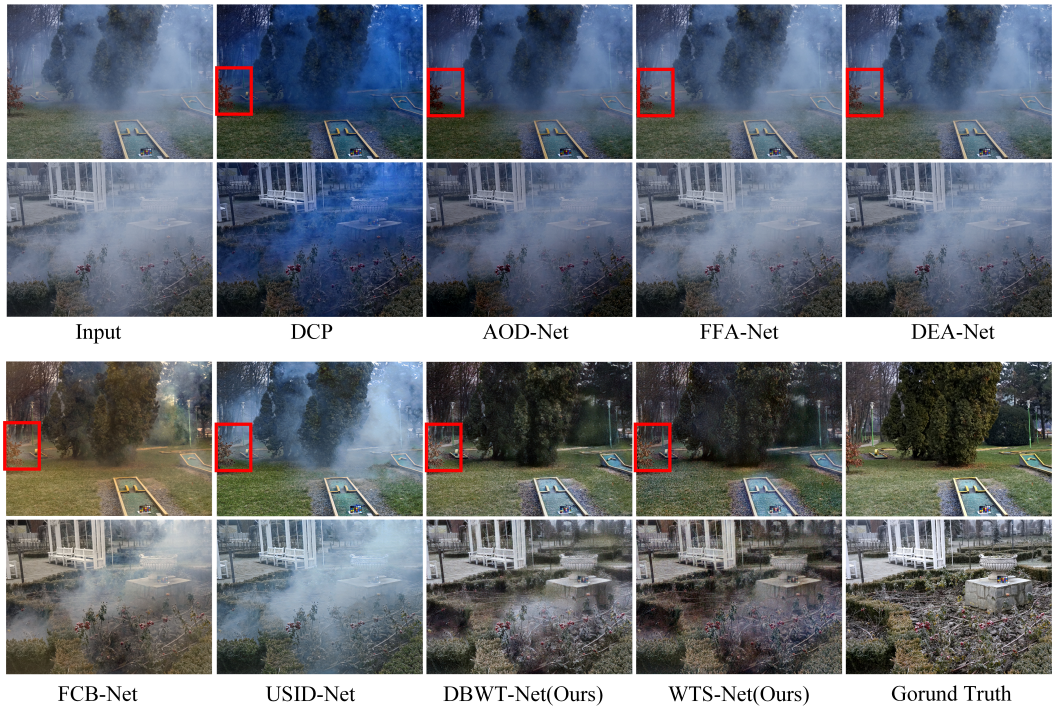

Figure 8. Visual comparison using the NH-HAZE dataset.

4.3. Unsupervised Dataset Evaluation

To verify feature transfer on real-world images, we train WTS-Net with 2061 synthetic pairs from RESIDE-OTS as the labeled dataset and 2061 real hazy images from RESIDE-URHI as the unlabeled dataset, reserving 500 URHI images for testing. WTS-Net is then trained for 80 epochs in Stage 1 and 45 epochs in Stage 2.

By including unlabeled real haze data in our network, we test and quantify the dehazing effectiveness on real hazy images. Table 2 summarizes the quantitative outcomes for RESIDE-URHI, and Figure 9 presents the corresponding visual comparisons. According to the quantitative results, MUSIQ and DBCNN achieve the highest scores, while our

method performs competitively across various metrics. Qualitatively, WTS-Net effectively removes haze while maintaining image hues close to those of the original. Additionally, WTS-Net excels in preserving fine texture details, demonstrating its ability to enhance image clarity while retaining structural information. Compared to other approaches, WTS-Net achieves superior dehazing performance in both global appearance and local details.

Table 2. Results for the RESIDE-URHI dataset. The ↑ symbol indicates that higher values are better, and the ↓ symbol indicates that lower values are better.

Method	CLIPIQA ↑	MUSIQ ↑	DBCNN ↑	NIQE ↓
CASM	0.4460	58.2828	0.4665	4.3039
USID-Net	0.4793	58.6729	0.4678	3.8499
WTS-Net (Ours)	0.4470	58.7700	0.5071	4.5336

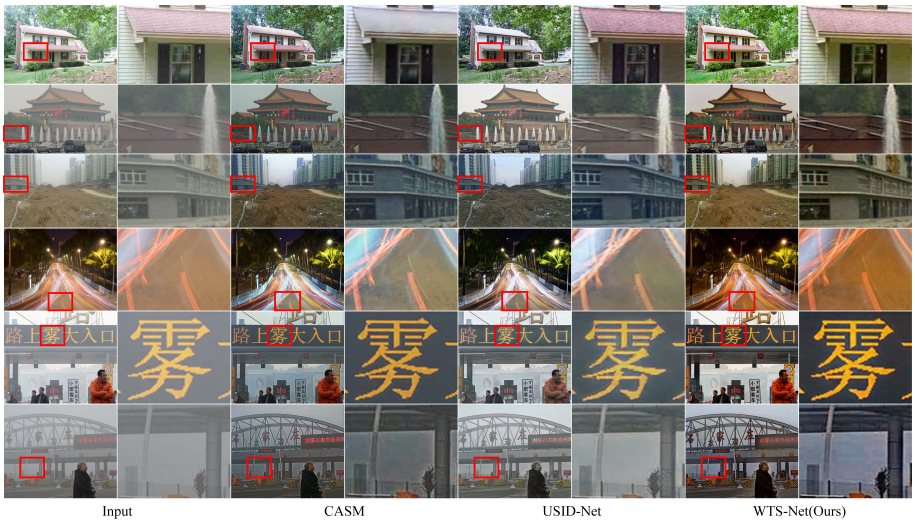

Figure 9. Visual comparison on the RESIDE-URHI dataset. The non-English term in the fifth row is a Chinese character meaning hazy.

To further evaluate WTS-Net's robustness, we test its performance on the RESIDE-HTST dataset. The quantitative findings are summarized in Table 3, and Figure 10 visualizes the comparison of performance on the RESIDE-HTST dataset. The results indicate that WTS-Net surpasses CSAM in PSNR and USID-Net in SSIM on synthetic data, confirming its superior dehazing performance. On real datasets, WTS-Net achieves the highest scores compared to CLIPIQA, DBCNN, NIQE, and BRISQUE. Subjectively, our method produces more accurate colors and improved deblurring results, outperforming other approaches.

Table 3. Results on the RESIDE-HTST dataset.

Method	Synthetic Images		Real Images			
	SSIM	PSNR	CLIPIQA ↑	MUSIQ ↑	DBCNN ↑	NIQE ↓
CASM	0.8801	28.1852	0.5457	65.9068	0.5449	3.5975
USID	0.8033	29.5860	0.5762	65.2448	0.5141	3.4019
WTS-Net (Ours)	0.8704	29.3823	0.5978	65.7613	0.5705	3.2224



Figure 10. Visual comparison of performance on the RESIDE-HTST dataset.

Overall, these results demonstrate that WTS-Net consistently excels across both synthetic and real-world haze scenarios.

4.4. Ablation Study

We design four ablation groups and evaluate them both quantitatively and qualitatively to validate our semi-supervised approach. They are as follows: (a) DBWT-Net, the base dual-branch wavelet-transform network; (b) Teacher–Student, with DBWT-Net serving as the backbone network in the teacher–student model; (c) Teacher–Student + CL, adding the Contrastive Loss model to the Teacher–Student model; (d) WTS-Net, the full model combining Teacher–Student, contrastive learning, and the trusted library. The labeled data come from RESIDE-OTS, while the unlabeled data come from RESIDE-URHI.

Table 4 summarizes the ablation study's quantitative results. (a) DBWT-Net achieves the best results as measured by PSNR (26.515) and SSIM (0.939) when trained only on synthetic data. (b) Introducing the teacher–student model alone degrades performance, reflecting the challenge of unstable pseudo-labels. (c) Adding contrastive learning recovers some performance, yielding a PSNR of 24.922 and a SSIM of 0.929, but still falls short of the standalone backbone. (d) The full WTS-Net, which combines the teacher–student model, contrastive loss, and our trusted library. It achieves a PSNR of 25.605 and an SSIM of 0.935, demonstrating the trusted library's role in stabilizing pseudo-label quality. Figure 11 shows the visualization results, indicating that the addition of these two modules enhances the network's dehazing performance, as observed from a qualitative perspective. From these visualizations, it is clear that each module, those being contrastive learning and the trusted library, incrementally sharpens details and removes haze. WTS-Net produces the most visually pleasing, artifact-free dehazed images.

Table 4. Results of the ablation study. The ✓ indicates that the corresponding module has been incorporated into the network.

	TS	CL	TL	PSNR	SSIM
DBWT-Net				26.515	0.939
Teacher–Student	✓			24.384	0.901
Teacher–Student + CL	✓	✓		24.922	0.929
WTS-Net	✓	✓	✓	25.605	0.935

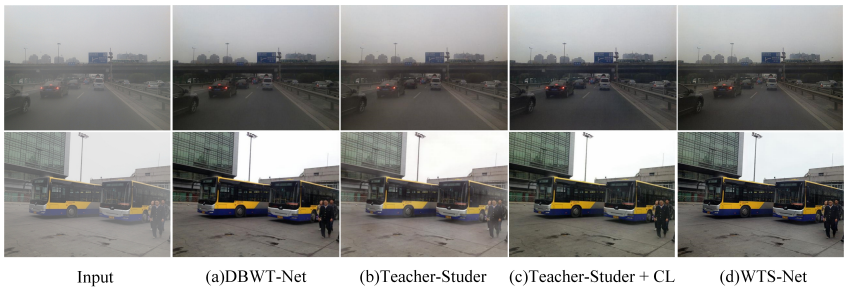

Figure 11. Vhlisual comparison of model ablation experiment.

Furthermore, to demonstrate how our approach lessens reliance on synthetic data and enhances feature adaptation from synthetic to real hazy images, we perform comparative experiments with three distinct training groups: (a) training DBWT-Net using only the synthetic dataset; (b) training WTS-Net using the full set of synthetic dataset; (c) training WTS-Net on a mixed dataset of 50% synthetic dataset and 50% real hazy dataset. The synthetic dataset is sourced from RESIDE-OTS, while the unlabeled dataset is obtained from RESIDE-URHI.

Figure 12 presents the visualization results, as follows: (a) the DBWT-Net trained solely on synthetic data yields severe color shifts and detail distortions when it is applied to real hazy images; (b) incorporating real samples dramatically improves dehazing quality, yielding natural colors and sharp textures; (c) when synthetic and real samples are balanced equally, dehazing performance nearly matches that of the "all synthetic + real" group, with no noticeable degradation in detail or tone. These findings demonstrate that by enforcing cross-domain consistency constraints, WTS-Net markedly lessens its reliance on synthetic data while effectively promoting feature adaptation from synthetic to real hazy images.

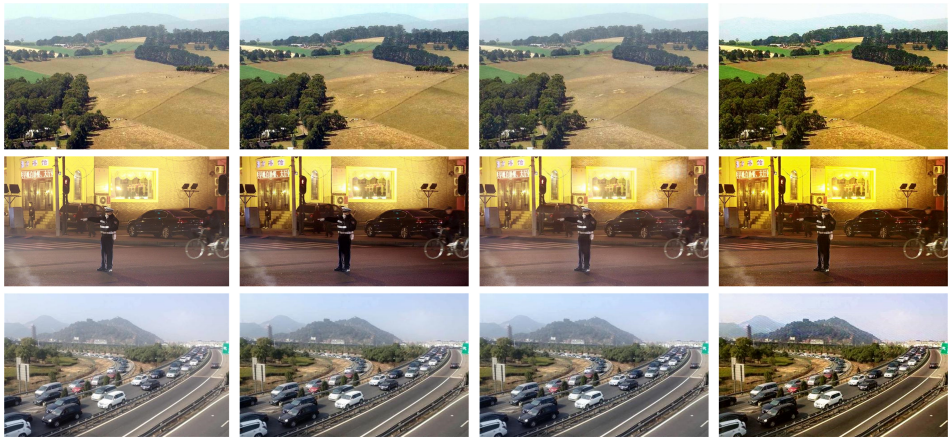

Figure 12. Visual comparison of results generated with different proportions of synthetic training data.

5. Conclusions

We introduce a semi-supervised dehazing algorithm that integrates discrete wavelet transform with a teacher-student model to address the challenges of insufficient utilization of frequency-domain information and reliance on labeled datasets, particularly synthetic data. Our dual-branch wavelet transform network uses a frequency-mixing module to extract and fuse high- and low-frequency features, while an attention mechanism captures global context. We then employ a two-stage training strategy and maintain a trusted

library of high-quality pseudo-labels selected via NR-IQA to stabilize and enhance the teacher network's guidance. This approach reduces dependency on synthetic datasets and facilitates feature adaptation from synthetic to real-world hazy images. Experimental results demonstrate that our method achieves superior performance on both synthetic hazy images and real-world hazy images, exhibiting strong robustness. In future work, we plan to analyze how varying ratios of labeled samples affect dehazing performance, aiming to lessen our dependence on synthetic data. We also intend to extend our approach to other image-restoration tasks.

Author Contributions: Proposed the research idea, designed the research plan, and drafted the initial manuscript, W.L.; Conducted experiments, collected and analyzed data, and reviewed and edited the manuscript, C.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (No. 62101312); Shaanxi University of Science and Technology Natural Science Preliminary Research Fund Project (No. 2019BJ-11).

Data Availability Statement: The data used in this study are publicly available.

Conflicts of Interest: The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. 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]
2. Berman, D.; Avidan, S. Non-local image dehazing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016
3. 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. [CrossRef] [PubMed]
4. 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.
5. 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]
6. Qin, X.; Wang, Z.; Bai, Y.; Xie, X.; Jia, H. FFA-Net: Feature fusion attention network for single image dehazing. *Aaai Conf. Artif. Intell.* **2020**, *34*, 11908–11915. [CrossRef]
7. Wu, H.; Qu, Y.; Lin, S.; Zhou, J.; Qiao, R.; Zhang, Z.; 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.
8. 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.
9. Li, Y.; Miao, Q.; Ouyang, W.; Ma, Z.; Fang, H.; Dong, C.; Quan, Y. LAP-Net: Level-aware progressive 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. 3276–3285.
10. Zhang, Y.; Zhou, S.; Li, H. Depth information assisted collaborative mutual promotion network for single image dehazing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 2846–2855.
11. Guo, T.; Mousavi, H.S.; Vu, T.H.; Monga, V. Deep wavelet prediction for image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 104–113.
12. Liu, P.; Zhang, H.; Zhang, K.; Lin, L.; Zuo, W. Multi-level wavelet-CNN for image restoration. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 773–782.
13. Fu, M.; Liu, H.; Yu, Y.; Chen, J.; Wang, K. Dw-gan: A discrete wavelet transform gan for nonhomogeneous dehazing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 203–212.

14. Li, L.; Dong, Y.; Ren, W.; Pan, J.; Gao, C.; Sang, N.; Yang, M.H. Semi-supervised image dehazing. *IEEE Trans. Image Process.* **2019**, *29*, 2766–2779. [CrossRef] [PubMed]
15. Tarvainen, A.; Valpola, H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *arXiv* **2017**, arXiv:1703.01780.
16. Hinton, G.; Vinyals, O. ; Dean, J. Distilling the knowledge in a neural network. *arXiv* **2015**, arXiv:1503.02531. [CrossRef]
17. Lee, S.; Jang, D.; Kim, D.S. Temporally averaged regression for semi-supervised low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 4208–4217.
18. Huang, S.; Wang, K.; Liu, H.; Chen, J.; Li, Y. Contrastive semi-supervised learning for underwater image restoration via reliable bank. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 18145–18155.
19. Zhang, H.; Patel, V.M. Densely connected pyramid dehazing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 3194–3203.
20. Cong, X.; Gui, J.; Zhang, J.; Hou, J.; Shen, H. A semi-supervised nighttime dehazing baseline with spatial-frequency aware and realistic brightness constraint. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 2631–2640.
21. Wang, W.; Yang, H.; Fu, J.; Liu, J. Zero-reference low-light enhancement via physical quadruple priors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 26057–26066.
22. Duong, M.T.; Lee, S.; Hong, M.C. DMT-Net: Deep multiple networks for low-light image enhancement based on retinex model. *IEEE Access* **2023**, *11*, 132147–132161. [CrossRef]
23. 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]
24. Ying, Z.; Niu, H.; Gupta, P.; Mahajan, D.; Ghadiyaram, D.; Bovik, A. From patches to pictures (PaQ-2-PiQ): Mapping the perceptual space of picture quality. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 3575–3585.
25. Wang, J.; Chan, K.C.; Loy, C.C. Exploring clip for assessing the look and feel of images. *Aaai Conf. Artif. Intell.* **2023**, *37*, 555–2563. [CrossRef]
26. Ke, J.; Wang, Q.; Wang, Y.; Milanfar, P.; Yang, F. Musiq: Multi-scale image quality transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 5148–5157.
27. Zhang, W.; Kede, M.; Yan, J.; Deng, D.; Wang, Z. Blind Image Quality Assessment Using a Deep Bilinear Convolutional Neural Network. *IEEE TRansactions Circuits Syst. Video Technol.* **2020**, *30*, 36–47. [CrossRef]
28. Mittal, A.; Soundararajan, R.; Bovik, A.C. Making a completely blind image quality analyzer. *IEEE Signal Process. Lett.* **2012**, *20*, 209–212. [CrossRef]
29. 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]
30. Zhang, L.; Zhang, L.; Bovik, A.C. A feature-enriched completely blind image quality evaluator. *IEEE Trans. Image Process.* **2015**, *24*, 2579–2591. [CrossRef] [PubMed]
31. Kang, L.; Ye, P.; Li, Y.; Doermann, D. Convolutional neural networks for no-reference image quality assessment. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1733–1740.
32. 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.
33. 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]
34. Wang, J.; Wu, S.; Yuan, Z.; Tong, Q.; Xu, K. Frequency compensated diffusion model for real-scene dehazing. *Neural Netw.* **2024**, *175*, 106281. [CrossRef] [PubMed]
35. Wang, X.; Chen, X.A.; Ren, W.; Han, Z.; Fan, H.; Tang, Y.; Liu, L. Compensation atmospheric scattering model and two-branch network for single image dehazing. *IEEE Trans. Emerg. Top. Comput. Intell.* **2024**, *8*, 2880–2896. [CrossRef]
36. Li, J.; Li, Y.; Zhuo, L.; Kuang, L.; Yu, T. USID-Net: Unsupervised single image-dehazing network via disentangled representations. *IEEE Trans. Multimed.* **2022**, *25*, 3587–3601. [CrossRef]

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.