arXiv:2305.04430v1 [cs.CV] 8 May 2023

拨开迷雾：一种基于快速傅里叶卷积和 ConvNeXt 的先进非均匀去雾方法

周涵[1], 董伟[2],刘洋逸[1],陈军[11]加拿大汉密尔顿市麦克马斯特大学电气与计算机工程系[2]
加拿大埃德蒙顿市阿尔伯塔大学电气与计算机工程系zhouh115@mcmaster.ca,
wdong1745376@gmail.com, {liu5, chenjun}@mcmaster.ca

摘要

雾霾通常会导致图像质量下降，表现为对比度低、颜色偏移和结构失真。我们观察到，许多基于深度学习的模型在去除均匀雾霾方面表现出卓越的性能，但它们通常无法应对非均匀去雾这一挑战。这种情况主要由两个因素造成。首先，由于浓雾分布复杂且不均匀，难以高保真度地恢复结构和色彩特征，尤其是在雾霾严重的区域。其次，现有用于非均匀去雾的小规模数据集不足以支持基于卷积神经网络（*CNN*）的模型可靠地学习雾霾图像及其对应无雾图像之间的特征映射。为应对这两个挑战，我们提出了一种新颖的双分支网络，该网络结合了二维离散小波变换（*DWT*）、快速傅里叶卷积（*FFC*）残差块以及预训练的 *ConvNeXt* 模型。具体而言，在 *DWT-FFC* 频率分支中，我们的模型利用 *DWT* 捕获更多高频特征。此外，借助 *FFC*残差块所提供的大感受野，我们的模型能够有效地探索全局上下文信息，并生成具有更佳感知质量的图像。在先验知识分支中，我们采用了在 *ImageNet* 上预训练的 *ConvNeXt*，而非 *Res2Net*。这使得我们的模型能够学习更多补充信息，并获得更强的泛化能力。通过大量实验和消融研究，验证了所提出方法的可行性和有效性。代码可在以下地址获取 *https://github.com/zhouh115/DWT-FFC*。

1. 引言

作为一种自然现象，雾霾通常会大幅降低能见度，导致雾霾图像模糊不清且对比度低、颜色偏移和结构失真。目标检测等各种智能应用 [41] 以及自动驾驶 [45], 都需要在雾霾条件下正常运行，因此有必要从雾霾图像中恢复缺失的信息。近年来，图像去雾已成为计算机视觉领域的热门研究课题 [9, 19, 23, 27, 37, 38, 47–49, 53, 56]。

早期的图像去雾方法 [21, 26] 主要基于大气散射模型（ASM）开发 [35]，以建立雾霾图像与无雾图像之间的对应关系。该模型可用下式（1）表示：

$$I(x) = J(x)t(x) + A(1 - t(x)). \quad (1)$$

其中，I和 J分别表示雾霾图像及其清晰图像； x 表示像素位置； A 表示全局大气光； $t(x)$为透射图，由大气散射参数 β 和场景深度 $d(x)$决定，具体如下：

$$t(x) = e^{-\beta d(x)}. \quad (2)$$

显然，在假设 ASM 有效的情况下，图像去雾归结为估计 t（x）和 A [19,21,34]。遗憾的是，该模型仅适用于理想化的均匀雾霾。因此，基于 ASM 的方法无法处理非均匀去雾任务。

近年来，受其在分类、目标检测及其他视觉任务中取得的显著成功的启发，深度学习也被应用于单幅图像去雾领域 [16, 17, 54], 。原则上，通过端到端监督训练，基于深度学习的去雾方法不再受 ASM 框架的限制。事实上，研究表明，这些方法在一定程度上能够处理复杂且非均匀的雾霾图像 [5, 6]。

基于深度学习的去雾方法通常依赖大量训练数据。然而，在真实世界中，获取大量图像对非常困难，甚至是不可能的 [15]。训练数据不足已成为阻碍

非均匀去雾方法发展的因素之一。为应对训练数据有限的问题并缓解过拟合，近期一些方法 [5, 6, 15] 采用预训练模型，例如在 ImageNet 上预训练的 Res2Net [12, 16], 进行迁移学习 [36]。然而，这些方法并未利用最先进的模型，例如 Vision Transformer [14] 及其后续模型 [18, 52], Swin Transformer [30] 和 ConvNeXt[31, 46], ，这些模型在 ImageNet 上的表现优于 Res2Net。因此，利用更强大的预训练模型有望提升图像去雾效果。

此外，大多数现有方法无法恢复边缘和精细纹理等高频成分。DW-GAN [15]利用离散小波变换（DWT）在下采样阶段提取高频特征，并将其传递至上采样过程。然而，该方法仍难以处理严重雾霾区域。

因此，要实现高质量图像去雾，需要解决两个问题。1）包括 NTIRE [1, 2], 发布的数据集在内，大多数用于非均匀雾霾去雾的数据集规模较小，不足以让基于 CNN 的模型学习雾霾图像与其对应无雾图像之间的映射。尽管采用预训练模型可以缓解训练数据有限所导致的过拟合问题，但仍应谨慎设计网络，以最大限度地提升其获取先验知识的能力。2）复杂且非均匀的雾霾模式给图像复原带来了巨大挑战，尤其是在浓雾区域 [6, 15]。

考虑到上述两个问题，我们提出了一种双分支生成对抗网络，其中每个分支分别针对一个问题进行设计。第一个分支旨在学习从雾霾图像到清晰图像的颜色和结构映射。它由三个离散小波变换块和三个 FFC [11, 43] 残差块组成。离散小波变换块用于获取高频知识和结构细 [15, 22, 42], 节，而 FFC 残差块通过同时利用频谱信息和空间信息 [32, 43], 提供能够覆盖整幅图像的宽广混合感受野。得益于更大的感受野，FFC 残差块使编码器能够充分利用全局上下文，从而提高恢复图像的感知质量，这对于高分辨率非均匀去雾尤为重要。第二个分支旨在实现迁移学习 [36, 44, 58]。具体而言，鉴于预训练 ConvNeXt 出色的分类性能，我们采用其前三个层来构建编码器。通常，预训练模型能够成功适应不同于其原始训练任务的其他任务 [6]。与 Res2Net 相比，ConvNeXt 在融合视觉 Transformer 的若干关键组件并保留卷积优势的同时，在 ImageNet 上表现出更优异的性能。

卷积网络 [14, 31]。对于第二个分支的解码器，采用像素重排层和通道/像素级注意力模块 [25]，逐步将图像恢复至其初始分辨率。随后，利用融合操作 [6] 聚合两个分支的输出。最后，引入由对抗损失 [15]引导的判别器，以确保最终重建结果的感知质量。

我们工作的主要贡献如下：1）我们引入 FFC残差块，并将其与二维离散小波变换相结合，以应对非均匀去雾问题。2）我们采用预训练的 ConvNeXt 模型，利用先验知识解决小规模数据集问题，并验证了该方法的有效性。3）我们开展了广泛的实验和消融研究，以论证整体设计的合理性，并展示其具有竞争力的性能。

2. 相关工作

单幅图像去雾。 单幅图像去雾是计算机视觉和图像处理中的一项具有挑战性的任务，因为它需要从单幅输入图像中去除不需要的大气雾霾。多年来，人们提出了多种有效的方法来解决这一问题，这些方法可分为两大类： *i.e.,* 基于物理的方法和基于 CNN 的解决方案。基于物理的方法主要依赖 ASM [33, 35] 和人工设计的先验知识，例如暗通道先验 [19], 颜色衰减先验 [57], 非局部先验 [8]。然而，由于其底层假设的适用范围有限，基于物理的方法通常不够稳健。

随着深度学习的快速发展，过去几年见证了其在单幅图像去雾中的广泛应用。早期基于深度学习的方法仍然采用 ASM。例如，DehazeNet [9] 设计了 CNN 模型来估计介质透射图，然后通过 ASM 利用该图获得去雾图像。随后，AOD-Net [23] 同时估计大气光和透射图，以生成恢复图像。近年来，研究人员提出了各种深度学习模型，可将雾霾图像直接映射为对应的清晰图像，而无需借助 ASM。例如，GCANet [10] 引入了门控上下文聚合网络，用于去除网格伪影并融合不同层级的特征表示。Qin 等人 [37] 提出了 FFA-net，通过通道注意力和像素注意力机制组，自适应地处理不同的特征和像素区域，从而增强灵活性。MSBDN [13]采用多尺度增强解码器，逐步恢复去雾图像。大多数以往工作都假设雾霾呈均匀分布，但这一假设通常无法代表真实世界场景，并且在存在浓雾或非均匀雾霾的场景中可能导致性能显著下降。为处理真实世界中的雾霾图像，三叉戟去雾网络（TDN）[5], , 其

由细节优化子网络、编码器-解码器子网络和雾霾密度图生成子网络组成，并在 NTIRE 2020 非均匀去雾挑战赛中被提出。在 NTIRE 2021 非均匀去雾挑战赛中 [6], TDN 被 DW-GAN 超越， [15] 后者采用离散小波变换提取低频和高频特征，这些特征有利于去雾任务。此外， [48] 提出了一种基于对比学习的新型去雾方法，利用了正、负图像信息。除 DW-GAN 外，上述所有方法都是以空间域为中心的去雾方法，未能直接利用雾霾退化在频域中的特性 [49]。

频域学习。 近年来，为了获得有效的频谱特征，频域学习日益受到研究者的关注。具体而言， [55] 利用基于小波的表示来促进高分辨率图像复原。 [51] 通过离散小波变换将图像分解为低频带和高频带，并提取各频带的特征。此外，Chi 等人 [11] 提出将 FFC 作为一种非局部操作单元，通过学习空间和频谱特征来同时扩大感受野。随后， [43]将 FFC 用作主要卷积操作，在大掩码图像修复任务中取得了出色的性能，且所提出的方法展现出很强的泛化能力。 [49] 揭示了雾霾退化与频率属性之间的联系，并设计了一种双分支网络，从空间和频谱两个角度引导图像去雾过程。

上述方法表明，频域信息能够显著提升图像复原方法的性能。因此，我们借鉴基于小波的分解思路，并在深度学习网络架构中引入非局部 FFC 操作单元。这样一来，我们的模型便能有效利用频谱信息，同时保留有意义的纹理细节，并确保最终复原的高分辨率图像具有一致的结构。我们将通过定量评估验证所提模型相较于现有方法的优越性。

迁移学习。 迁移学习旨在利用从相关但不同任务中获取的知识，增强目标模型在特定任务上的能力，从而降低对目标域中大量数据的需求 [36, 44, 58]。一些现有方法利用通过在 ImageNet 上预训练获得的大量先验知识，来辅助图像复原任务。例如，2020 年 NTIRE 去雾挑战赛的两位冠军都将预训练的 Res2Net 模型用作知识迁移的基础模块，成功缓解了小尺度训练数据集导致的过拟合问题。尽管其能力令人印象深刻，Res2Net 在图像去雾方面的效果有限，尤其是在处理具有非均匀雾霾的高分辨率图像时。我们旨在通过实验分析证实这一点。此外，值得注意的是，即使在图像分类领域，Res2Net 等卷积神经网络的性能也已被视觉变换器 [14]所超越。Liu 等人 [31] 通过引入多个促成变换器卓越性能的关键组件，模仿视觉变换器对标准 ResNet [20] 进行了重新设计。由此产生的架构被命名为 ConvNeXt，其性能超越了 Swin 变换器 [30]，并对视觉变换器比卷积神经网络更加准确和高效这一广泛认同的观点提出了挑战。然而，据我们所知，基于 ConvNeXt 的图像复原仍是一个尚未得到充分探索的领域。

3. 提出的方法

本节首先介绍我们提出的网络的详细结构（如图 1所示）。该网络基于离散小波变换、快速傅里叶卷积残差块和预训练 ConvNeXt 参数构建。随后，我们详细介绍离散小波变换和快速傅里叶卷积残差块，并分别分析它们对整个网络的意义。最后，我们讨论用于训练模型的损失函数。

3.1. 网络框架

许多双分支方法在 NTIRE 2020 和 2021 非均匀去雾挑战赛中取得了巨大成功 [5, 6]。观察到这一点后，我们设计了一个双分支神经网络（如图1所示）。

DWT-FFC分支。 受 [6], 启发，我们构建了 DWT-FFC 频率分支，将其设计为一个编码器-解码器网络，用于学习雾霾图像与清晰图像之间的特征映射，并在每个特征尺度上采用大量跳跃连接。除传统卷积技术外，我们还采用离散小波变换来实现特征提取。借助离散小波变换，我们的模型能够检测高频和低频特征（第3.2 节提供了详细说明）。如图2所示，低频表示与常规卷积输出拼接，而高频特征则传递至上采样模块，逐步恢复无雾图像。为了使重建图像更加逼真且符合感知效果，如图3所示，我们提出采用快速傅里叶卷积残差块（FFC，详情见第 3.3节），利用空间和频谱信息进行图像去雾。

然而，仅采用 DWT-FFC 频率分支的模型由于缺乏大量训练数据，无法在非均匀雾霾挑战中取得令人信服的性能。因此，我们进一步引入第二个分支，该分支使用在大型数据集上预训练的强大网络，以获取先验知识。

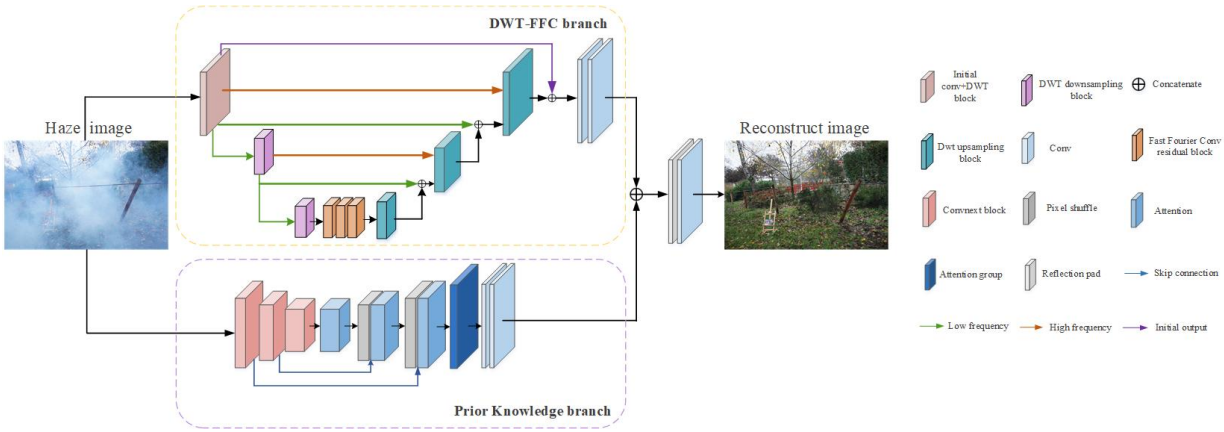

图1. 我们提出的方法的网络结构。

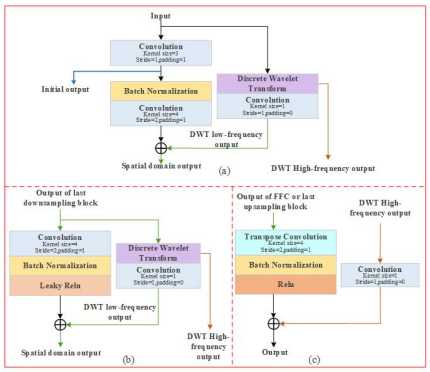

图2. DWT-FFC分支中常规卷积与DWT之间连接方式的示意图。绿色线条表示DWT低频特征，棕色线条表示DWT高频特征。

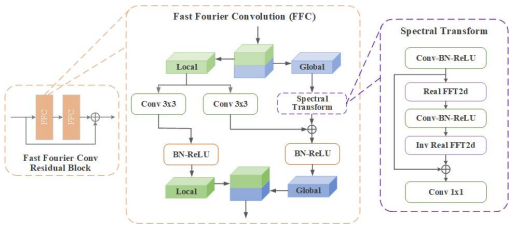

图3. FFC残差块的网络结构。一个FFC残差块包含两个FFC单元，每个单元包含一个频谱变换。

先验知识分支。 我们的第二个分支——先验知识分支，旨在将大型数据集图像分类任务中的额外信息引入当前的非均匀去雾挑战中。具体而言，我们将该分支设计为类似于 U-Net 的结构 [39], 并采用经过 ImageNet 预训练的 ConvNeXt 作为编码器的骨干网络。我们仅采用 ConvNeXt 的前三个阶段。

这样，使用预训练权重的模型可以比使用随机初始参数的模型表现更好。对于解码器，受 [37], 启发，我们采用若干上采样层，每个上采样层包含一个像素重排块和一个注意力模块。在此，我们引入像素重排块，以降低计算负担，并使特征图尺寸逐步恢复至原始分辨率；注意力块则使模型能够识别动态雾霾模式。此外，如图3所示，我们采用跳跃连接技术来增强泛化能力。最后，通过一种简单有效的融合方式组合各分支的输出，生成最终恢复的清晰图像 [50]。

3.2. 离散小波变换

二维离散小波变换包含一个低通滤波器（f_{LL}）和三个高通滤波器（f_{LH}, f_{HL}, f_{HH}），这些滤波器等效于参数固定且步幅为 2 的卷积运算。例如，在我们的方法中采用的 Haar 离散小波变换里，$f_{LL} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, $f_{LH} = \begin{pmatrix} -1 & -1 \\ 1 & 1 \end{pmatrix}$, $f_{HL} = \begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix}$, $f_{HH} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$。与每个滤波器进行卷积后，输入可分解为四个子分量： $i.e., x_{LL}$, x_{LH}, x_{HL}和 x_{HH}。具体而言， x_{LL} 可表示为式（3），其中 i 和 j表示像素索引，另外三个子带的表达式与 x_{LL}类似。

$$x_{LL} = x(2i-1, 2j-1) + x(2i-1, 2j) + x(2i, 2j-1) + x(2i, 2j) \tag{3}$$

此外，我们将 x_{LL} 与各个特征尺度上的常规卷积输出相结合，使所提出的模型能够学习空间信息和频率信息。

3.3. 快速傅里叶卷积

快速傅里叶卷积基于逐通道快速傅里叶变换，其感受野覆盖整幅图像或整个特征图。FFC 将输入拆分为两个并行分支： *i) local branch*并行执行两个常规卷积， *ii) global branch*则同时利用一个常规卷积和一个谱变换来捕获全局上下文。具体而言，谱变换包含三个步骤：

a) 对输入张量应用 *Real FFT2d*，并将实部和虚部拼接：

Real FFT2d: $\mathbb{R}^{H\times W\times C} \to \mathbb{C}^{H\times \frac{W}{2}\times C}$,
ComplexToReal: $\mathbb{C}^{H\times \frac{W}{2}\times C} \to \mathbb{R}^{H\times \frac{W}{2}\times 2C}$;

b) 在频域中应用 ReLU 激活、批归一化和卷积运算：

ReLU $\cup$ *BN* $\cup$ 卷积*1×1*: $\mathbb{R}^{H\times \frac{W}{2}\times 2C} \to \mathbb{R}^{H\times \frac{W}{2}\times 2C}$;

c) 应用逆 FFT 变换以恢复空间结构：

实数转复数:$\mathbb{R}^{H\times \frac{W}{2}\times 2C} \to \mathbb{C}^{H\times \frac{W}{2}\times C}$,
逆实数 *FFT2d*: $\mathbb{C}^{H\times \frac{W}{2}\times C} \to \mathbb{R}^{H\times W\times C}$。

如图 3所示，FFC 利用融合操作将局部分支和全局分支的输出结合起来。最后，通过将输入与输出相连，两个 FFC 单元构成一个 FFC 残差块。在我们提出的方法中，我们总共采用了三个 FFC 残差块，使网络能够从整体图像或特征图的角度去除雾霾。

3.4. 损失函数

我们将真实图像定义为 I^{gt}，并分别用 I^{hazy} 和 $\tilde{I}$ 表示雾霾图像和去雾图像。我们用 G 和 D 表示所提出的方法和判别器。

平滑 L1 损失。平滑 L1 损失可通过公式 (4) 和公式 (5) 计算，其中 N 表示像素总数， $I_i^{gt}(x)$ 和 $\tilde{I}_i(x)$ 分别表示真实图像和去雾图像第 i 个通道中像素 x 的强度。

$$\mathcal{L}_{\text{smooth}-\text{L1}} = \frac{1}{N}\sum_{x=1}^{N}\sum_{i=1}^{3} f(I_i^{gt}(x) - \tilde{I}_i(x)) \tag{4}$$

其中

$$f(\gamma) = \begin{cases} 0.5\gamma^2 & \text{if } |\gamma| < 1 \\ |\gamma| - 0.5 & \text{otherwise} \end{cases} \tag{5}$$

感知损失。我们引入在 ImageNet 上预16 [40]训练的 VGG 作为损失网络 ϕ。感知损失可定义为式（6）：

$$\mathcal{L}_{\text{PL}} = \sum_{j=1}^{3} \frac{1}{C_j H_j W_j} \left\| \phi_j(G(I_i^{hazy})) - \phi_j(I_i^{gt}) \right\|_2^2 \tag{6}$$

其中， ϕ_j 表示骨干网络中第 j 层的激活值， C_j、 W_j 和 H_j 分别表示对应特征图的通道数、宽度和高度。

消融研究可以证明，感知损失对我们提出的方法至关重要。

多尺度结构相似性损失。我们在损失函数中引入多尺度结构相似性（MS-SSIM） [15] 。首先，根据式 i 计算像素的结构相似性 7 :

$$\begin{aligned} SSIM(i) &= \frac{2\mu_D\mu_C + T_1}{\mu_D^2 + \mu_C^2 + T_1} \cdot \frac{2\sigma_{DC} + T_2}{\sigma_D^2 + \sigma_C^2 + T_2} \\ &= l(i) \cdot s(i) \end{aligned} \tag{7}$$

其中 T_1 和 T_2 表示两个较小的常数， D 和 C 分别表示重建图像和清晰图像中以当前像素为中心的两个固定大小窗口。应用高斯滤波器后，我们可以计算均值 μ_D、 μ_C标准差 σ_D、 σ_C 以及协方差 σ_{DC}。多尺度结构相似性损失如式（8）所示，其中 S 表示尺度总数， α 和 β 为默认参数。

$$\mathcal{L}_{\text{MS}-\text{SSIM}} = 1 - \prod_{s=1}^{S} (l^{\alpha}(i) \cdot cs_s^{\beta_s}(i)) \tag{8}$$

对抗损失。L对抗损失按式（9）计算，其中 $D(G(I^{hazy}))$ 表示判别器将恢复图像判定为真实图像的概率 [17]。

$$\mathcal{L}_{\text{adv}} = 1 - \sum_{s=1}^{N} -logD(G(I^{hazy})) \tag{9}$$

总损失。用于监督我们所提出方法训练的总损失如公式（10）所示：

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{smooth}-\text{L1}} + \alpha\mathcal{L}_{\text{MS}-\text{SSIM}} + \beta\mathcal{L}_{\text{PL}} + \gamma\mathcal{L}_{\text{adv}} \tag{10}$$

其中， $\alpha = 0.2$、 $\beta = 0.01$、 $\gamma = 0.0005$ 分别是各个损失函数的超参数。

4. 实验

在本节中，我们首先介绍数据集。随后，讨论实验设置和评估标准。此外，我们还对所提出的网络进行了消融研究。与此同时，我们从定量和定性两个方面，将我们的方法与当前最先进的图像去雾方法进行比较。最后，我们介绍了我们在 NTIRE 2023 高分辨率非均匀去雾挑战赛中的图像去雾结果。

4.1. 数据集

我们使用 RESIDE（室内训练集，ITS） [24], NH-HAZE [2], NH-HAZE2 [6],和 HD-NH-HAZE [7] 进行实验。ITS 包含一万多个训练样本，其合成客观测试集（SOTS）可用于测试。NH-HAZE 共包含 55 张图像。我们采用官方测试数据进行模型评估，其余数据则用于模型训练。NH-HAZE2 仅包含25 个训练数据、5 个验证数据和5 个测试数据。鉴于验证数据和测试数据均未公开，我们从训练数据中提取前20 张图像用于训练，

方法	L_1	L_{PL}	L_{ssim}	L_D	PSNR	SSIM
(1) 单个DWT分支	✓	✓	✓	✓	18.35	0.6491
(2) 单个FFC分支	✓	✓	✓	✓	19.03	0.6834
(3) DWT-FFC分支	✓	✓	✓	✓	19.38	0.6991
(4) 先验知识分支	✓	✓	✓	✓	20.35	0.7161
(5) DWT-FFC分支+Res2Net	✓	✓	✓	✓	21.67	0.7301
(6) 我们的方法	✓	✓	✓	✓	**22.20**	**0.7458**
(7) 我们的方法	✓	✓	✓		22.07	0.7427
(8) 我们的方法	✓	✓			21.81	0.7344
(9) 我们的方法	✓				21.71	0.7299

表 1. 消融研究结果。第一行用于研究网络各组件的必要性；第二行用于说明训练所用损失函数的合理性。**粗体** 数字表示最佳结果。

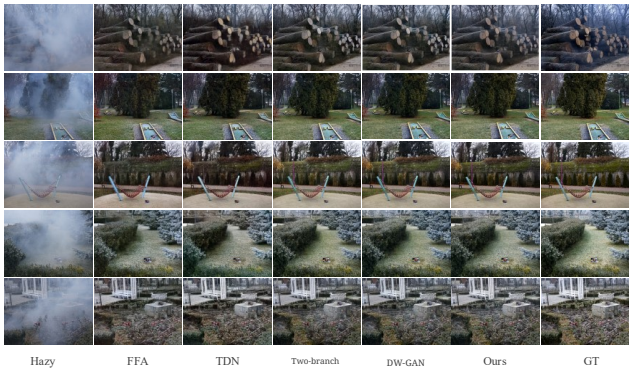

图 4. NH-HAZE 上的对比结果。

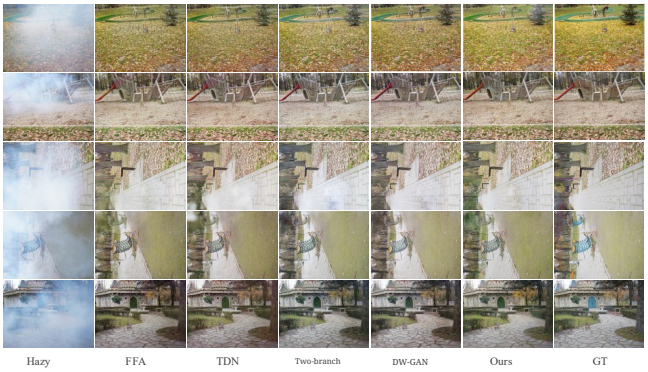

图 5. NH-HAZE2 上的对比结果。

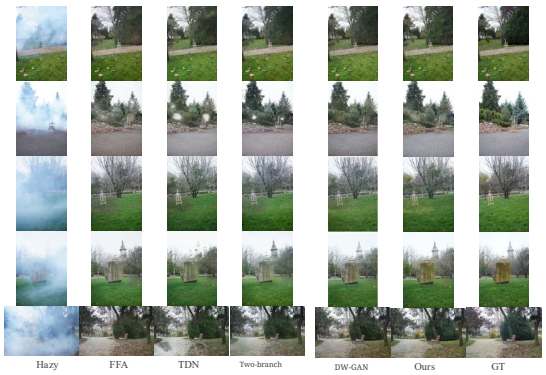

图 6. HD-NH-HAZE 上的对比结果。我们的结果几乎接近真实值，且高度逼真，几乎没有明显的不协调感。

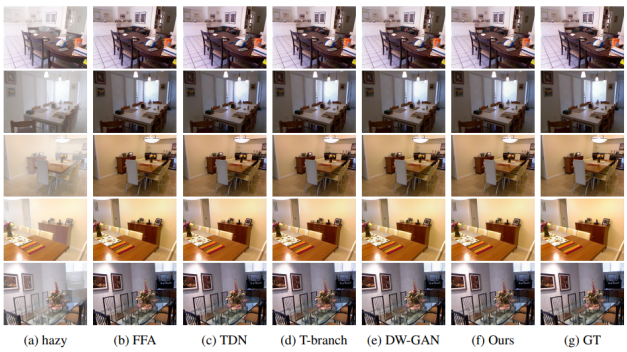

图 7. ITS 上的对比结果。所有方法生成的去雾图像在视觉上都令人愉悦。

并将剩余的 5 张图像用作测试样本。HD-NH-HAZE数据集以雾霾浓重且分布不均匀的场景为特点。该数据集包含 40 个训练数据、5 个验证数据和 5 个测试数据，每张图像的分辨率为 6000×4000。鉴于验证集和测试集目前尚未公开，我们使用第 1 至 35 张图像作为训练数据，而第 36 至 40 张图像则保留用于评估。此外，我们利用伽马校正构建了一个组合数据集，其中包含120 对基于 NH-HAZE、NH-HAZE2和 HD-NH-HAZE 数据集的图像（更多详情见第4.5节）。我们使用 HD-NH-HAZE 中相同的5 张图像进行测试，并使用其余115 张图像训练模型。

方法	ITS		NTIRE20		NTIRE21		HD-NH-HAZE		组合数据集	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
FFA	36.55	0.9888	19.50	0.6441	20.56	0.8106	20.23	0.7103	20.43	0.7112
TDN	34.87	0.9804	20.73	0.6727	20.44	0.8014	20.06	0.7132	20.30	0.7140
Two-branch	37.76	0.9905	21.55	0.7149	21.75	0.8277	21.15	0.7323	21.39	0.7327
DW-GAN	36.33	0.9866	21.69	0.7161	22.13	0.8314	21.52	0.7325	21.68	0.7371
ours	**37.86**	**0.9907**	**22.64**	**0.7298**	**22.82**	**0.8738**	**22.20**	**0.7458**	**22.26**	**0.7469**

表 2. ITS、NH-HAZE、NH-HAZE2、HD-NH-HAZE 和组合数据集上的定量比较。**粗体** 数字表示最佳结果，双下划线标注的结果表示次佳结果。

4.2. 实验设置

为丰富模型的训练数据，我们采用随机裁剪，将图像块裁剪为 384×384大小，并随机旋转 90、180 或 270 度，同时进行垂直翻转和水平翻转。优化过程使用 Adam 优化器，其默认参数为 β_1和 β_2 （分别为 0.9 和 0.999 ）。我们的训练方法采用专门的衰减策略：初始学习率为 e^{-4}，在训练至 3000、5000 和 8000 个 epoch 时，学习率分别衰减至原来的 0.5 倍，最终训练在 10000 个 epoch 后结束。判别器采用相同的优化器和训练策略。实验使用一块 RTX 2080 Ti GPU。为定量评估模型的性能，我们采用两种常用指标：峰值信噪比（PSNR）和结构相似性（SSIM）。由于 GPU 资源有限，训练好的模型无法处理整张 HD-NH-HAZE 图像，因此我们采用基于分块的测试策略来解决这一问题：1）首先将图像划分为 9 个图块，每个图块的大小为 1600×2432；2）然后使用训练好的权重输出每个图块的去雾结果；3）最后，通过对重叠区域取平均，将这些输出合并为与雾霾输入图像大小相同的最终结果。

4.3. 消融研究

我们进行了全面的消融研究，旨在验证所提出方法中每个组件的必要性。遵循消融研究的原则，我们设计并构建了六种不同的网络，每种网络都采用独特的模块组合。这些网络用于展示各个模块的相对重要性，并强调各组成部分对整体框架的重要作用。(1) 单个DWT分支：仅使用三次DWT下采样和三次DWT上采样，不包含FFC。(2) 单个FFC分支：仅使用FFC模块，不包含DWT下采样块或高频跳跃连接。(3) DWT-FFC分支：将DWT下采样和上采样与FFC模块组合为单个分支。(4) 先验知识分支：仅使用迁移学习分支来复原去雾图像。(5) DWT-FFC 频率分支和Res2Net：该网络与最终方法的唯一区别在于迁移学习分支的编码器。(6) 我们的方法：由DWT-FFC 频率分支和基于ConvNext的先验知识分支组成的双分支结构。

通过分别比较表1 中的 (3) 与 (1)、(2)，我们可以得出结论：DWT 和 FFC 均有助于提升图像去雾性能，具体表现为峰值信噪比和结构相似性的提高，而且 FFC 的作用比 DWT 更为显著。这一观察结果与我们对 DWT（第 3.2 节）和 FFC（第 3.3 节）的讨论一致。此外，我们可以观察到，所提出的双分支网络的性能远优于任何单分支网络（分别比较 (6) 与 (3)、(4)）。这一发现验证了我们设计双分支网络的初衷。最后，我们将模型与另一种引入预训练 Res2Net 权重的双分支网络进行比较。表1中 (6) 与 (5) 的峰值信噪比和结构相似性提升，证明了预训练 ConvNeXt 权重的有效性。此外，为验证训练中使用的总损失函数的合理性，我们采用第 3.4 节介绍的几种损失函数组合训练模型。如表1中的 (6)–(9) 所示，使用公式（10）定义的损失函数时，模型性能最佳；舍弃任何一项损失都会降低模型的峰值信噪比和结构相似性。

4.4. 与最先进模型的比较

本节将从定量和定性两方面，将我们提出的方法与最先进的模型进行比较。这些方法包括 NTIRE 2020 非均匀去雾挑战赛的获胜方案（TDN [27]）、NTIRE 2021 非均匀去雾挑战赛的获胜方案（DW-GAN），以及一种通过集成学习实现的双分支图像去雾方法（本文简称 Two-branch），还有 FFA [37]。

定量比较结果。比较结果如表2所示。我们的方法性能



图8. 我们的方法在 NTIRE 2023 HR 非均匀去雾挑战赛上的测试结果。

在所有数据集上，我们的方法均显著优于其他模型。特别是，在这四个非均匀数据集上，我们的模型在峰值信噪比方面平均比排名第二的方法高出 0.725 dB。另一方面，在 ITS 上，与其他模型相比，我们的方法优势并不明显（峰值信噪比仅比 Two-branch 高出 0.09 dB）。这一现象的原因在于训练数据的数量以及雾霾模式的复杂程度。ITS 是一个大规模合成数据集。考虑到其可用训练数据量超过其他四个数据集的 100 倍，预训练 ConvNeXt 权重对整个网络的贡献已相应减弱。此外，ITS 中的雾霾不像其他数据集中的雾霾那样复杂，因此，旨在解决非均匀雾霾和浓雾问题的离散小波变换和 FFC残差块的优势无法得到充分发挥。

定性比较结果。 本节从感知和视觉角度，将我们模型的性能与其他方法进行比较。比较结果如图所示 7、图 4、图 5 和图 6。对于 ITS，所有方法都能生成视觉效果令人满意且与真实图像非常接近的去雾图像。大量训练数据和相对简单的雾霾模式，使这些先进模型能够学习从雾霾图像到清晰图像的特征和颜色映射。对于非均匀数据集，采用不同方法会产生明显的视觉差异。具体而言，在 NH-HAZE 上，FFA 和 TDN 生成的模糊结果表明，这两种方法在非均匀去雾任务中的能力有限；而 Two-branch 和 DW-GAN 方法则存在一定程度的颜色和结构偏移。在 NH-HAZE2上，由于结果中仍有雾霾，TDN 的表现相当差；FFA、Two branch 和 DW-GAN 方法能够生成与真实图像高度接近的结果，但某些区域仍存在令人不悦的颜色偏移，例如第一张测试图像中的草地和最后一张图像中的道路。对于 HD-NH-HAZE数据集，FFA、TDN 和 Two-branch 方法处理的图像存在一些明显缺陷：出现了毫无意义且极不协调的白色空洞。DW-GAN 展现了其基于极度雾霾图像进行图像恢复的能力，但结果中包含若干局部模糊区域，并且因基于分块的测试策略而出现明显的碎片化现象。总体而言，在所有数据集上，尽管我们的方法可能会产生伪影，或无法去除某些小区域的雾霾，例如 NH-HAZE2 第一张测试图像中的树木以及 HD-NH-HAZE 最后一张图像中的树木，但图像绝大多数区域的颜色和结构细节都与真实图像非常接近，结果也十分逼真，几乎没有明显的不协调感。

4.5. NTIRE2023高分辨率非均匀挑战赛

基于 NH-HAZE、NH-HAZE2和 HD-NH-HAZE 构建的组合数据集。该挑战赛仅提供 40 个训练样本，训练数据并不充足。为了增加训练数据的数量并提升泛化能力，我们分别对 NH-HAZE 和 NH-HAZE2应用逐通道伽马校正。具体而言，在这两个数据集的每个通道中，我们使用各不相同的伽马参数进行伽马校正，使该通道的平均强度更接近 HD-NH-HAZE。基于此，我们假设这两个数据集与 HD-NH-HAZE 具有相似的雾霾模式，并且在组合数据集上训练的网络可能比仅在 HD-NH-HAZE 上训练的网络表现更好。表2 中的结果表明了这一趋势，并验证了这种预处理方法的有效性。

在 NTIRE 2023 高分辨率非均匀图像去雾挑战赛中的表现。根据报告 [7],，我们的模型在该挑战赛中夺冠，并且在峰值信噪比（22.87）和结构相似性（0.71）方面属于感知质量最高的方案之一。我们的模型测试结果如图所示。8，这表明我们的模型具备先进的去雾能力，能够有效去除雾霾，生成视觉效果出色且结构一致的图像。

5. 结论

本文提出了一种用于高分辨率图像去雾的新型双分支网络。DWT-FFC 频率分支利用离散小波变换提取复杂的高频特征，同时采用 FFC 扩大感受野，从而提升输出的感知质量。先验知识分支用于提供补充信息，以解决小尺度数据集上的过拟合问题，进而增强模型的泛化能力。大量实证评估有力地表明，我们的模型在合成数据集和真实场景中均表现出色。此外，我们的模型在保真度和感知质量方面均优于最新的最先进技术。

参考文献

[1] C.O. Ancuti、C. Ancuti、M. Sbert 和 R. Timofte。采用浓雾图像与无雾图像的图像去雾基准。在 IEEE国际图像处理会议 (ICIP) 上，第 1014–1018 页，2019。2[2] C.O. Ancuti、C. Ancuti 和 R. Timofte。采用浓雾图像与无雾图像的图像去雾基准。在 IEEE国际图像处理会议 (ICIP) 上，第 1014–1018 页，2019。2, 5[3] C. Ancuti、C.O. Ancuti、R. Timofte 等。NTIRE 2018 图像去雾挑战赛：方法与结果。载于 IEEE/CVF计算机视觉与模式识别会议研讨会论文集，第 891–901 页，2018。1[4] C.O. Ancuti、C. Ancuti、R. Timofte 等。NTIRE 2019图像去雾挑战赛报告。载于 IEEE/CVF计算机视觉与模式识别会议研讨会论文集，第 2241–2253 页，2019。1[5] C.O. Ancuti、C. Ancuti、F.A. Vasluianu、R. Timofte 等。NTIRE 2020非均匀去雾挑战赛。载于 IEEE/CVF计算机视觉与模式识别会议研讨会论文集，2020。1, 2, 3[6] C.O. Ancuti、C. Ancuti、F.A. Vasluianu、R. Timofte 等。NTIRE 2021 非均匀去雾挑战赛报告。载于 IEEE/CVF计算机视觉与模式识别会议研讨会论文集，2021。1, 2, 3, 5[7]Codruta O Ancuti、Cosmin Ancuti、Florin-Alexandru Vasluianu 和 Radu Timofte。NTIRE 2023 非均匀去雾挑战赛。载于 IEEE/CVF计算机视觉与模式识别会议研讨会论文集，2023。5, 8[8] D. Berman、T. Treibitz 和 S. Avidan。非局部图像去雾。载于 CVPR，第 1674–1682 页，2016。2[9] B. Cai、X. Xu、K. Jia、C. Qing 和 D.Tao。单幅图像去雾的端到端系统。IEEE图像处理汇刊，25(11):5187–5198，2016。1, 2[10] D. Chen、M. He、Q. Fan、J. Liao、L. Zhang、D. Hou、L. Yuan 和 G. Hua。用于图像去雾和去雨的门控上下文聚合网络。载于 IEEE 冬季应用计算机视觉会议，第 1375–1383 页，2019。1, 2[11] L. Chi、B. Jiang 和 Y. Mu。快速傅里叶卷积。《神经信息处理系统进展》，33:4479–4488，2020。2, 3[12] J. Deng、W. Dong、R. Socher、L. Li、K. Li 和 F. Li。ImageNet：大规模分层图像数据库。载于 IEEE计算机视觉与模式识别会议，第 248–255 页，2009。2[13] H. Dong、J. Pan、L. Xiang、Z. Hu、X. Zhang、F. Wang 和 M. Yang。采用密集特征融合的多尺度增强去雾网络。载于 CVPR，第 2157–2167 页，2020。1, 2[14] A. Dosovitskiy、A. Kolesnikov、L. Beyer、D. Weissenborn、X. Zhai、T. Unterthiner、M. Dehghani、M. Minderer、G. Heigold、S. Gelly、J. Uszkoreit 和 N. Houlsby。一张图像胜过 16×16 个词：用于图像识别的变换器及其尺度。在国际学习表征会议上，2021。2, 3

[15] M. Fu、H. Liu、Y. Yu、J. Chen 和 K. Wang。Dw-gan：一种用于非均匀去雾的离散小波变换 GAN。载于 IEEE/CVF计算机视觉与模式识别会议研讨会论文集，第 203–212 页，2021 年。1、2、3、5[16] S. Gao、M. Cheng、K. Zhao、X. Zhang、M. Yang 和 P. Toll。Res2net：一种新型多尺度骨干架构。IEEE模式分析与机器智能汇刊，43(2)：652–662，2021 年。1、2[17] I. Goodfellow、P. A. Jean、M. Mehdi、B. Xu、W.F. David、O. Sherjil、C. Aaron 和 B. Yoshua。生成对抗网络。计算机协会，63(11)：139–144，2020 年。1、5[18] K. Han、A. Xiao、E. Wu、J. Guo、C. Xu 和 Y. Wang。Transformer in Transformer。神经信息处理系统进展，34：15908–15919，2021 年。2[19] K. He、J. Sun 和 X. Tang。利用暗通道先验进行单幅图像去雾。IEEE模式分析与机器智能汇刊，33(12)：2341–2353，2016 年。1、2[20] K. He、X. Zhang、S.Ren 和 J. Sun。用于图像识别的深度残差学习。载于 CVPR，第 770–778 页，2016 年。3[21] M. Ju、Z. Gu 和 D. Zhang。基于改进大气散射模型的单幅图像去雾。神经计算，260：180–191，2017 年。1[22] L.Dai、X. Liu、C. Li 和 J. Chen。Awnet：用于图像 ISP 的注意力小波网络。载于欧洲计算机视觉会议研讨会（ECCVW），2020 年。2[23] B. Li、X. Peng、Z. Wang、J. Xu 和 D. Feng。利用暗通道先验进行单幅图像去雾。载于 CVPR，第 4770–4778 页，2017 年。1、2[24] B. Li、W. Ren、D. Fu、D. Tao、D. Feng、W. Zeng 和 Z. Wang。单幅图像去雾及更广泛任务的基准测试。IEEE图像处理汇刊，28(5)：492–505，2019 年。1、5[25] J. Li、Z. Pan、Q. Liu 和 Z.Wang。用于显著目标检测的通道注意力堆叠式 U 形网络。IEEE多媒体汇刊，23：1397–1409，2020 年。2[26] Y. Li、Q. Miao、J. Song、Y. Quan 和 W. Li。基于雾霾物理特性和自适应天空区域检测的单幅图像去雾。神经计算，182：221–234，2016 年。1[27] J. Liu、H. Wu、Y. Xie、Y. Qu 和 L. Ma。三叉戟去雾网络。载于 IEEE/CVF计算机视觉与模式识别会议研讨会论文集，2020 年。1、7[28] X. Liu、Y. Ma、Z. Shi 和 J. Chen。Griddehazenet：用于图像去雾的基于注意力的多尺度网络。载于 IEEE/CVF国际计算机视觉会议（ICCV），2019 年。1[29] X. Liu、Z. Shi、Z. Wu、J. Chen 和 G. Zhai。Griddehazenet+：一种用于单幅图像去雾的增强型多尺度网络，具有任务内知识迁移能力。IEEE智能交通系统汇刊，2022 年。1[30] Z. Liu、Y. Lin、Y. Cao、H. Hu、Y. Wei、Z. Zhang、S. Lin 和 B. Guo。Swin Transformer：分层视觉 Transformer

采用移位窗口。载于 IEEE/CVF国际计算机视觉会议论文集，第 10012–10022 页，2021 年。2、3[31] Z. Liu、H. Mao、C. Wu、C. Feichtenhofer、T. Darrell 和 S. Xie。用于 2020 年的卷积网络。载于 CVPR，第 11976–11986 页，2022 年。2、3[32] Z. Lu、J. Jiang、J. Huang、G. Wu 和 X. Liu。Glama：用于通用图像修复的空间与频率联合损失。载于 IEEE/CVF计算机视觉与模式识别会议研讨会论文集，第 1301–1310 页，2022 年。2[33] E.J. McCartney。大气光学：分子和粒子的散射。1976 年。2[34] G. Meng、Y. Wang、J. Duan、S. Xiang 和 C. Pan。基于边界约束和上下文正则化的高效图像去雾。载于 IEEE国际计算机视觉会议论文集，第 617–624 页，2013 年。1[35] W. E. K. Middleton。透过大气观察。多伦多大学出版社，1952 年。1、2[36] S.J. Pan 和 Q. Yang。迁移学习综述。IEEE 知识与数据工程汇刊，22(10)：1345–1359，2010 年。2、3[37] X. Qin、Z. Wang、Y. Bai、X. Xie 和 H. Jia。FFA-Net：用于单幅图像去雾的特征融合注意力网络。载于 AAAI人工智能会议论文集，第 11908–11915 页，2020 年。1、2、4、7[38] W. Ren、L. Ma、J. Zhang、J. Pan、X. Cao、W. Liu 和 M. Yang。用于单幅图像去雾的门控融合网络。载于 CVPR，第 3253–3261 页，2018 年。1[39] O. Ronneberger、P. Fischer 和 T. Brox。UNet：用于生物医学图像分割的卷积网络。载于医学图像计算与计算机辅助干预国际会议，第 234–241 页，2015 年。4[40] K. Simonyan 和 A. Zisserman。用于大规模图像识别的超深卷积网络。arXiv 预印本 arXiv:1409.1556，2014 年。5[41] V. A. Sindagi、P. Oza、R. Yasarla 和 V.M. Patel。针对雾霾和雨天条件的基于先验的领域自适应目标检测。载于 ECCV，第 763–780 页，2020 年。1[42] X. Song、D. Zhou、W. Li、H. Ding、Y. Dai 和 L. Zhang。WSAMF-Net：用于单幅图像去雾的基于小波空间注意力的多流反馈网络。IEEE视频技术电路与系统汇刊，33(2)：575–588，2023 年。2[43] R. Suvorov、E. Logacheva、A. Mashikhin、A. Remizova、A. Ashukha、A. Silvestrov、N. Kong、H. Goka、K. Park 和 V. Lempitsky。采用傅里叶卷积的分辨率鲁棒大掩码图像修复。载于 IEEE/CVF冬季计算机视觉应用会议论文集，第 2149–2159 页，2022 年。2、3[44] C. Tan、F. Sun、T. Kong、W. Zhang、C. Yang 和 C. Liu。迁移学习综述。载于《人工神经网络与机器学习——ICANN 2018：第 27 届国际人工神经网络会议论文集》，第 270–279 页，2018 年。2、3[45] H. Wang、Y. Xu、Z. Wang、Y. Cai、L. Chen 和 Y. Li。CenterNet-Auto：一种多目标视觉检测算法基于改进的 CenterNet 用于自动驾驶场景。IEEE新兴计算智能主题汇刊，第 1–11 页，2023 年。1[46] S. Woo、S. Debnath、R. Hu、X. Chen、Z. Liu、I.S. Kweon 和 S. Xie。ConvNeXt V2：利用掩码自编码器协同设计并扩展卷积网络。2023 年。2[47] H. Wu、J. Liu、Y. Xie、Y. Qu 和 L. Ma。用于非均匀去雾的知识迁移去雾网络。载于《IEEE/CVF计算机视觉与模式识别会议研讨会论文集》，2020 年。1[48] H. Wu、Y. Qu、S. Lin、J. Zhou、R. Qiao、Z. Zhang、Y. Xie 和 L. Ma。用于紧凑型单幅图像去雾的对比学习。载于 CVPR，第 10551–10560 页，2022 年。1, 3 [49] H. Yu、N. Zheng、M. Zhou、J. Huang、Z. Xiao 和 F. Zhao。用于图像去雾的频率与空间双重引导。载于 ECCV，第 181–198 页，2022 年。1, 3[50] Y. Yu、H. Liu、M. Fu、J. Chen、X. Wang 和 K. Wang。通过集成学习实现非均匀去雾的双分支神经网络。载于《IEEE/CVF计算机视觉与模式识别会议研讨会论文集》，第 193–202 页，2021 年。4[51] Y. Yu、F. Zhan、S. Lu、J. Pan、F. Ma、X. Xie 和 C. Miao。WaveFill：用于图像修复的基于小波的生成网络。载于 IEEE/CVF 国际计算机视觉会议论文集，第 14114–14123 页，2021 年。3[52] L. Yuan、Y. Chen、T. Wang、W. Yu、Y. Shi、Z. Jiang、Francis E.H. Tay、J. Feng 和 S. Yan。Tokens-to-Token ViT：在 ImageNet 上从头开始训练视觉变换器。载于 IEEE/CVF 国际计算机视觉会议论文集，第 558–567 页，2021 年。2[53] H. Zhang 和 V.M. Patel。密集连接金字塔去雾网络。载于《IEEE/CVF计算机视觉与模式识别会议研讨会论文集》，第 3194–3203 页，2018 年。1[54] Z. Zhao、P. Zheng、S. Xu 和 X. Wu。深度学习目标检测：综述。《IEEE神经网络与学习系统汇刊》，30(11):3212–3232，2019 年。1[55] Z. Zhong、T. Shen、Y. Yang、Z. Lin 和 C. Zhang。用于小波域超分辨率的团结构联合子带学习。《神经信息处理系统进展》，31，2018 年。3[56] Z. Zhou 和 Z. Shi。CGGAN：用于单幅图像去雾的上下文引导生成对抗网络。《IET图像处理》，14(15):3982–3988，2020 年。1[57] Q. Zhu、J. Mai 和 L. Shao。使用颜色衰减先验的快速单幅图像去雾算法。《IEEE多媒体汇刊》，24(11):3522–3533，2015 年。2[58] F. Zhuang、Z. Qi、K. Duan、D. Xi、Y. Zhu、H. Zhu、H. Xiong 和 Q. He。迁移学习综合综述。《IEEE会刊》，109(1):43–76，2020 年。2, 3