

DOI:10.19651/j.cnki.emt.2416080

# 融合外部注意力的扩散模型巡检图像去雾<sup>\*</sup>

周景 田兆星 王满意

(华北电力大学控制与计算机工程学院 北京 102206)

**摘要:**为降低雾天对输电线路巡检图像造成的影响,针对目前主流去雾算法存在计算成本高、图像去雾后检测性能差和难以部署的问题,提出了一种雾天输电线路巡检图像去雾方法 Diff-EaT。该方法采用融合 Transformer 的扩散模型结构,为降低体征提取模块中多头自注意力在 ViT 中的计算复杂度,使用多头外部注意力代替多头自注意力以减少计算负荷和增强特征学习。同时,设计了一个混合尺度门控前馈网络,在输入特征的深度可分离卷积之后集成了选通机制以改善局部信息捕获。在合成数据集和真实数据集上进行对比试验,定量指标和定量指标都证明其有效性,复原图像细节更加清晰。在去雾检测系统中,对真实巡检图像去雾后使用 YOLOv7 进行检测,mAP@0.5、召回率、查准率分别提升 6.92%、9.58%、4.11%,本文方法去雾后有效提高检测置信度,去雾检测系统可应用于实际场景。同时在消融实验中,证明了改进的有效性。

**关键词:**扩散模型;输电线路巡检;图像增强;Transformer;外部注意力

中图分类号: TP391; TN98 文献标识码: A 国家标准学科分类代码: 520.6040

## Diffusion model for foggy inspection image defogging by incorporating external attention

Zhou Jing Tian Zhaoxing Wang Manyi

(School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China)

**Abstract:** To reduce the impact of foggy days on transmission line inspection images, a foggy day transmission line inspection image de-fogging method, Diff-EaT, is proposed for the current mainstream de-fogging algorithms that have high computational costs, poor detection performance after image de-fogging, and difficult to deploy. The method adopts a fusion Transformer's diffusion model structure, and to reduce the computational complexity of the multi-head self-attention in the feature extraction module in the ViT, multi-head external attention is used instead of multi-head self-attention to reduce the computational load and enhance feature learning. Meanwhile, a mixed-scale gated feed-forward network is designed to integrate a pick-and-pass mechanism after the depth-separable convolution of input features to improve local information capture. Tested on synthetic and real datasets, quantitative and quantitative metrics prove their effectiveness with clearer details of recovered images. In the defogging detection system, the real inspection images are defogged and then detected using YOLOv7, mAP@0.5, recall rate, and checking accuracy rate are improved by 6.92%, 9.58%, and 4.11%, respectively, and this paper's method effectively improves the detection confidence after defogging. Defogging detection systems can be applied in real-world scenarios. Also in ablation experiments to demonstrate the effectiveness of its improvements.

**Keywords:** diffusion model; transmission line inspection; image enhancement; Transformer; external attention

## 0 引言

输电线路巡检是保障电力系统安全运维的重要手段,近几年无人化视觉检测为输电线路巡检开启新道路。无人机巡检时,不免遇到雾霾天气,严重影响所捕获拍摄图像的

质量<sup>[1]</sup>和降低图像分类<sup>[2]</sup>、目标检测<sup>[3]</sup>等高层计算机视觉任务的性能,对输电线路巡检工作的开展造成影响。因此,从输电线路巡检图像中恢复出损失的背景信息和色彩信息,具有重要的研究意义和应用价值<sup>[4]</sup>。为了减轻雾天对图像细节的影响,早期的研究集中于通过物理模型去雾,从

收稿日期:2024-05-20

\*基金项目:国家自然科学基金(52179014)项目资助

图像因雾霾影响造成退化的原因考虑,反演退化过程。为了解决这个不适定的逆问题,Zhu 等<sup>[5]</sup>构建了饱和度差、亮度和深度三者之间的线性模型,通过透射图去除雾霾。然而,基于物理模型的方法只适用在特定场合,在复杂的雾天下难以实现通用性。随着深度学习时代的到来,卷积神经网络(convolutional neural network,CNN)在图像去雾领域取得不错的效果,研究人员采用 CNN 作为主干,并将此问题建模为像素级图像回归过程。Cai 等<sup>[6]</sup>第一次将卷积神经网络引入到图像去雾任务中,提出了一种端到端的可训练去雾网络,利用多尺度卷积操作提取雾霾特征,大幅提升了图像去雾性能。Dong 等<sup>[7]</sup>提出了一个双视觉注意网络的联合图像去雾和透射率估计算法,充分利用了透射率估计和去雾过程中捕获信息的相关性,实现了图像中雾气浓度信息的预测;Wang 等<sup>[8]</sup>提出了一种分离特征和协同网络下的去雾模型,利用神经网络提取不同深度的空间信息及细节特征,使得复原图像颜色自然,细节保持良好。但卷积架构具有固有的局限性,不能很好的恢复复杂和长距离的特征效果。CNN 的体系结构对其权重施加局部性和权重共享两个硬约束。具体来说,卷积运算依赖于固定内核从局部区域收集信息,这削弱了对长范围依赖关系建模的能力。

最近,扩散模型 DDPM<sup>[9]</sup>因其在一系列视觉任务中的有效性而受到广泛关注,如图像恢复、图像超分辨率、图像去噪,并且取得非常优秀的性能和效果。扩散模型在建模图像像素分布方面能力强大,有利于高质量的图像恢复。Chen 等<sup>[10]</sup>研究发现在扩散模型中,去噪过程比扩散过程更至关重要。DDPM 扩散模型在去噪网络中均使用 U-Net 的网络结构。Chen 等<sup>[11]</sup>研究表明,使用 Vision Transformer(ViT)<sup>[12]</sup>的类条件扩散模型取得了比 U-Net 更好的性能,这归咎于多头注意力可以更好地对非局部信息进行建模。Bao 等<sup>[13]</sup>证明了 ViT 在扩散模型中的可用性。然而,ViT 的计算复杂度依赖于输入分辨率,带来了昂贵的训练代价和缓慢运行速度的问题,限制了其在高分辨率巡检图像恢复中的使用。Uformer<sup>[14]</sup>是一种基于 Swin Transformer 的统一图像恢复算法,其将注意力限制在固定窗口内,分担了计算负荷。然而,这些方法仍然有很高的计算需求,影响训练和推理时间。DDPM 虽然有效,但需要大量时间进行培训和取样。在噪声估计部分直接使用 ViT 的网络结构可以显着减少推理时间。作为回应,Guo 等<sup>[15]</sup>引入了外部注意力,这是一种适用于各种视觉任务的轻量级和通用机制。在此基础上,提出了噪声估计外部注意力(external attention transformer,NE-EAT),它大大减轻了计算负荷,同时采用多头外部注意力(multi-head external attention,EA)来增强特征学习。此外,Chen 等<sup>[16]</sup>的研究证明了门控机制在改善 ViT 中前馈网络的非线性表示方面的有效性,从而提高了其整体性能。

综上所述,为了解决扩散模型 U-ViT 在雾天输电线路

巡检工作中的计算成本高、图像去雾后检测性能差和难以部署的实际问题。针对上述问题,本文提出一种融合外部注意力的扩散模型巡检图像去雾方法 Diff-EaT。首先在特征提取阶段由于 ViT 使用全局注意力机制增加计算负担,本文采用多头外部注意力机制来降低计算负荷,并且增强特征学习。同时本文采用了混合尺度门控前馈网络来替代 MLP 前馈网络,该网络在输入特征的深度可分离卷积之后集成了选通机制,以生成用于图像去伪的上级特征。通过对比实验、消融实验表明去雾效果优异,将真实雾天场景下的巡检图像输入去雾检测系统,检测性能显著提升,可应用于实际巡检工作中。

## 1 去噪扩散概率模型

去噪扩散概率模型 DDPM 分为前向扩散过程和逆向扩散过程。前向扩散过程可以描述为逐渐破坏数据结构的前向加噪处理,通过不断向输入的数据中添加低水平的噪声(高斯噪声),直到全部变成高斯噪声。

该过程可以视为马尔科夫链过程,满足:

$$q(x_{1:T} | x_0) = \prod_{t=1}^T q(x_t | x_{t-1}) \quad (1)$$

$$q(x_t | x_{t-1}) = N(x_t, \sqrt{1-\beta_t}x_{t-1}, \beta_t I) \quad (2)$$

其中,  $q(x_{1:T} | x_0)$  代表先验分布,  $N$  代表高斯分布,  $I$  代表噪声, 不同时间  $t$  的  $\beta_t$  是预先定义好的, 由时间  $1 \sim T$  的递增, 可以是线性也可以是非线性, 满足  $\beta_1 < \beta_2 < \dots < \beta_T$ 。据以上公式, 可以通过重参数化采样得到:

$$x_t \sim N(0, I), \alpha_t = 1 - \beta_t, \bar{\alpha}_t = \prod_{i=1}^t \alpha_i \quad (3)$$

经过推导可以得到  $x_t$  和  $x_0$  系为:

$$q(x_t | x_0) = N(x_t; \sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)I) \quad (4)$$

逆向过程称为去噪过程, 如果得到逆向过程  $q(x_{t-1} | x_t)$  可以通过随机噪声, 逐步还原出一张清晰输电线路检测图。扩散模型使用  $p_\theta(x_{t-1} | x_t)$  逆向过程  $q(x_{t-1} | x_t)$ 。

$$q(x_{t-1} | x_t, x_0) = N(x_{t-1} | \tilde{\mu}_t(x_t, x_0), \tilde{\beta}_t I) \quad (5)$$

$$p_\theta(x_{t-1} | x_t) = N(x_{t-1} | \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)) \quad (6)$$

$$p_\theta(x_{0:T}) = p(x_T) \prod_{t=1}^T p_\theta(x_{t-1} | x_t) \quad (7)$$

其中,  $p_\theta(x_{t-1} | x_t)$  代表后验分布,  $\mu_\theta(x_t, t)$  代表参数化估计均值。扩散模型中不计算方差, 通过神经网络 U-Net 拟合均值  $\mu_\theta$ , 从而得到  $x_{t-1}$ 。

$$\mu_\theta = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_{\theta(x_t, t)}) \quad (8)$$

其中,  $t$  和  $x_t$  是已知, 只需训练去噪网络对参数  $\epsilon_{\theta(x_t, t)}$  的预测。

扩散模型作为一种新兴的生成式模型, 更易进行高质量的图像生成, 其适用性和泛化性更强。因此更适用于处理复杂的雾天输电线路巡检图像信息, 更适合应用到实际

工作中。本文方法 DF-EAT 体框架如图 1 所示,包括两个过程:上面部分是前向扩散,此过程为无参数学习,在前向过程中不断加入高斯噪声;下面部分为逆向扩散过程,进行参数优化,图中  $y$  表示图片信息,  $t$  表示时间信息。

## 2 DiffT-Net 去噪网络

### 2.1 整体架构

DiffT-Net 是一个去噪扩散网络,网络模型如图 2 所示,使用 DiT-Res 作为主体。将有雾图输入到模型里面,根据神经网络学习高斯条件分布进行反演。

输入巡检有雾图像  $I \in \mathbb{R}^{H \times W \times 3}$  和时间步长  $t$  作为输入。在每一层,DiT-Res 模块接收  $X_m$  和  $t$ 。 $t$  表示时间步

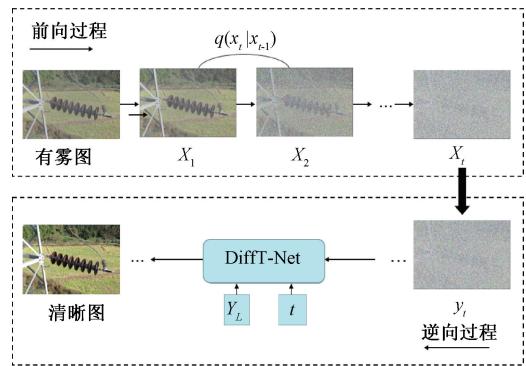


图 1 Diff-EaT 模型

Fig. 1 Diff-EaT model

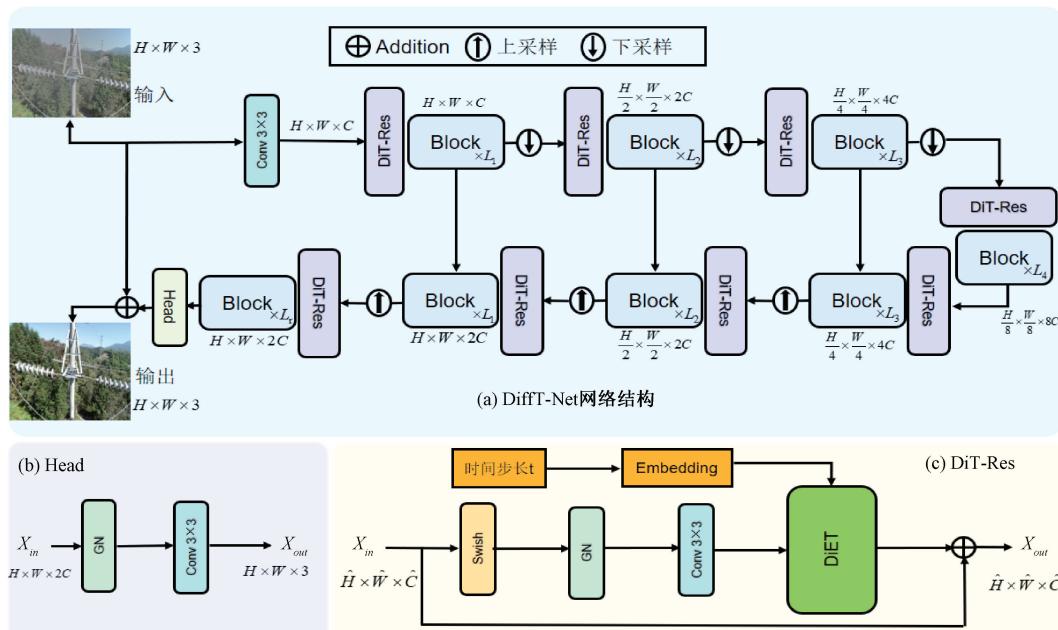


图 2 DiffT 网络架构图

Fig. 2 DiffT network architecture

长的时间令牌,通过将时间位置嵌入馈送到具有激活的 MLP 中获得时间令牌。在前半过程中每经过 DiT-Res 模块进行一次下采样,在每个下采样的过程中会对特征图的尺寸减半,这样能降低参数量,提高计算效率。最后经过上采样输出,使用 GN 函数激活,并通过卷积输出清晰图像。

DiT-Res 使用 DiET 模块作为网络组成的基础部分,具体结构如图 2(c)所示。为了解决随着网络加深而过拟合的问题,将 DiT-Res 模块的输入设置为双信息输入,分别为特征图信息和时间  $t$  信息。特征图信息经过归一化、激活函数和卷积处理。时间  $t$  信息进行正弦编码,经过激活函数和全连接层。之后将两个信息进行相加操作,进而信息融合。使用残差结构,让模型的梯度不易爆,训练起来更加准确同时也减少了计算资源的浪费。具体公式如下:

$$h(x) = F(x) + x \quad (6)$$

其中,  $h(x)$  代表输入,  $F(x)$  表示残差项,  $x$  代表输

入。简而言之,在残差模块中采用跳跃链接,使输入与输出相加的方法,填补可能丢失的特征信息。

### 2.2 外部注意力模块(DiET)

DiET 模块具体结构如图 3(a)所示。对于输入图像,首先将其转换成空间表示并对其编码为  $T$  个令牌的序列。然后将每个补丁嵌入到输入中。 $x_t$  表示时间步长的时间令牌,通过将时间嵌入馈送到具有 swish 激活的小 MLP 来获得时间令牌。这个时间令牌被传递到 DiET 网络中的所有层。最后通过投影特征和时间令牌嵌入在一个共享的空间,一起馈送到一系列 Transformer 中进行处理。这样有利于捕获了长距离的空间和时间的依赖性。DiET 网络具体表示为:

$$\begin{aligned} \hat{x}_s &= MHEA(LN(x_s), x_t) + x_s \\ x_s &= MSFN(LN(\hat{x}_s)) + \hat{x}_s \end{aligned} \quad (7)$$

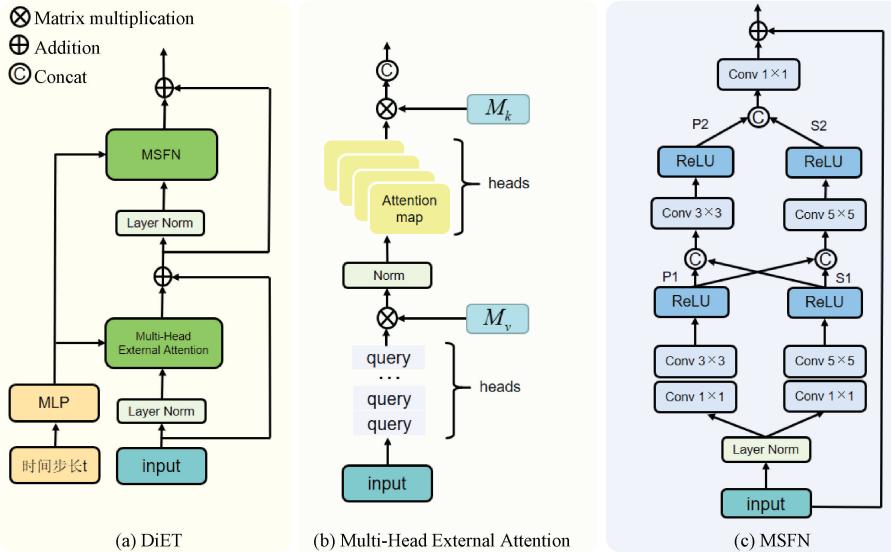


图 3 Diff Transformer 网络结构图

Fig. 3 Structure of Diff Transformer network

## 2.3 多头外部注意力模块(EA)

Peebles 等<sup>[17]</sup>研究发现 ViT 对长距离的建模能力和架构搭建相比于 CNN 更具有优势, 在提高一系列计算机视觉任务的性能方面发挥了重要作用。然而, ViT 中的自注意力机制是基于全局的, 计算和存储器开销随着图像的空间分辨率成二次方地增长, 不适合处理在输电线路巡检过程中拍摄到的高分辨率图像, 这会造成巨大的计算成本。为减少计算成本, 在 DiET 中引入外部注意力(EA)。如图 3(b)所示。

EA 计算输入像素和两个外部存储器单元  $M_k \in \mathbb{R}^{S \times d}, M_v \in \mathbb{R}^{S \times d}$  之间的注意力, 通过:

$$\begin{aligned} A &= \text{Norm}(F M_k^T) \\ F_{out} &= A M_v \end{aligned} \quad (8)$$

其中,  $M_k$  和  $M_v$  是两个可训练的参数, 其用作整个训练数据集的记忆, 并且不依赖于输入。 $A$  是从这个学习到的知识库级别的先验知识推断出的注意力地图。EA 的计算复杂度为  $O(dSN)$ , 其中  $d$  和  $S$  是超参数。因此, EA 计算复杂度为  $O(N)$  具有线性复杂度。

与自注意力机制(SA)中使用的余弦相似性不同, EA 选择双重归一化来分别规范化列和行, 可以通过这种方法来规避由矩阵乘法产生的注意力权重对输入特征尺寸的敏感性。双重归一化公式化为:

$$\begin{aligned} (\hat{\alpha})_{i,j} &= F M_k^T \\ \hat{\alpha}_{i,j} &= \exp(\hat{\alpha}_{i,j}) / \sum \exp(\hat{\alpha}_{k,j}) \\ \alpha_{i,j} &= \hat{\alpha}_{i,j} / \sum_k \hat{\alpha}_{i,k} \end{aligned} \quad (9)$$

为了捕捉令牌之间的不同关系, 在不同的输入通道上多次计算 EA, 提高了单头注意的能力, EA 可以考虑到不

同样本元素之间的关系, 可以学到不同领域像素节点之间的关联信息。多头外部注意力可以表述为:

$$h_i = \text{ExternalAttention}(F_i, M_k, M_v) \quad (10)$$

$$F_{out} = \text{MultiHead}(F, M_k, M_v) = \text{Concat}(h_1, \dots, h_H)W_o \quad (11)$$

其中,  $h_i$  是第  $i$  个头部  $m$ ,  $H$  是头部的数量,  $W_o$  是使输入和输出的维度一致的线性变换矩阵。

## 2.4 混合尺度门控前馈网络模块(MSFN)

先前的研究通常在常规前馈网络中引入单尺度卷积以提高局部性。然而, 这些研究都忽略了雾霾的多尺度相关性。事实上, 丰富的多尺度表示已经充分证明了其在去除雾霾方面有更好的有效性。本文将通过在传输过程中插入两个多尺度深度卷积路径来设计 MSFN, 具体网络结构参见图 3(c)。给定一个输入张量  $\mathbf{X}_{l-1} \in \mathbb{R}^{H \times W \times C}$ , 经过层归一化后, 首先利用  $1 \times 1$  卷积以  $r$  的比例扩展通道维数, 然后将其送入两个并行分支。在特征变换过程中, 采用  $3 \times 3$  和  $5 \times 5$  深度卷积增强多尺度局部信息提取。以这种方式, 所开发的 MSFN 的整个特征融合过程被公式化为:

$$\begin{aligned} \hat{\mathbf{X}}_l &= f_{1 \times 1}^c(LN(\mathbf{X}_{l-1})) , \mathbf{X}_l^{s_1} = \sigma(f_{5 \times 5}^{dwc}(\hat{\mathbf{X}}_l)) \\ \mathbf{X}_l^{p_2} &= \sigma(f_{3 \times 3}^{dwc}[\mathbf{X}_l^{s_1}, \mathbf{X}_l^{s_1}]), \\ \mathbf{X}_l^{p_2} &= \sigma(f_{3 \times 3}^{dwc}[\mathbf{X}_l^{s_1}, \mathbf{X}_l^{s_1}]) \\ \mathbf{X}_l^i &= f_{1 \times 1}^c[\mathbf{X}_l^{p_2}, \mathbf{X}_l^{s_2}] + \mathbf{X}_{l-1} \end{aligned} \quad (12)$$

其中,  $\sigma(\cdot)$  表示 ReLU 激活,  $f_{n \times m}^c$  表示  $n \times m$  的卷积,  $[\cdot]$  是通道级联。

## 3 实验与分析

### 3.1 数据集

本文在输电线路巡检图像上实现去雾实验, 其中真实

雾天数据集 64 张以及通过合成雾气的方法生成薄雾数据集 2 000 张和重雾数据集 2 000 张。薄雾和重雾两个数据集分别用来训练出输电线路图像去雾模型, 其中 1 800 张输电线路雾图用于训练, 200 张用于测试。薄雾图像和重雾图像, 如图 4 所示。



图 4 薄雾与重雾数据集

Fig. 4 Thin and heavy fog datasets

雾天的成像原理, 大气散射模型的数学表达式为:

$$I(x) = J(x)t(x) + A(1 - t(x)) \quad (13)$$

式中:  $I(x)$  为含雾气图像;  $J(x)$  为原始图像;  $t(x)$  为透射率;  $A$  为全局大气光值。中心点合成雾是指通过设定图片的一个点为雾化中心点进行雾合成扩散, 距离雾化中心点的距离越远, 雾合成的效果越弱。大气透视率  $t(x)$  计算式为:

$$\begin{aligned} t(x) &= e^{-\beta d(x)} \\ d(x) &= \alpha \cdot \sqrt{(i - i_{ce})^2 + (j - j_{ce})^2} + c \\ (i_{ce}, j_{ce}) &= (\frac{H}{2}, \frac{W}{2}) \\ c &= \sqrt{\max(H, W)} \end{aligned} \quad (14)$$

式中:  $\beta$  为散射系数;  $d(x)$  为图像各点到雾化中心的距离;  $\alpha$  为比例系数;  $(i_{ce}, j_{ce})$  为图像雾化中心;  $(H, W)$  为图像的高和宽;  $c$  为雾化尺寸。 $\beta$  可以控制雾气的浓度,  $\beta$  值越大表示雾气浓度越大; 雾化尺寸  $c$  决定生成雾的大小,  $c$  值越大表示雾气范围越大。

通过和真实雾天场景样本对比, 本文将比例系数  $\alpha$  设置为  $-0.04$ , 为增加雾天场景多样性, 本文轻雾图像的  $\beta$  值在  $0.06 \sim 0.08$  的范围内随机选取, 浓雾图像的  $\beta$  值在  $0.13 \sim 0.16$  的范围内随机选取, 然后调整全局大气光值  $A$ ,  $A$  值在  $0.5 \sim 0.8$  的范围内随机选取, 通过上式生成雾图。

### 3.2 设置与训练

本文实验平台使用 Windows10 操作系统, 是在 Python3.8 编程语言和 Pytorch1.8.1 框架上使用一个单一的 NVIDIA GeForce RTX 2080GPU 运行以及 16 GB 的内存, Cuda 为 11.1。本文在训练过程中选择 Adam 优化器。batch size 为 8, 初始学习速率设定为  $1 \times 10^{-4}$ 。在微调阶段, 学习速率被调整到  $1 \times 10^{-5}$  和  $1 \times 10^{-6}$  之间的值。直到网络被训练误差在 50 个 epochs 内几乎没有下降, 最终获得的最佳模型训练了 4 200 轮次, 如果继续训练会导致该去雾扩散模型过拟合。实验参数如表 1 所示。

表 1 实验参数设置

Table 1 Experimental parameter settings

参数名称	参数值
迭代次数	4 200
批处理大小	8
优化方法	Adam 优化器
权重衰减	$1 \times 10^{-5} \sim 1 \times 10^{-6}$
初始学习率	$1 \times 10^{-4}$
图片大小	5 472 × 3 648

### 3.3 评价标准

针对衡量图片恢复质量本文采用峰值信噪比 (PSNR)、结构相似度指数 (SSIM) 作为复原图像质量评估指标。图像的 PSNR 主要用于衡量图像中信号与噪声的占比, 在于复原领域使用的目的在于衡量处理后复原图像与真实图像之间的相似度, 数值越高说明复原性能越好。PSNR 数学定义如下:

$$PSNR = 10 \lg \frac{255^2}{\frac{1}{3HW} \sum_{i=1}^H \sum_{j=1}^W \|y(i, j) - o(i, j)\|_2^2} \quad (15)$$

其中,  $H$  和  $W$  分别代表图像的高和宽,  $o(i, j)$  和  $y(i, j)$  分别代表坐标为  $(i, j)$  的原雾天图像像素和增强后的图像像素。SSIM 将图像的结构信息定义为独立于亮度、对比度的, 用于衡量处理后图像与原图的结构相似性的指标。SSIM 数值范围为  $0 \sim 1$ , 越大表示图像增强性能越好。数学表达式:

$$SSIM = \frac{(2\mu_o\mu_y + C_1)(2\delta_{oy} + C_2)}{(\mu_o^2 + \mu_y^2 + C_1)(\delta_o^2 + \delta_y^2 + C_2)} \quad (16)$$

其中,  $\mu_o$  和  $\mu_y$  分别是原始图像和滤波后图像的均值,  $\delta_o$  和  $\delta_y$  分别是原始图像和滤波后图像的方差,  $\delta_{oy}$  是原始图像和滤波后图像的协方差,  $C_1$  和  $C_2$  是两个常数, 为与现有方法统一, 分别设置为  $C_1 = 6.50$ ,  $C_2 = 58.52$ 。

### 3.4 合成数据集对比试验

#### 1) 定性比较

本小节使用本文算法与其他 5 中主流的去雾方法 CAP<sup>[18]</sup>、GFN<sup>[19]</sup>、RefineDNet<sup>[20]</sup>、Uformer<sup>[14]</sup>、U-ViT<sup>[13]</sup> 在薄雾和重雾数据集上分别进行对比试验。CAP 算法是一种基于物理模型的去雾算法, 通过建立暗通道模型来估计大气光强度和透射率, 进而恢复无雾的场景颜色, 可以对图像做出简单快速的处理, 但对细粒度的细节恢复较差。GFN 算法依赖于编码器和解码器组成的端到端神经网络, 与本文算法相一致使用类似 Unet 结构, 同样使用了前后跳跃链接。RefineDNet 算法采用弱监督学习, 不过度依赖成对的图像对进行训练, 同时结合了 CAP 算法和 GFN 算法的优点, 使用了物理先验和基于学习的方法。Uformer 算法是近年被提出的基于 U 形的 Swin Transformer 方法,

通过使用移动窗口的注意力机制快速节省计算量,性能优异。U-ViT 算法是将 ViT 和扩散模型相结合这归咎于 Transformer 注意力机制对远距离信息的利用,与本文同样使用注意力机制和扩散模型的结构,但本文采用多头外部注意力机制和混合尺度门控网络,不同于其使用的标准多头自注意力机制以及单尺度门控前馈网络。不同算法恢复巡检图像的视觉效果如图 5 所示。上面两行为薄雾数据集远景、近景实验结果,下面两行为重雾数据集远景、近景实验结果。

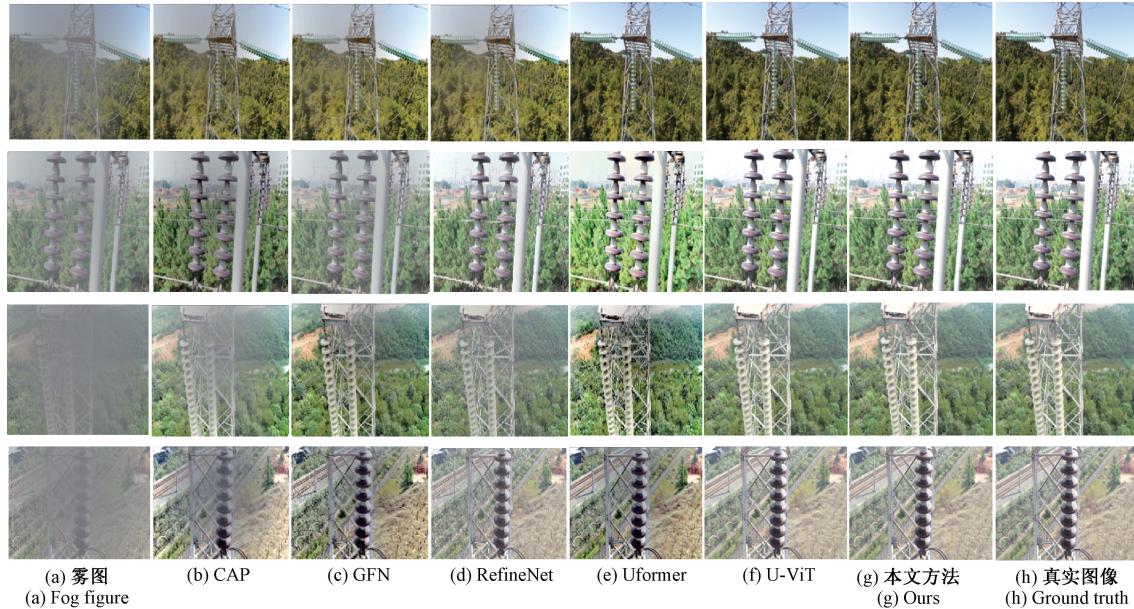


图 5 不同方法去雾结果图

Fig. 5 Graph of the results of different methods of defogging

## 2) 定量比较

表 2 表明了不同方法在薄雾和重雾数据集的定量评估。

表 2 不同算法去雾对比实验

Table 2 Comparative of different algorithms for defogging

方法	薄雾数据集		重雾数据集	
	PSNR	SSIM	PSNR	SSIM
CAP	30.08	0.880 5	16.38	0.747 2
GFN	31.03	0.932 9	28.61	0.922 6
RefineNet	31.88	0.933 0	25.83	0.899 5
Uformer	31.92	0.930 8	29.81	0.920 2
U-ViT	32.28	0.940 6	30.23	<b>0.931 9</b>
本文方法	<b>32.30</b>	<b>0.941 3</b>	<b>31.68</b>	<b>0.944 7</b>

如表 2 所示,本文方法在薄雾数据集和重雾数据集均高于其他方法,在薄雾数据集上 PSNR 和 SSIM 相较于 U-ViT 基础模型分别提升 0.38 dB 和 0.000 7 dB,在重雾数据集上 PSNR 和 SSIM 分别提升 1.45 dB 和 0.012 8 dB。

在图 5 中,本文算法能够有效的去除较大较厚面积的雾气遮挡,恢复背景细节信息。相较于本文算法其他算法无法有效恢复背景细节信息,如 CAP 算法,图中依然有大面积的可视雾气遮挡。特别地,Uformer 算法在处理薄雾时性能出色,但在重雾数据集的表现上无法恢复细小细节,如第 3 行对防震锤这种小目标恢复较为欠缺,U-ViT 算法复原图像较为真实,但运行速度缓慢,而本文算法对细小纹理恢复更加真实更加贴近真实图像,同时花费较少时间。

这说明本文的去雾扩散模型相较于传统 CNN 方法和 Transformer 方法更具有优势。

## 3.5 真实数据集对比试验

为验证本文方法在真实雾天巡检场景下的图像恢复表现,在含有 64 张真实雾天输电线路巡检的数据集上进行实验。实验使用真实数据集在去雾检测系统上进行去雾后使用 YOLOv7<sup>[21]</sup> 进行目标检测。通过去雾前和去雾后的平均精确率 mAP@0.5 (mean average precision)、召回率 Recall、查准率 Precision 进行评估。实验表明,通过使用 Diff-Eat 方法后进行检测,mAP@0.5、召回率、查准率较原雾天巡检图像分别提升 6.92%、9.58%、4.11%,具体实验结果如表 3 所示。

表 3 实验检测结果

Table 3 Experimental test results

名称	mAP@0.5	Recall	Precision
原图	88.62	80.60	93.15
去雾图	95.54	90.18	97.26

为了更加直观显示去雾后检测的性能以及能否达到预期,本文在图 6 显示了真实场景雾天巡检图像去雾前后检测的视觉效果。可以发现,在雾天检测时出现了错检和漏检的情况,与雾天图像相比,本文的方法可以有效地衰减雾天造成的影响,错检和漏检的情况大幅减少,尤其是

对防震锤小目标的检测。如图 6 第一列,在雾天时检测出现大量错检,对复原后的图像进行检测,此情况没有发生。这主要得益于 Diff-EaT 方法对图像背景细节的把控和恢复,并且生成计算机易于理解的像素语言。经证实该系统可以应用到实际巡检任务中,达到巡检工作的预期目标。



图 6 真实场景去雾检测系统结果对比  
Fig. 6 Real scene defogging detection system results

### 3.6 前馈网络消融实验

为了评估所提出的 MSFN 的有效性,本文将其与 3 个基线进行比较:1)传统前馈网络(FN),2)Dconv 前馈网络(DFN),3)门控 Dconv 前馈网络(GDFN)。Rain200H 的定量分析结果见表 4。虽然 GDFN 在两个相同尺度的深度卷积流中引入了门控机制以带来性能优势,但它仍然忽略了多尺度知识以进行去盲。通过增加不同尺度的局部特征提取和融合,MSFN 确实可以更好地提高性能,并实现比 GDFN 高 0.21 dB 的 PSNR 增益。

表 4 前馈网络消融实验结果

Table 4 Feedforward network ablation experiment results

前馈网络	PSNR	SSIM
FN	29.43	0.9020
DFN	30.25	0.9171
GDFN	31.29	0.9398
MSFN	<b>31.68</b>	<b>0.9347</b>

### 3.7 注意力机制消融实验

作为 DiET 的主要构成部分,外部注意力机制能更加关注到特征部分,同时减少计算负荷。本文在薄雾数据集上进行消融实验,图 7 展示了使用 U-ViT 的自我注意力和

外部注意力对图像进行复原后进行注意力图检测,可以发现改进后的外部注意力能够更好的特征聚合,生成更加真实的图像信息,易于计算机理解,为后续检测任务提供支持与可能性。

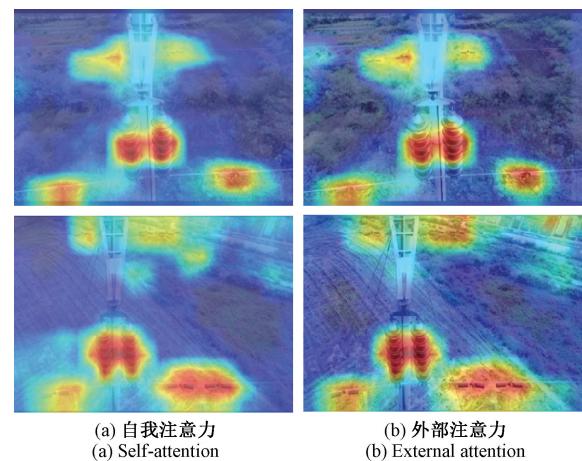


图 7 外部注意力图与自我注意力图对比

Fig. 7 External attention map vs. self-attention map

如表 5 所示,描述了以 FCN 为特征主体的体系结构以进行消融实验。批次大小和总迭代次数分别为 8 次和 3 000 次。实验分别对 U-ViT 的自我注意力机制和外部注意力机制、归一化方法进行探究。结果表明,外部注意力

比自我注意力提供更好的准确性,并且归一化方法对外部注意力产生较大的正向影响,对图像复原有改善作用。同

时减少了大量计算成本,提高了计算效率,适合部署在移动设备中,满足输电线路巡检的工作需求。

表 5 外部注意力消融实验结果

Table 5 Results of external attention ablation experiments

方法	Backbone	Norm	PSNR	SSIM	$MACs/\times 10^9$
FCN	ResNet-50	—	29.43	0.9020	29.41
FCN+SA	ResNet-50	Batch Norm	30.25	0.9171	32.32
FCN+SA	ResNet-50	Layer Norm	31.29	0.9198	31.26
FCN+EA	ResNet-50	Batch Norm	31.45	0.9214	19.76
FCN+EA	ResNet-50	Layer Norm	31.57	0.9276	18.87

### 3.8 步长消融实验

为了探究采样步长的影响,对薄雾验证集的 100 张图像进行了  $4\times$  噪声 SR ( $\sigma n=0.05$ ) 实验,采样步骤  $T \in [10, 15, 20, 50, 100, 200, 500, 1000]$ 。超参数分别固定为  $\lambda=8.0$  和  $\lambda=0.3$ 。定量结果如图 8 中所示。从该图中观察到,虽然 PSNR 与采样步长呈对数线性关系,但对于  $T \in [100, 500]$ , LPIPS 得分最低。LPIPS<sup>[22]</sup> 为学习感知图像块相似度,数值越低越好。因此,Diff-EaT 可以用少于 100 个采样步长生成去雾的图像,本文中采样步长的默认数量为 100。

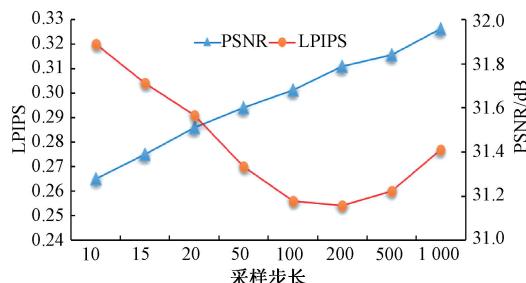


图 8 采样步长的影响

Fig. 8 Effect of sampling step size

## 4 结 论

针对雾天对输电线路巡检成像的影响,本文提出了一种融合外部注意力的扩散模型雾天巡检图像复原方法 Diff-EaT。该方法采用融合外部注意力的扩散模型结构,采用多头外部注意力代替多头自注意力以减少计算负荷和增强特征学习。同时,设计了一个混合尺度门控前馈网络,在输入特征的深度可分离卷积之后集成了选通机制以改善局部信息捕获。多个定性定量实验表明本文所提方法有效提高图像复原质量,对后续检测视觉任务提供支持,可应用于实际场景。在接下来的工作中,将进一步探索视频增强方法,以实现更高效的视频复原。

## 参考文献

[1] 吴正平,程洁莹,雷帮军,等.基于特征注意力的快速非均匀雾图像去雾算法[J].国外电子测量技术,2023,

- 42(9):9-18.
- WU ZH P, CHENG J Y, LEI B J, et al. Defogging algorithm for fast non-uniform fog images based on feature attention[J]. Foreign Electronic Measurement Technology, 2023, 42(9):9-18.
- [2] 陈燕,杨艳,杨春兰,等.基于阶段聚焦损失和并行增广策略的遥感图像场景分类[J].电子测量与仪器学报,2023,37(1):116-122.
- CHEN Y, YANG Y, YANG CH L, et al. Remote sensing image scene classification based on phase focus loss and parallel augmentation strategy[J]. Chinese Journal of Electronic Measurement and Instrumentation, 2023, 37(1):116-122.
- [3] 周道先,张吟龙,徐高飞,等.基于形变卷积和深层聚合网络的水下文物检测[J].仪器仪表学报,2023,44(11):185-195.
- ZHOU D X, ZHANG Y L, XU G F, et al. Underwater cultural relics detection based on deformation convolution and deep aggregation network[J]. Chinese Journal of Scientific Instrument, 2023, 44 ( 11 ): 185-195.
- [4] 王子昭,景明利,史金钢,等.一种改进 CBAM 机制和细节恢复的单幅图像去雾算法[J].电子测量技术,2023,46(2):161-168.
- WANG Z ZH, JING M L, SHI J G, et al. A single image defogging algorithm with improved CBAM mechanism and detail recovery [J]. Electronic Measurement Technology, 2023, 46(2):161-168.
- [5] ZHU Q S, MAI J M, SHAO L. A fast single image haze removal algorithm using color attenuation prior[J]. IEEE Transactions on Image Processing, 2015, 24 ( 11 ): 3522-3533.
- [6] CAI B L, XU X M, JIA K, et al. DehazeNet: An end-to-end system for single image haze removal[J]. IEEE Transactions on Image Processing, 2016, 25 ( 11 ): 5187-5198.

- [7] DONG Y, LIU Y H, ZHANG H, et al. FD-GAN: Generative adversarial networks with fusion-discriminator for single image dehazing[C]. AAAI Conference on Artificial Intelligence, 2020: 10729-10736.
- [8] WANG W K, HU H P, CAO SH F, et al. Adaptive single image defogging based on sky segmentation[J]. Multimedia Tools and Applications, 2023, 82: 46521-46545.
- [9] HO J, JAIN A, ABBEEL P. Denoising diffusion probabilistic models [J]. Advances in Neural Information Processing Systems, 2020, 33: 6840-6851.
- [10] CHEN X L, LIU ZH, XIE S N, et al. Deconstructing denoising diffusion models for self-supervised learning[J]. ArXiv preprint arXiv:2401.14404, 2024.
- [11] CHEN J N, LU Y Y, YU Q H, et al. TransUNet: Transformers make strong encoders for medical image segmentation [J]. ArXiv preprint arXiv: 2102.04306, 2021.
- [12] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth  $16 \times 16$  words: Transformers for image recognition at scale [J]. ArXiv preprint arXiv:2010.11929, 2020.
- [13] BAO F, NIE SH, XUE K W, et al. All are worth words: A ViT backbone for diffusion models [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 22669-22679.
- [14] WANG ZH D, CUN X D, BAO J M, et al. Uformer: A general u-shaped transformer for image restoration[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 17683-17693.
- [15] GUO M H, LIU ZH N, MU T J, et al. Beyond self-attention: External attention using two linear layers for visual tasks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45 (5): 5436-5447.
- [16] CHEN X, LI H, LI M Q, et al. Learning a sparse transformer network for effective image deraining[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 5896-5905.
- [17] PEEBLES W, XIE S N. Scalable diffusion models with transformers[C]. IEEE/CVF International Conference on Computer Vision and Pattern Recognition, 2023: 4195-4205.
- [18] ZHU Q S, MAI J M, SHAO L. A fast single image haze removal algorithm using color attenuation prior[J]. IEEE Transactions on Image Processing, 2015, 24 (11): 3522-3533.
- [19] ZHANG D Y, LIANG ZH W, SHAO J. Joint image deblurring and super-resolution with attention dual supervised network[J]. Neurocomputing, 2020, 412: 187-196.
- [20] ZHAO SH Y, ZHANG L, SHEN Y, et al. RefineDNet: A weakly supervised refinement framework for single image dehazing [J]. IEEE Transactions on Image Processing, 2021, 30: 3391-3404.
- [21] WANG C Y, BOCHKOVSKIY A, LIAO H Y. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464-7475.
- [22] BLAU Y, MICHAELI T. The perception-distortion tradeoff[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 6228-6237.

### 作者简介

**周景**(通信作者),副教授,博士,主要研究方向为人工智能、计算机视觉。

E-mail: zhouding108@ncepu.edu.cn

**田兆星**,硕士研究生,主要研究方向为计算机视觉、图像去噪。

E-mail: 1337132263@qq.com

**王满意**,硕士研究生,主要研究方向为安全运维、目标检测。

E-mail: wmy@ncepu.edu.cn