

基于小波双分支建模的频率感知单幅图像去雾

高梓涵, 魏伟波*, 潘振宽

青岛大学计算机科学技术学院, 山东 青岛

收稿日期: 2026年3月17日; 录用日期: 2026年4月7日; 发布日期: 2026年4月21日

摘要

在单幅图像去雾任务中, 实现有效去雾的同时保持颜色保真度和边缘结构细节仍然是一个具有挑战性的问题。现有的基于深度学习的去雾方法在复杂有雾条件下往往容易出现整体颜色失真、视觉伪影以及边缘模糊等问题。为克服上述局限性, 本文基于频率感知图像去雾框架提出了TripleD-Net图像去雾模型, 该模型利用离散小波变换对图像的低频分量与高频分量进行分离, 低频分量主要表征图像的整体结构和颜色一致性, 因此通过全局上下文建模机制对其进行建模, 以增强对全局语义信息的感知能力; 相比之下, 高频分量包含丰富的边缘结构与纹理细节, 使用多尺度卷积框架对其进行处理, 从而有效保留精细的结构信息。通过对低频分支与高频分支的共同优化, 所提出的方法在颜色一致性、边缘清晰度以及伪影抑制等方面均表现出更优的恢复效果。在RESIDE数据集上的大量实验结果表明, TripleD-Net在定量指标和定性视觉对比方面均优于近期具有代表性的去雾方法。

关键词

图像去雾, 频率感知建模, 深度学习, 小波变换

Frequency-Aware Single Image Dehazing via Wavelet-Based Dual-Branch Modeling

Zihan Gao, Weibo Wei*, Zhenkuan Pan

College of Computer Science and Technology, Qingdao University, Qingdao Shandong

Received: March 17, 2026; accepted: April 7, 2026; published: April 21, 2026

Abstract

In single-image dehazing tasks, achieving effective haze removal while preserving color fidelity and edge structural details remains a challenging problem. Existing deep learning-based dehazing methods

*通讯作者。

文章引用: 高梓涵, 魏伟波, 潘振宽. 基于小波双分支建模的频率感知单幅图像去雾[J]. 图像与信号处理, 2026, 15(2): 271-281. DOI: 10.12677/jisp.2026.152023

often suffer from overall color distortion, visual artifacts, and edge blurring under complex hazy conditions. To overcome these limitations, this paper proposes TripleD-Net, an image dehazing model based on a frequency-aware framework. The model employs the Discrete Wavelet Transform (DWT) to separate image components into low-frequency and high-frequency bands. Since low-frequency components primarily represent the overall structure and color consistency of the image, they are modeled via a global context modeling mechanism to enhance the perception of global semantic information. In contrast, high-frequency components contain rich edge structures and texture details; thus, a multi-scale convolutional framework is utilized to process them, effectively preserving fine structural information. Through the joint optimization of the low-frequency and high-frequency branches, the proposed method demonstrates superior restoration performance in terms of color consistency, edge sharpness, and artifact suppression. Extensive experiments on the RESIDE dataset show that TripleD-Net outperforms recent representative dehazing methods in both quantitative metrics and qualitative visual comparisons.

Keywords

Image Dehazing, Frequency-Aware Modeling, Deep Learning, Wavelet Transform

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

雾是一种在日常环境中频繁出现的大气现象，严重的雾会显著降低能见度并增加事故发生的风险。在图像中，雾的存在会严重影响对图像的使用。有雾情况下采集的低质量图像容易导致后续在高级视觉任务(如目标检测与语义分割)中产生误差。为解释这一图像退化过程，Cantor [1]在 1978 提出了大气散射模型(Atmospheric Scattering Model, ASM)，为图像去雾技术提供了物理基础。该模型表明，随着场景深度或散射强度的增加，透射率呈指数衰减，从而导致图像出现模糊和泛白现象。

自提出以来，大气散射模型已成为大多数传统去雾方法所依赖的主流物理模型。早期研究主要基于先验假设来估计雾相关参数，例如暗通道先验[2] (Dark Channel Prior, DCP)、颜色衰减先验[3] (Color Attenuation Prior, CAP)。暗通道先验是一种具有代表性的去雾方法。其核心机理源于对自然图像统计规律的观察，在绝大多数非天空局部区域内，至少有一个色彩通道的像素值趋于低值。基于这一特性，该算法通过精准估算场景透射率及全局大气光参数，最终实现清晰图像的重建。颜色衰减先验利用有雾环境下图像亮度与颜色饱和度之间的统计关系来估计场景深度信息。该方法基于这样一种观察：伴随场景深度的延伸，图像往往呈现出亮度递增而色彩饱和度显著衰减的趋势。基于这一物理现象，构建亮度、饱和度与深度间的线性映射关系，能够有效推算透射率参数，进而实现无雾图像的复原。与暗通道先验方法相比，CAP 方法计算效率较高，在一些自然场景中能够获得较好的去雾效果，但在复杂光照条件下仍可能出现颜色偏差等问题。这些先验在一定程度上对解空间进行约束，以增强场景可见性。然而，去雾性能在很大程度上取决于所假设先验与真实数据分布之间的一致性。当这些假设条件不满足时，恢复图像往往会出现颜色失真与伪影问题。

近年来，大量不依赖大气散射模型的深度学习方法被广泛应用于图像去雾问题中，并取得了显著成效。在基于卷积神经网络(CNN)的自监督去雾研究中，主流策略通常依赖于成对的有雾与无雾图像数据，旨在通过学习图像域间的映射规律来获取优异的去雾性能。具体而言，Chen 等[4]构建了门控上下文聚合

网络(GCANet), 该模型通过对不同层级的特征赋予自适应权重, 成功实现了多尺度特征的高效融合。此外, Liu 及其团队[5]提出了 GridDehazeNet, 这是一种融合了注意力机制的多尺度估计架构, 进一步提升了去雾效果。该方法在网格结构中实现了不同尺度之间的密集交互, 有效缓解了传统编码器-解码器结构或层次化多尺度网络中存在的信息瓶颈问题。Wu 等[6]引入了具有双重约束的对比正则化方法, 在保持内容一致性的同时, 提升了结构质量与感知质量, 从而有效缓解了仅使用 L1 或 L2 损失函数时常见的细节缺失与颜色失真问题。Liu 等[7]提出了一种融合生成对抗网络(GAN)的图像自增强去雾方法。该方法通过两个生成对抗网络分别学习图像的加雾过程与去雾过程, 并在加雾阶段引入场景深度估计与随机散射系数采样, 以模拟不同浓度的雾气分布。卷积核高频信号具有较强的敏感性, 使得 CNN 在提取边缘、纹理等细节信息方面表现尤为突出。然而, CNN 受限于其固定且有限的感受野, 在长距离上下文依赖关系建模方面能力不足。

注意力机制[8] [9]与 Vision Transformer [10] [11] (Vision Transformer, ViT)的引入进一步推动了去雾研究的发展。Song 等[12]通过改进 Swin Transformer 中的归一化层与激活函数, 解决了去雾过程中存在的特征退化问题。Dong 等[13]在 DCNv4 的基础上提出了 DCNFormer, 通过在空间聚合模块中去除 Softmax 归一化操作, 加速了模型的推理过程。尽管注意力机制能够提升性能, 但其计算与存储开销较大, 且 ViT 在输入分辨率上的二次复杂度进一步加重了计算负担。此外, 近期研究表明, 注意力机制更倾向于强调全局上下文信息[14], 在建模细粒度等高频细节方面可能效果有限。

基于上述观察, 本文在统一的去雾框架中融合了 CNN 与 ViT 的互补优势。其中, CNN 负责捕获局部高频细节信息, 而 ViT 侧重于建模全局低频信息。为实现这种频率感知表示, 首先采用离散小波变换 [15] (Discrete Wavelet Transform, DWT)将输入图像分解为一个低频子带与三个高频子带。通过同时处理这四个子带, 本方法能够更加高效地利用计算资源, 在降低图像模糊程度的同时实现更优的去雾性能。

2. TripleD-Net 的算法设计与实现

图 1 展示了 TripleD-Net 的整体网络架构。对于输入的有雾图像, 首先通过一个 3×3 卷积操作提取浅层特征。设 $F_0 \in \mathbb{R}^{C \times H \times W}$ 表示输入的有雾图像, 其中 H 和 W 分别表示图像大小, C 表示通道数。该过程可表示为:

$$F_1 = \text{Conv}_{3 \times 3}(F_0) \quad (1)$$

其中, $\text{Conv}_{3 \times 3}$ 表示一个常规卷积模块(Conv-BatchNorm-Sigmoid-ReLU)。该层不仅用于特征提取, 同时还对通道维度进行扩展, 以满足后续模块的输入需求。

随后, 特征 F_1 依次经过两次下采样和两次上采样操作, 将整个网络划分为三个阶段。在每一个阶段中均引入频率分离处理模块对特征进行处理, 但不同阶段中该模块的具体结构设计有所差异。并且, 使用 Fusion 模块对特征进行融合, 设输入的两个特征图分别为 x_1 和 x_2 , 首先通过线性映射函数 $f(\cdot)$ 将 x_1 投影得到特征表示 \hat{x}_1 。随后, 对 \hat{x}_1 与 x_2 的逐元素相加结果施加全局平均池化(GAP), 并通过多层感知机(MLP, 即 Linear-ReLU-Linear 结构)进行特征变换, 再经 Softmax 函数归一化, 并通过分割操作获得对应的融合权重 a_1 和 a_2 , 其计算过程如下所示:

$$\{a_1, a_2\} = \text{Split}\left(\text{Softmax}\left(F_{MLP}\left(\text{GAP}\left(\hat{x}_1 + x_2\right)\right)\right)\right) \quad (2)$$

最后, 利用得到的权重对两个分支特征进行加权融合, 同时引入来自 x_2 的残差连接以增强信息传递能力, 得到最终输出:

$$y = a_1 \cdot \hat{x}_1 + a_2 \cdot x_2 + x_2 \quad (3)$$

该融合策略能够根据输入特征的全局语义信息自适应地调整各分支的重要性，从而有效提升特征表达能力并增强网络的表示鲁棒性。

2.1. 频率分离处理模块(FDPB)

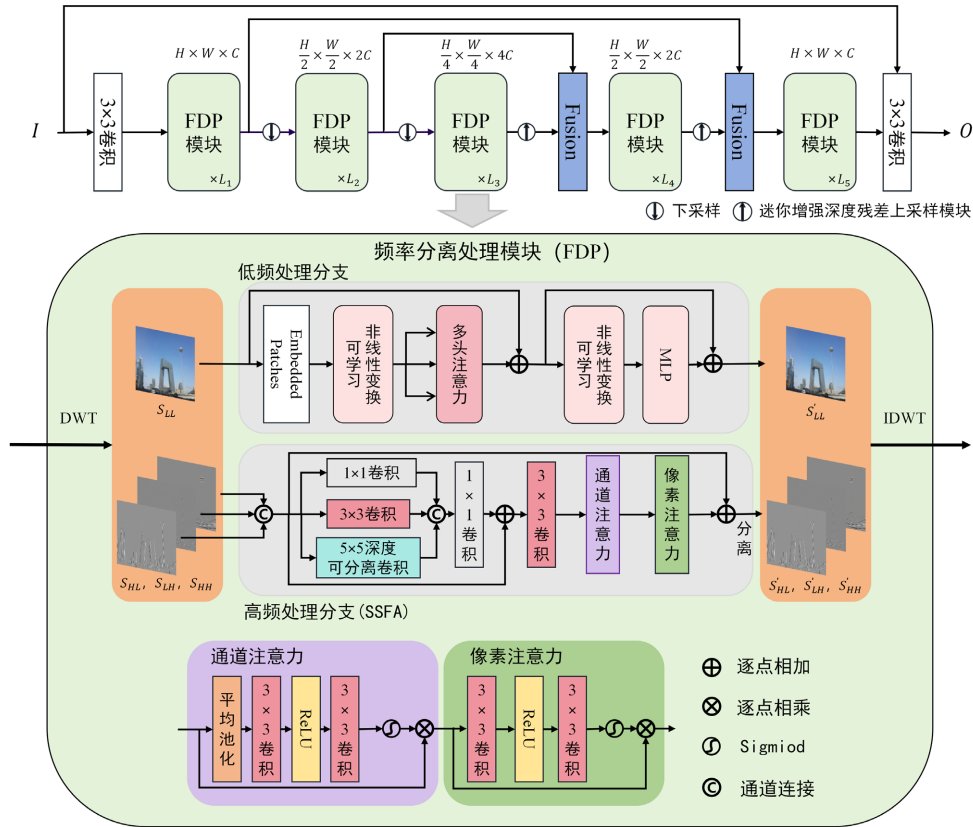


Figure 1. Design of the TripleD-Net network architecture
图 1. TripleD-Net 网络架构设计

频率分离处理模块(FDPB)的整体结构如图 1 所示。设 $F_1 \in \mathbb{R}^{C \times H \times W}$ 表示该模块的输入特征，则首先对 F_1 施加离散小波变换，以获得四个子带 $S_k, k \in \{LL, HL, LH, HH\}$ 。其中，低频子带 S_{LL} 作为低频分支的输入，而三个高频子带 S_{LH} 、 S_{HL} 和 S_{HH} 被拼接后送入高频分支进行处理。

由于通过 DWT 分解得到的低频与高频子带在统计分布上存在显著差异，若将其直接输入网络，可能导致特征分布不平衡与训练过程不稳定。为缓解这一问题，本文设计了一种轻量化的子带归一化 (Subband Normalization, SubNorm) 模块，用以增强不同频率子带之间的特征一致性。具体而言，首先将每个子带调整为均值为零、方差为一的分布形式，随后通过可学习的仿射变换进行进一步细化，其表达式为：

$$\tilde{S} = \gamma_k \cdot \frac{S_k - \mu_k}{\sigma_k + \epsilon} + \beta_k \quad (4)$$

其中， γ_k 和 β_k 为可训练参数， μ_k 与 σ_k 分别表示子带 S_k 的均值与标准差， ϵ 为防止数值不稳定而引入的微小常数，通常取 1×10^{-5} 或 1×10^{-6} 。该设计在保证训练过程稳定性的同时，能够保持不同频率子带之间的相对重要性。

2.1.1. 低频处理分支

低频处理分支的整体结构如图 1 所示, 该模块通过引入改进的 DehazeFormer 模块。考虑到多头注意力机制类似于低通滤波器的机制, 其难以充分利用特征表示中所包含的高频细节信息, 同时还会带来较大的计算开销, 因此将改进的 DehazeFormer 只应用于低频子带, 从而将计算成本降低至原来的 1/4。

为更深入理解 Transformer 中归一化层的作用机理, 本文对 LayerNorm 层的输入与输出数值进行了分析。如图 2 所示, 在浅层网络中, 输入与输出之间的关系近似呈线性变化, 而在深层网络中逐渐演化为类似 tanh 的非线性关系[16]。受此现象启发, 将原有的 RescaleNorm 模块替换为一种可学习非线性变换(Learnable Nonlinear Transform, LNT)模块, 以进一步加速 DehazeFormer 并降低其计算复杂度。LNT 模块对输入特征施加可学习的非线性映射, 其形式定义为:

$$\text{LNT}(x) = \tanh(\alpha \cdot x) \cdot \gamma + \beta \quad (5)$$

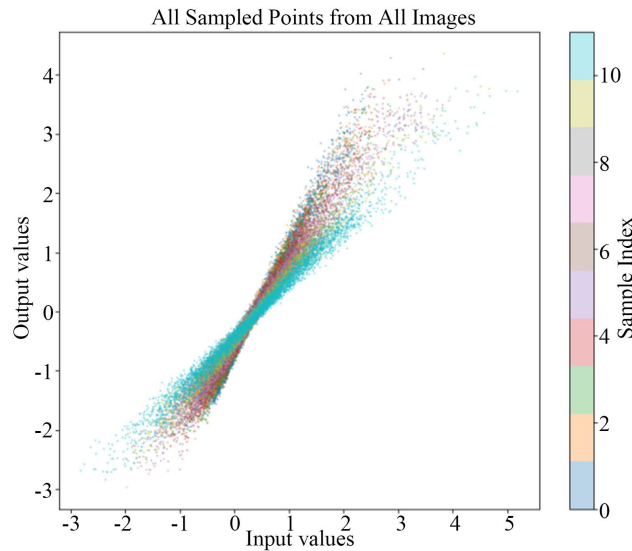


Figure 2. Comparison of input and output of the LayerNorm layer in the Transformer model
图 2. Transformer 模型中 LayerNorm 层的输入和输出比较

设 $X \in \mathbb{R}^{C \times H \times W}$ 表示输入特征图, 其中 $\alpha \in \mathbb{R}$ 为可训练的标量参数, 用于调节函数的激活幅度; $\gamma, \beta \in \mathbb{R}^{1 \times C \times 1}$ 分别为按通道学习的缩放因子与偏置项。低频处理模块中的其余结构与原始方法保持一致。因此, 该分支的整体计算过程可表示为:

$$\hat{x} = \text{MultiHead Attention}(\text{LNT}(S_{LL})) + S_{LL} \quad (6)$$

$$y = F_{MLP}(\text{LNT}(\hat{x})) + \hat{x} \quad (7)$$

其中, S_{LL} 表示通过离散小波变换获得的低频子带, 其空间尺寸仅为对应层输入特征图的 1/4。由于低频表示主要保留了输入特征中的整体结构与轮廓信息, 在该分支中引入改进后的 DehazeFormer 模块能够有效建模全局退化特征, 同时显著降低计算开销。

2.1.2. 高频处理分支

高频处理模块的结构如图 1 所示, 其核心由空间与尺度特征注意力模块(Spatial and Scale Feature Attention, SSFA)构成。该设计受到特征注意力块(Feature Attention Block, FAB) [17]的启发, FAB 通过融合通道注意力与像素注意力机制提升信息处理的灵活性, 从而引导网络关注高密度雾区域与显著通道特征。

然而，由于其固定的感受野大小，FAB 在处理细节与纹理丰富的自然图像时，往往难以充分提取多尺度特征。

为解决上述问题，本文提出了 SSFA 模块，将多尺度表示提取(Multi-scale Representation Extraction, MRE)与注意力机制相结合。MRE 模块由三条并行卷积分支组成：标准的 1×1 卷积、标准的 3×3 卷积以及 5×5 深度可分离卷积。通过并行卷积操作，网络能够在不同感受野尺度上建模特征信息。

设 $x \in \mathbb{R}^{C \times H \times W}$ 表示输入到 MRE 模块的特征图，则提取到的多尺度特征 F_a, F_b, F_c 可表示为：

$$F_a = \text{Conv}_{1 \times 1}(x), F_b = \text{Conv}_{3 \times 3}(x), F_c = \text{DSConv}_{5 \times 5}(x) \quad (8)$$

来自三条分支的特征首先在通道维度上进行拼接，并通过 1×1 卷积进行通道压缩。随后，引入一个 3×3 卷积层并结合残差连接，得到 MRE 模块的最终输出，其表达式为：

$$F_{MRE} = \text{Conv}_{3 \times 3}(x + \text{Conv}_{1 \times 1}(F_a, F_b, F_c)) \quad (9)$$

在后续的注意力模块中，同时引入通道注意力与像素注意力机制。通道注意力模块通过全局平均池化并依次经过 ReLU 与 Sigmoid 运算来获得通道权重。像素注意力模块则通过 ReLU 激活函数与 Sigmoid 函数，在每个通道的每个像素位置生成对应的特征权重。SSFA 模块的最终输出形式为：

$$F_{SSFA} = x + F_{PA}(F_{CA}(F_{MRE})) \quad (10)$$

经过 FDP 模块处理后，不同频率子带的特征通过逆小波变换进行重构，得到 FDP 的输出结果，并传递至后续模块进行处理。

2.2. 损失函数

本文采用 L1 损失作为优化目标，用于直接计算复原图像 \hat{I} 与其对应真实图像 I_{gr} 之间差异的绝对值，其定义如下：

$$L = \|\hat{I} - I_{gr}\|_1 \quad (11)$$

与 L2 损失相比，L1 损失对异常值具有更强的鲁棒性，并能够获得更为精细的图像重建效果，因此更适用于对边缘与纹理细节保持要求较高的图像去雾任务。实验结果表明，相较于感知损失或对抗损失，L1 损失在去雾任务中能够带来更加稳定的收敛过程以及更优的边缘保持性能。

2.3. 实现细节

低频处理分支中多头注意力模块的数量遵循 DehazeFormer-B [12] 的配置方案。在本文设计中，高频处理模块在第三阶段设置为 6 个，而在其余阶段仅设置为 1 个。该设计灵感来源于 AECR-Net 的实验结论：在低分辨率阶段仅使用 6 个 FAB 模块即可获得比在高分辨率阶段使用 57 个 FAB 模块的 FFA-Net 更高的精度[6]。

图 3(a)给出了 Mini Enhanced Deep Residual Upsample (MEDRU) 的整体结构示意图，图 3(b)对比了反卷积上采样(左)与 MEDRU 模块(右)在重建效果上的差异。

为提升去雾过程中图像重建质量，本文在模型重建阶段引入了 Mini Enhanced Deep Residual Upsample (MEDRU) 模块[18]作为一种可学习的上采样组件。MEDRU 结合深度残差学习框架与亚像素卷积，实现高质量的空间分辨率恢复。如图 3(a)所示，MEDRU 模块主要由三部分组成：用于浅层特征提取的初始卷积层、由 8 个残差块构成的深层特征精炼模块，以及基于像素重组的上采样模块以重建高分辨率输出。

每个残差块由两层卷积操作与 ReLU 非线性激活函数构成，并通过引入跳跃连接来优化梯度传播能力，同时促进特征的有效复用。

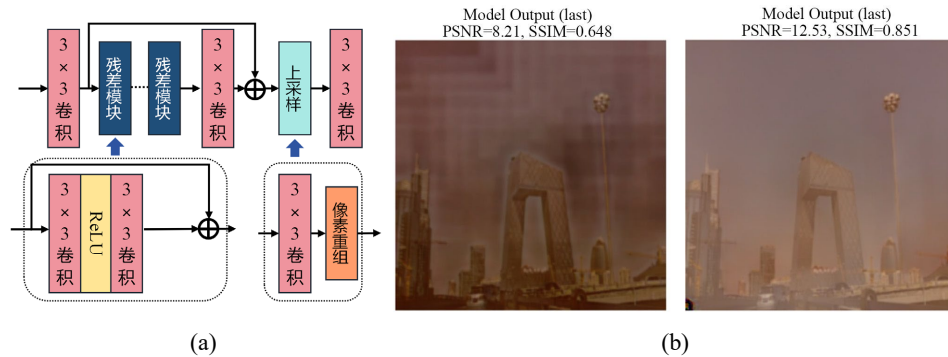


Figure 3. Architecture of MEDRU module and the comparison of the effects between deconvolutional upsampling (left) and the MEDRU module (right)

图 3. 迷你增强深度残差上采样模块结构图和反卷积上采样(左)与 MEDRU 模块(右)效果比较

3. 实验结果与分析

3.1. 数据集与评价指标

本文所有实验均基于 RESIDE 数据集[19]进行, 该数据集是图像去雾领域广泛使用的基准数据集。RESIDE 包含多个子集, 其中室内训练集(Indoor Training Set, ITS)、室外训练集(Outdoor Training Set, OTS)以及合成客观测试集(Synthetic Objective Testing Set, SOTS)被广泛采用。ITS 与 OTS 用于模型训练, SOTS 用于模型测试。SOTS 数据集进一步划分为 SOTS-indoor 与 SOTS-outdoor 两个子集, 分别用于评估在 ITS 与 OTS 上训练得到的模型性能。

ITS 数据集包含 1399 张干净的室内图像, 每张图像对应 10 张基于物理散射模型生成的合成雾图像。相比之下, OTS 数据集提供约 313,950 张合成雾图像用于训练, 覆盖了大量的室外场景。用于评估的 SOTS-indoor 和 SOTS-outdoor 各包含 50 张测试图像, 分别覆盖室内与室外环境。

在性能评估方面, 本文采用计算机视觉中常用的两种图像质量评价指标: 峰值信噪比(Peak Signal-to-Noise Ratio, PSNR)与结构相似度(Structural Similarity, SSIM) [20], 所有指标均基于原始大小的 RGB 图像计算, 未进行像素裁剪。

3.2. 实验设置

实验使用 PyTorch 1.12.2 在 NVIDIA Tesla A100 GPU 上进行。训练阶段中, 图像首先通过裁剪调整为 256×256 大小, 然后通过随机翻转与旋转进行数据增强。模型优化采用 Adam 优化器[21], 参数设置为 $\beta_1 = 0.9$ 、 $\beta_2 = 0.999$ 。初始学习率为 1×10^{-4} , 并按照余弦退火策略进行调整, 训练批次大小固定为 16。所提出的网络在 ITS 与 OTS 数据集上分别独立训练, 训练步数分别为 3×10^5 和 6×10^5 步。

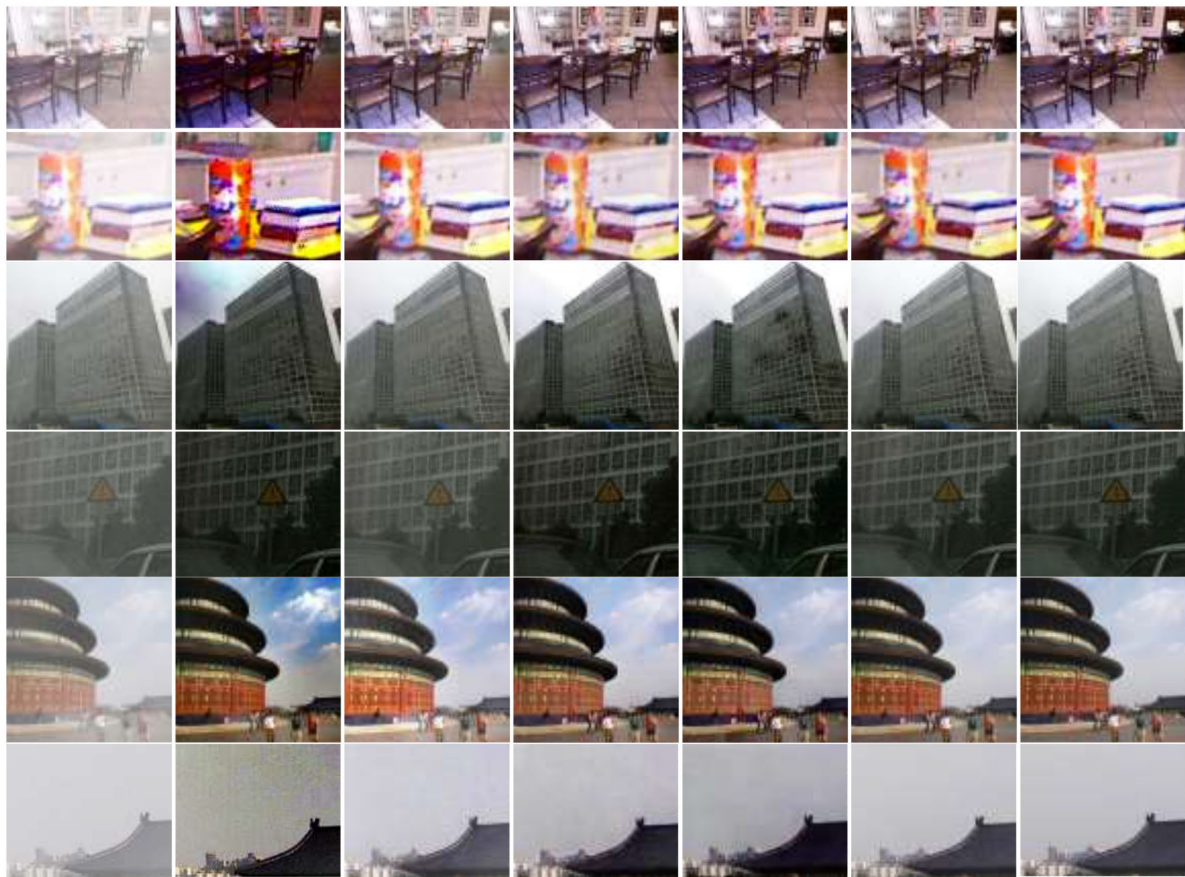
3.3. 实验结果

本文将所提出的 TripleD-Net 与六种具有代表性的去雾方法进行对比, 包括 DCP [2]、FFA-Net [17]、MSBDN [22]、AECR-Net [6]、Dehamer [23]以及 DehazeFormer [12]。定量比较结果汇总于表 1 中, 可见 TripleD-Net 在 SOTS-outdoor 数据集上表现更佳。实验结果表明, 所提出的方法在参数量更少、计算成本更低的条件下实现了较强的性能, 这进一步验证了该方法在图像质量提升任务中的有效性与高效性。

为了对去雾结果进行定性评估, 本文在 SOTS-indoor 与 SOTS-outdoor 测试集中选取具有代表性的图像进行可视化, 结果如图 4 所示。

Table 1. Results of comparative experiments
表 1. 对比实验结果

模型	SOTS-outdoor		SOTS-indoor		#Param.	MACs
	PSNR	SSIM	PSNR	SSIM		
(TPAMI') DCP	19.14	0.8605	16.61	0.8546	-	-
(AAAI') FFA-Net	33.57	0.9840	36.39	0.9886	4.456 M	287.8 G
(CVPR') MSBDN	34.81	0.9857	32.77	0.9812	31.35 M	<u>41.54 G</u>
(CVPR') AECR-Net	-	-	<u>37.17</u>	<u>0.9901</u>	2.611 M	52.20 G
(CVPR') Dehamer	<u>35.18</u>	<u>0.9860</u>	36.63	0.9881	132.4 M	48.93 G
(TIP') DehazeFormer-M	34.95	0.9840	37.84	0.994	4.63 M	48.64 G
(Ours) TripleD-Net	35.92	0.9892	35.04	0.9814	<u>4.33 M</u>	38.13 G



(a) Hazy Inputs (b) DCP [2] (c) FFA-Net [17] (d) MSBDN [22] (e) Dehamer [23] (f) Ours (g) GT

Figure 4. Visualization results of comparative experiments
图 4. 对比实验可视化结果

3.4. 消融实验

为了系统性地评估 TripleD-Net 中各关键模块的作用，本文进行了消融实验，通过在基准架构中逐步添加或移除模块来分析其作用。基线模型为简化版 TripleD-Net，仅保留基本的编码器 - 解码器结构，在

第一层和第二层使用普通卷积，在第三层使用 DehazeFormer 块，且未引入任何所提出的增强模块。分析重点关注以下四个模块：DWT、MFA、MEDRU，以及 LNT。各变体在 PSNR 和 SSIM 上的定量比较结果列于表 2，并得出如下观察结论：

DWT: 引入 DWT 模块使网络能够将输入特征分离为低频与高频分量。低频分量主要保留图像的结构信息及整体上下文信息，而高频分量包含更多噪声与细节信息。通过减少输入至 DehazeFormer 块的高频信息量，网络能够更专注于建模全局退化模式，从而实现训练效率的提升并改善整体去雾效果。

SSFA: 空间与尺度特征注意力模块(Spatial and Scale Feature Attention, SSFA)增强了网络在多尺度感受野下学习与提取特征的能力，同时关注视觉上重要的区域。当 SSFA 模块引入后，定量性能进一步提升，表明该模块的引入使得图像质量进一步改善、空间一致性进一步增强。

LNT: 引入 LNT 归一化模块为归一化层提供了自适应非线性变换能力。与使用静态归一化(如 LayerNorm)的基准相比，LNT 在 PSNR 与 SSIM 上均表现出一致提升，说明动态归一化更适合处理雾图像中严重退化的特征分布。

MEDRU: 用 MEDRU 模块替换标准上采样层可显著提升 PSNR，验证了基于超分辨率的上采样在重建过程中能够保持特征保真度，同时减轻传统插值或反卷积带来的信息损失。具体对比实验中，分别从训练好的 Base 模型和 Base + MEDRU 模型中选取一层上采样与一层下采样，其结果如图 3(b)所示。

综上，消融实验表明，所有模块均对最终性能具有实质性贡献，完整的 TripleD-Net 架构(整合四个模块)能够实现最佳性能。这充分证明了各模块在定量指标与视觉感知性能提升中的互补性与必要性。

Table 2. Result of ablation study
表 2. 消融实验结果

模型	Setting			PSNR (dB)
	Level 1	Level 2	Level 3	
Baseline	Conv	Conv	DehazeFormer	30.53
Base + DWT	Conv + DWT	Conv + DWT	DehazeFormer + DWT	31.32
Base + SSFA	Conv	Conv	DehazeFormer + SSFA + DWT	31.89
Base+ LNT	Conv	Conv	LNT-DehazeFormer	31.08
Base + MEDRU (apply MEDRU in Decoder)	Conv	Conv	DehazeFormer	32.22
Base + DWT+ SSFA	Conv + DWT	Conv + DWT	FDPB	32.67
Base + DWT + SSFA + MEDRU (apply MEDRU in Decoder)	Conv + DWT	Conv + DWT	FDPB	34.38
TripleD-Net	FDPB	FDPB	FDPB	35.92

4. 局限性分析

尽管本文提出的 TripleD-Net 在多个基准数据集上取得了具有竞争力的性能，并在颜色恢复、结构保持以及伪影抑制方面表现出良好的效果，但仍存在一定的局限性，有待进一步研究与改进。

例如从实验结果可以看出，所提出方法在室外场景(SOTS-outdoor)上的表现优于室内场景(SOTS-indoor)。其主要原因在于：室内数据通常具有更复杂的光照条件、更强的颜色偏移以及更短的景深分布，使得基于全局建模的低频分支难以准确捕捉真实退化模式。同时，室内场景中纹理重复性较高，而通过离散小波变换的高频分支信息量较少，无法满足区分真实细节与噪声方面面临的需求，从而在一定程度

上影响整体性能。未来可以设计新的频率分解方法，高频分支处理优化来进行室内去雾效果的改进。

此外，本文通过将 Transformer 结构限制在低频分支中有效降低了计算复杂度，但模型整体仍然包含多分支结构与频率变换操作，在资源受限设备上的部署仍存在一定挑战。因此，在保证性能的前提下进一步压缩模型规模、提升推理效率，也是一个具有实际意义的研究方向。

5. 结论

本文提出了一种基于频率感知建模的单幅图像去雾网络 TripleD-Net。该方法通过引入离散小波变换 (DWT) 对输入特征进行频率分解，将图像信息划分为低频结构信息与高频细节信息，并在统一框架中分别进行针对性建模。针对低频子带，本文引入了改进的 DehazeFormer 模块以增强全局上下文建模能力，并通过可学习非线性变换 (LNT) 替代传统归一化操作，从而在降低计算复杂度的同时提升特征表达能力。对于高频子带，本文设计了空间与尺度特征注意力模块 (SSFA)，通过多尺度卷积分支结合通道注意力与像素注意力机制，实现了对边缘结构与纹理细节的有效提取。此外，在重建阶段采用 MEDRU 上采样模块以进一步提升图像重建质量。在 RESIDE 数据集上的大量实验结果表明，所提出的 TripleD-Net 在 SOTS-outdoor 与 SOTS-indoor 测试集上均取得了具有竞争力的性能，并在参数规模与计算复杂度较低的情况下实现了更优的去雾效果。综上所述，TripleD-Net 通过频率分离建模与多尺度特征学习的协同设计，有效提升了去雾结果的颜色一致性、结构清晰度以及视觉质量。

基金项目

本研究得到了国家自然科学基金(No. 12472040)的资助。

参考文献

- [1] Cantor, A. (1978) Optics of the Atmosphere—Scattering by Molecules and Particles. *IEEE Journal of Quantum Electronics*, **14**, 698-699. <https://doi.org/10.1109/jqe.1978.1069864>
- [2] He, K., Sun, J. and Tang, X. (2011) Single Image Haze Removal Using Dark Channel Prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **33**, 2341-2353. <https://doi.org/10.1109/tpami.2010.168>
- [3] Zhu, Q., Mai, J. and Shao, L. (2015) A Fast Single Image Haze Removal Algorithm Using Color Attenuation Prior. *IEEE Transactions on Image Processing*, **24**, 3522-3533. <https://doi.org/10.1109/tip.2015.2446191>
- [4] Chen, D., He, M., Fan, Q., Liao, J., Zhang, L., Hou, D., et al. (2019) Gated Context Aggregation Network for Image Dehazing and Deraining. 2019 *IEEE Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, 7-11 January 2019, 1375-1383. <https://doi.org/10.1109/wacv.2019.00151>
- [5] Liu, X., Ma, Y., Shi, Z. and Chen, J. (2019) Griddehazenet: Attention-Based Multi-Scale Network for Image Dehazing. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, 27 October 2019-2 November 2019, 7314-7323. <https://doi.org/10.1109/iccv.2019.00741>
- [6] Wu, H., Qu, Y., Lin, S., Zhou, J., Qiao, R., Zhang, Z., et al. (2021) Contrastive Learning for Compact Single Image Dehazing. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 10551-10560. <https://doi.org/10.1109/cvpr46437.2021.01041>
- [7] 刘万军, 程裕茜, 曲海成. 基于生成对抗网络的图像自增强去雾算法[J]. 系统仿真学报, 2024, 36(5): 1093-1106.
- [8] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L. and Polosukhin, I. (2017) Attention Is All You Need. 31st *Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, 4-9 December 2017, 5998-6008.
- [9] 符程程, 魏为民, 杨同, 等. 结合特征增强注意力的混合卷积去雾网络[J]. 现代电子技术, 2026, 49(1): 27-33.
- [10] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T. and Houlsby, N. (2020) An Image Is Worth 16×16 Words: Transformers for Image Recognition at Scale. arXiv: 2010.11929.
- [11] 李玉洁, 马子航, 王艺甫, 等. 视觉 Transformer (ViT) 发展综述[J]. 计算机科学, 2025, 52(1): 194-209.
- [12] Song, Y., He, Z., Qian, H. and Du, X. (2023) Vision Transformers for Single Image Dehazing. *IEEE Transactions on Image Processing*, **32**, 1927-1941. <https://doi.org/10.1109/tip.2023.3256763>
- [13] Dong, W., Zhou, H., Wang, R., Liu, X., Zhai, G. and Chen, J. (2024) DehazeDCT: Towards Effective Non-Homogeneous

- Dehazing via Deformable Convolutional Transformer. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, 17-18 June 2024, 6405-6414. <https://doi.org/10.1109/cvprw63382.2024.00642>
- [14] Park, N. and Kim, S. (2022) How Do Vision Transformers Work? arXiv: 2202.06709.
- [15] Ahmed, N., Natarajan, T. and Rao, K.R. (1974) Discrete Cosine Transform. *IEEE Transactions on Computers*, **23**, 90-93. <https://doi.org/10.1109/t-c.1974.223784>
- [16] Zhu, J., Chen, X., He, K., LeCun, Y. and Liu, Z. (2025) Transformers without Normalization. 2025 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 10-17 June 2025, 14901-14911. <https://doi.org/10.1109/cvpr52734.2025.01388>
- [17] Qin, X., Wang, Z., Bai, Y., Xie, X. and Jia, H. (2020) FFA-Net: Feature Fusion Attention Network for Single Image Dehazing. *Proceedings of the AAAI Conference on Artificial Intelligence*, **34**, 11908-11915. <https://doi.org/10.1609/aaai.v34i07.6865>
- [18] Lim, B., Son, S., Kim, H., Nah, S. and Lee, K.M. (2017) Enhanced Deep Residual Networks for Single Image Super-Resolution. 2017 *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, 21-26 July 2017, 136-144. <https://doi.org/10.1109/cvprw.2017.151>
- [19] Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., *et al.* (2019) Benchmarking Single-Image Dehazing and Beyond. *IEEE Transactions on Image Processing*, **28**, 492-505. <https://doi.org/10.1109/tip.2018.2867951>
- [20] Wang, Z., Bovik, A.C. and Simoncelli, E.P. (2004) Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, **13**, 600-612. <https://doi.org/10.1109/tip.2003.819861>
- [21] Kingma, D.P. and Ba, J. (2014) Adam: A Method for Stochastic Optimization. arXiv: 1412.6980.
- [22] Dong, H., Pan, J., Xiang, L., Hu, Z., Zhang, X., Wang, F., *et al.* (2020) Multi-Scale Boosted Dehazing Network with Dense Feature Fusion. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 13-19 June 2020, 2157-2167. <https://doi.org/10.1109/cvpr42600.2020.00223>
- [23] Guo, C., Yan, Q., Anwar, S., Cong, R., Ren, W. and Li, C. (2022) Image Dehazing Transformer with Transmission-Aware 3D Position Embedding. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 5812-5820. <https://doi.org/10.1109/cvpr52688.2022.00572>