

TF-Net：一种优化泰勒公式模块的 图像去噪算研究*

文◆北京科技大学天津学院 张洪峰 孙沛叶 刘俊培 孙墨林 刘云豪

引言

在图像去雾任务中，由于注意力机制的二次计算复杂度，Transformer 网络的应用受到了限制。为应对这个挑战，本文提出了一种新颖的 Transformer 变体，采用泰勒扩展来近似注意力机制，实现了线性轻量级的复杂度。为进一步提高模型性能，本文引入多尺度注意力细化模块，用于校正泰勒扩展引入的误差。此外，本文对提出的变形器进行了优化，引入多分支架构和多尺度嵌入方法。多分支架构通过不同尺度的重叠变形卷积嵌入特征，多尺度补丁嵌入的设计基于不同大小的感受野、多层次语义信息、灵活的感受野形状大小的理念。本文的模型被称为由泰勒公式扩展的多分支变形器（TF-Net），在补丁嵌入阶段更灵活地嵌入从粗到细的特征，以最小的计算成本捕捉长距离像素之间的相互作用。通过多个去雾基准测试，TF-Net 在较轻的计算负担下实现了最先进的性能，解决了注意力机制模块的计算复杂度问题，有效提高了图像去雾任务的性能。

1 研究背景

单幅图像去雾是一项图像修复任务，旨在从受雾霾影响的图像中估算出无雾霾图像。从基于卷积神经网络 (Convolutional Neural Networks, CNN)^[1] 的方法开始，这项任务在去雾方面取得了巨大胜利，深度去雾网络的性能得到了显著提高，但 Transformer 的感受野仍受到限制。

为应对这些挑战，提出了一种创新方法 TF-Net。首先，针对 Transformer 计算复杂度的问题，引入小空间窗口内的自注意机制，适应像素级的去雾任务，减少计算量，提高模型效率。其次，对视觉 Transformer 进行改进，引入灵活的尺度变换器，以适应不同尺度的变化。

模型经过创新性的改进，使其更适用于图像去雾任务，并且仍然保持了 Transformer 在计算机视觉任务中的通用性。TF-Net 是一种改进的方法，相较于 SOTA 方法，在 SOTS-Indoor 数据集上进行训练。在 TF-

Net 中，圆圈的大小与模型参数的数量成正比。TF-Net 在整个特征图上跨越空间维度应用自注意力机制，并且保持线性的计算复杂度。通过对 Softmax 函数进行泰勒展开计算自注意力的权重，应用矩阵乘法的关联法则，将自注意力的计算复杂度降到最低。这种策略具有以下优点。（1）保留了 Transformer 模型对数据间长距离依赖关系的建模能力，避免了因窗口分割而导致的感受野减小的问题。（2）TF-Net 提供了更强的值近似，与传统的 Transformer 模型相似。（3）TF-Net 使得 Transformer 模型关注像素级的交互而不是信道级的交互，能更精细地处理特征。

为了解决泰勒展开公式中忽略余数形式带来的误差问题，提出了多尺度注意力细化 (Multi-Scale Attention Refinement, MSAR) 模块^[2]。该模块利用图像内部的局部相关性，通过对查询和关键字的局部信息进行卷积操作，生成带有缩放因子的特征图。特征图的通道数量与多头自注意

*【基金项目】天津市教委科研计划项目“一种光伏板综合维护检测车研制”（项目批准号 2023KJ251）

【作者简介】张洪峰（1986—），男，山东潍坊人，硕士研究生，讲师，研究方向：先进制造技术研究与自动控制。

力 (Multi-headed Self-attention, MSA) 的头数相等, 因此每个头都有对应的缩放因子。

本文提出了一种名为 TF-Net 的“编码器—解码器”骨干网络, 基于多尺度补丁嵌入和泰勒公式多分支“编码器—解码器”结构。多尺度补丁嵌入模块具有不同大小的感受野、多层次的语义信息和灵活的感受野形状。考虑到每个标记的生成都应遵循局部相关性先验, 故对可变形卷积的偏移量进行了截断。为了降低计算复杂度和参数数量, 采用了深度可分离方法。将不同尺度的标记输入 TF-Net 中, 并进行融合, 多尺度补丁嵌入模块能够生成不同尺度和维度的标记, 而多分支结构能够同时处理这些标记, 从而捕捉到更强大的特征。TF-Net 网络结构图如图 1 所示, (a)TF-Net 是基于多尺度补丁嵌入的多分支分层设计组成的模型, (b)多尺度补丁嵌入技术能够将粗到细的补丁嵌入到模型

中, (c)TF-Net 具有线性计算复杂度, 能够高效计算, (d)MSAR 模块能够对泰勒扩展中的误差进行补偿。

2 去雾方法

关于自我关注的高效设计, 变形器的计算复杂度对计算资源的要求较高。因此, 一些研究采用基于滑动窗口的自我关注, 这种设计限制了 Transformer 对数据中长距离依赖关系建模的能力。

采用不使用卷积核函数的策略, 直接对 Softmax 函数进行泰勒展开, 确保了对 Softmax 函数和数值的准确近似, 并降低计算复杂度, 使模型在处理长距离像素之间的依赖关系时更加高效。

在高级视觉领域中, 金字塔结构是一种常见且简单的方法。IFormer 通过截取结构融合高频和低频信息; CrossViT 和 MPViT 采用多分支处理多尺度的图像块, 以获得多尺度的感知范围。在低级视觉领域, MSP-Former 利用多尺度投影帮助 Transformer 模型表示复杂的退化环境, Giqe 通过多分支处理不同尺寸的特征图。但这些方法很少进一步探索多尺度补丁和多分支架构, 且卷积核的采样点数量固定。相比之下, 多尺度可变形卷积不仅具有灵活的采样点。

3 TF-Net 方法

本文的目标是建立一个基于 Transformer 的高效、轻量级去雾网络。为降低计算复杂度, 使用 Softmax-attention 的泰勒扩展满足关联法则, 并采用 U-net 结构。为了补偿泰勒扩展误差的影响, 提出了一个 MSAR 模块, 首先描述 TF-Net 整体架构, 然后介绍 3 个核心模块, 即多尺度补丁嵌入、泰勒扩展自注意力机制和 MSAR 模块。

给定一幅模糊图像 I , 其大小为 $3 \times h \times w$, 使用卷积操作进行浅层特

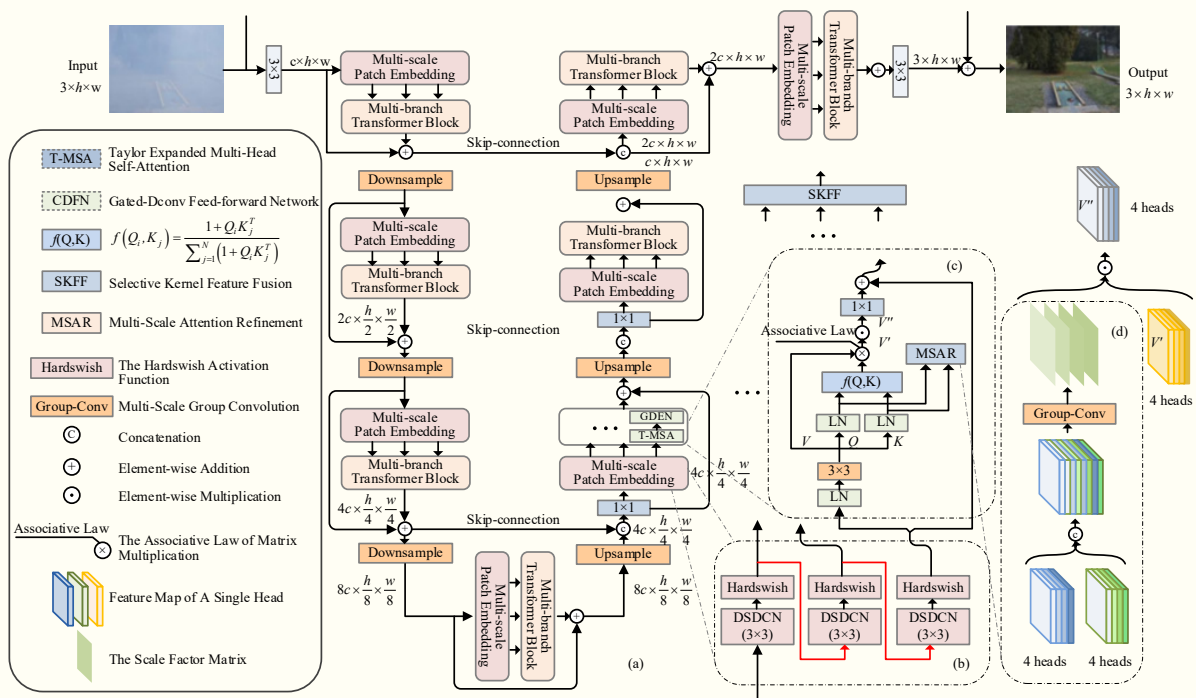


图 1 TF-Net 网络结构图

征提取, 得到 F , 其大小为 $c \times h \times w$ 。采用“四级编码器—解码器”网络进行深度特征提取, 每个阶段都包含一个残差块, 其中包含多尺度补丁嵌入和多分支变换器。使用多尺度补丁嵌入技术生成多尺度标记, 并将其输入多个变换器分支中, 每个变换器分支都包含多个变换器编码器。在多分支 Transformer 块的末尾应用 SKFF 模块, 对不同分支生成的特征进行融合。由于泰勒变换器具有卓越的性能和多分支设计, 压缩通道和变换器块的数量。在每个阶段, 使用像素取消洗牌和像素洗牌操作对特征进行下采样和上采样。使用跳跃连接聚合编码器和解码器的信息, 并通过 1×1 卷积层进行降维, 为了恢复精细的结构和纹理细节, 在“编码器—解码器”结构之后使用一个残差块。最后, 使用 3×3 卷积层减少通道, 并输出残差图像 $R \in R \times h \times w$ 。为进一步压缩计算量和参数, 在图像处理层应用深度可分离卷积。DSDCN 网络结构示意图如图 2 所示, 偏移由 $K \times K$ 深度卷积和点卷积生成, 输出由 $K \times K$ 深度可变形卷积和点卷积生成。这意味着在生成偏移量时, 使用了一个 $K \times K$ 的深度卷积和一个点卷积, 而在生成输出时, 使用了一个 $K \times K$ 的深度可变形卷积和一个点卷积。

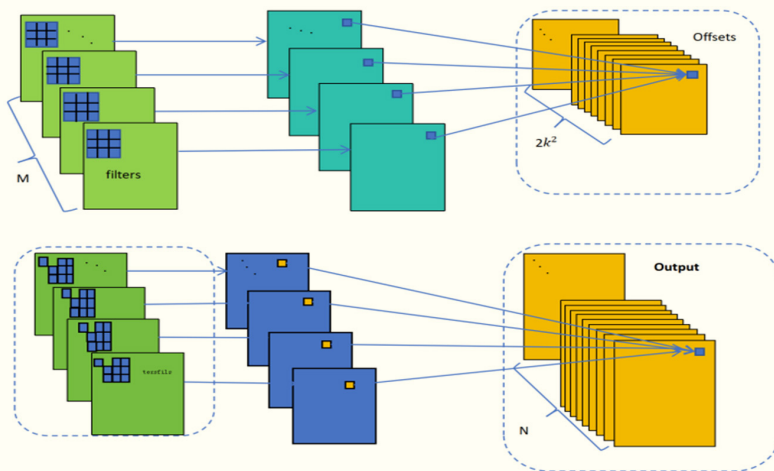


图 2 DSDCN 网络结构示意图

3.1 多尺度补丁嵌入模块

本文提出了一种创新的多尺度补丁嵌入方法, 采用堆叠扩展接收场的操作, 堆叠多个具有小核的可变形卷积层, 不仅增加了网络的深度, 提供了多层次的语义信息, 还有助于减少参数和计算负担。为了激活可变形卷积层的输出, 在所有可变形卷积层后面使用了 Hardswish 激活函数。采用新的多尺度补丁嵌入方法, 更好地捕捉不同尺度和层次的视觉特征, 提高对视觉元素多样性的建模能力。与深度可分离卷积的策略类似, 提出了深度可分离和可变形卷积 (Deep-Separable and Deformable Convolutions, DSDCN), 将可变形卷积 (DCN) 的各个部分分解为深度卷积和点卷积, 标准 DCN 和 DSDCN 对于 $h \times w$ 图像的计算代价如下。

$$\Omega(\text{DCN})=2MK^4 hw+MNK^2 hw+4MK^2 hw \quad (1)$$

$$\Omega(\text{DSDCN})=8MK^2 hw+MN hw \quad (2)$$

其中, M 和 N 分别为输入和输出通道数, K 为卷积核大小。DCN 和 DSDCN 的参数如下。

$$P(\text{DCN})=2MK^4+MNK^2 \quad (3)$$

$$P(\text{DSDCN})=4MK^2+MN \quad (4)$$

与 DCN 相比, DSDCN 显著降低了计算复杂度和参数数量。考虑到图像的局部相关性以及斑块嵌入可以捕捉到特征图的基本元素, 因此视觉元素应更加关注局部区域。通过截断偏移量控制补丁嵌入层的感受野范围, 在实践中选择的偏移量为 $[-3, 3]$ 。根据视觉对象的形状, 通过学习选择感受野的大小, 其上限为 9×9 , 相当于 $\text{BF}=4$ 的扩张卷积, 下限为 1×1 。当并行设置多尺度补丁嵌入时, 不同分支的感受野大小依次为 $x \in [1, 9], y \in [x, x+8], z \in [y, y+8]$ 。

3.2 泰勒扩展自注意力机制

假设 Queries (Q), Keys (K), Values (V) 是一个 $h \times w$ 的特征向量序列, 维数为 D , 其中 h 和 w 分别是图像的高度和宽度。变换器 Transformer 的计算公式如下。

$$V' = \text{Soft max} \left(\frac{Q^T K}{\sqrt{D}} \right) V^T \quad (5)$$

由于 $Q \in Rhw \times D, K \in Rhw \times D, V \in Rhw \times D$, Softmax 自我关注的计算复杂度为 O , 导致了昂贵的计算成本。要将自我关注的计算复杂度从 $O(h^2w^2)$ 降至 $O(hw)$, 首先将式 (5) 的广义注意力方程改写为如下形式。

$$V'_i = \frac{\sum_{j=1}^N f(Q_i, K_j) V_j}{\sum_{j=1}^N f(Q_i, K_j)} \quad (6)$$

其中, 下标为 i 的矩阵是矩阵第 i 行的向量, f 表示任何相似性函数。设 $f(Q_i, K_j) = \left| \exp \left(\frac{Q_i^T K_j}{\sqrt{D}} \right) \right|$ 时, 式 (6) 退化式为 (5)。如果本文应用泰勒公式对 $\exp \left(\frac{Q_i^T K_j}{\sqrt{D}} \right)$ 在 0 处进行一阶泰勒展开, 可将式 (6) 改写为

$$V'_i = \frac{\sum_{j=1}^N (1+Q_i^T K_j + o(Q_i^T K_j)) V_j}{\sum_{j=1}^N (1+Q_i^T K_j + o(Q_i^T K_j))} \quad (7)$$

此外，本文从向量 \tilde{Q}_i, \tilde{K}_i 的归一化中生成 \tilde{Q}_i and \tilde{K}_i 以近似 $\exp\left(\frac{Q_i^T K_j}{\sqrt{D}}\right)$ 。当 \tilde{Q}_i and \tilde{K}_i 的规范小于 1 时，使注意力图的值全部为正，在实践中得知规范归一化为 0.5 时效果最好。因此，有一个近似值，其一阶泰勒展开在定义域 $[-0.25, 0.25]$ ，忽略 Peano 的余项形式，得到自注意力的泰勒展开表达式如下。

$$V'_i = \frac{\sum_{j=1}^N (1 + \tilde{Q}_i^T \tilde{K}_j) V_j^T}{\sum_{j=1}^N (1 + \tilde{Q}_i^T \tilde{K}_j)} \quad (8)$$

本文激活函数图如图 3 所示。

最后，对式 (8) 应用矩阵乘法关联律，如下所示。

$$V'_i = \text{Taylor-Attention}(Q_i, K_i, V_i) = \frac{\sum_{j=1}^N V_j^T + \tilde{Q}_i^T \sum_{j=1}^N \tilde{K}_j V_j^T}{N + \tilde{Q}_i^T \sum_{j=1}^N \tilde{K}_j} \quad (9)$$

与 MDTA 类似，对 Q, K, V 采用深度卷积生成的方法，以强调局部上下文。使用多头结构，即从层级的顶部到底部，头部的数量逐渐增加。标准多头注意力机制和 T-MSA 模块对于一个 $h \times w$ 的补丁图像的计算复杂度如下。

$$\Omega(\text{MSA}) = 4hwD^2 + 2h^2w^2D \quad (10)$$

$$\Omega(\text{T-MSA}) = 18hwD + 7hwD^2 \quad (11)$$

通常情况下， $h \times w$ 远大于 D ，因此 T-MSA 提供了更多测试高分辨率图像的可能性，并确保其结果接近 MSA。

3.3 MSAR 模块

在 T-MSA 中对 Softmax 进行一阶泰勒展开时，没有考虑到 Peano 的提醒形式，导致近似误差的存在。对于泰勒展开的 n 阶余项 $\frac{(\tilde{Q}_i \tilde{K}_j)^n}{n!}$ ($n \geq 2$)，无法利用矩阵乘法组合规律使 T-MSA 的计算复杂度线性化。然而，余项与 Q 和 K 矩阵相关，考虑到图像具有局部相关性，学

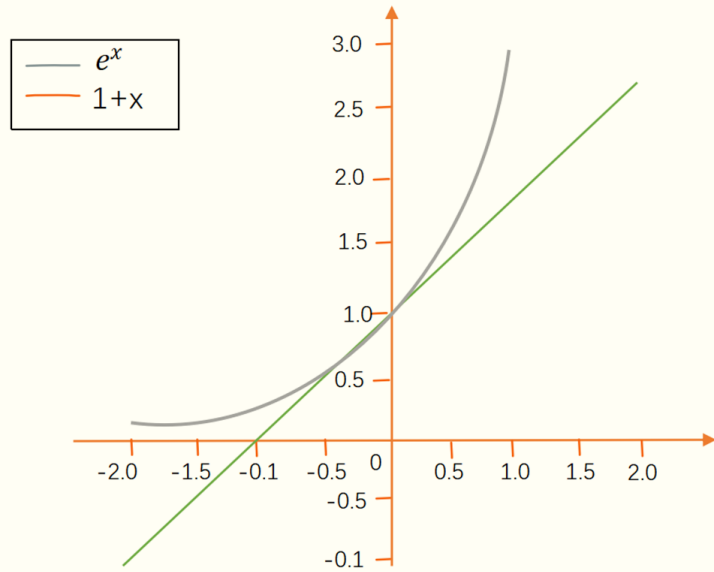


图 3 本文激活函数图

习 Q 和 K 矩阵的局部信息修正不准确的输出 V' 。此外，Conv-attention 模块允许泰勒展开更好地处理高频信息，对于多头注意力机制 $Q_m \in \mathbb{R}^{\text{head} \times \frac{D}{\text{head}} \times N}$ 和 $K_m \in \mathbb{R}^{\text{head} \times \frac{D}{\text{head}} \times N}$ ，将它们重塑为 $\hat{Q}_m \in \mathbb{R}^{\text{head} \times \frac{D}{\text{head}} \times H \times W}$ 和 $\hat{K}_m \in \mathbb{R}^{\text{head} \times \frac{D}{\text{head}} \times H \times W}$ 。其中，头表示头的数量，将 \hat{Q}_m 和 \hat{K}_m 沿通道维度串联，生成张量 $T \in \mathbb{R}^{\text{head} \times \frac{2D}{\text{head}} \times H \times W}$ ，然后通过多尺度分组卷积层生成门控张量 $G \in \mathbb{R}^{\text{head} \times 1 \times H \times W}$ 如下。

$$G = \text{Sigmoid}(\text{Concat}(T_1 W_1^G, \dots, T_{\text{head}} W_{\text{head}}^G)) \quad (12)$$

其中， $T_{\text{head}} \in \mathbb{R}^{\frac{D}{\text{head}} \times H \times W}$ 为 T 的 head-th head， W_{head}^G 为不同核的卷积。由于网络的不同层次具有不同数量的头，需要为不同数量的头选择相应的多尺度分组卷积。采用 T-MSA 方法和 MSAR 模块，计算出改进后的 T-MSA 模块。

$$\hat{X} = X + \text{Cat}(H_1 \odot G_1, \dots, H_{\text{head}} \odot G_{\text{head}}) W^P \quad (13)$$

$$H_i = \text{Taylor-Attention}(QW_i^Q, KW_i^K, VW_i^V)$$

其中， X 和 \hat{X} 表示输入和输出特征图。投影是参数矩阵 $\hat{W}^P \in \mathbb{R}^{D \times D}$ 、 $W_i^Q \in \mathbb{R}^{D \times \frac{D}{\text{head}}}$ 、 $W_i^K \in \mathbb{R}^{D \times \frac{D}{\text{head}}}$ 和 $W_i^V \in \mathbb{R}^{D \times \frac{D}{\text{head}}}$ 。

4 实验设计

4.1 实验细节

在 TF-Net 的实现过程中，提供了一种 TF-Net 架构，为了增加数据的多样性，使用随机裁剪和随机翻转的数据增强技术。初始学习率设置为 $2e-4$ ，并通过余弦退火法逐渐降低至 $1e-6$ ，仅选择使用 L1 损失作为损失函数。

本文对数据集进行了评估，包括合成数据集 (RESIDE) 和真实世界数据集 (O-HAZE、Dense-Haze)。在 RESIDE 数据集中，使用了 ITS 子集和 OTS 子集进行评估，ITS 子集包含了 13990 对室内图像，OTS 子集包含

表 1 各种方法对去马赛克基准的定量比较

Methods	SOTS-Indoor		SOTS-Outdoor		O-HAZE		Dense-Haze		Overhead	
	PSNR \uparrow	SSIM \uparrow	PSNR \uparrow	SSIM \uparrow	PSNR \uparrow	SSIM \uparrow	PSNR \uparrow	SSIM \uparrow	#Param	MACs
DCP	16.63	0.819	19.14	0.816	16.79	0.654	12.73	0.443	-	0.6G
DehazeNet	19.83	0.822	24.76	0.928	17.58	0.771	13.85	0.431	0.02M	0.6G
GFN	22.31	0.881	21.56	0.845	18.17	0.672	-	-	0.51M	14.9G
GDN	32.17	0.985	30.87	0.983	18.93	0.673	14.97	0.537	0.97M	21.5G
PFDN	32.69	0.977	-	-	-	-	-	-	11.28M	51.5G
MSBDN	33.68	0.986	33.49	0.983	24.37	0.750	15.14	0.556	31.36M	41.54G
FFA-Net	36.40	0.990	33.58	0.985	22.13	0.771	15.71	0.550	4.47M	287.8G
AECR-Net	37.18	0.991	-	-	-	-	15.81	0.467	2.62M	52.2G/216.0G
MAXIM-2S	38.12	0.992	34.20	0.986	-	-	-	-	14.11M	156.4G
SGID-PFF	38.53	0.992	30.21	0.976	20.97	0.742	12.50	0.518	13.88M	141.1G
Restormer	38.89	0.992	-	-	23.59	0.769	15.79	0.549	26.11M	60.4G
Dehamer	36.64	0.989	35.19	0.987	25.12	0.778	16.43	0.561	132.60M	38.6G
Ours(-B)	40.72	0.993	37.43	0.990	25.06	0.789	16.67	0.561	2.69M	88.2G
Ours(-L)	42.65	0.995	38.10	0.992	25.32	0.783	16.45	0.567	7.44M	

注：“-”表示没有结果。最佳和次佳结果分别以粗体和下划线标出。

了 313950 对室外图像，并使用 SOTS 子集对模型进行评估。真实世界数据集包括 O-HAZE 和 Dense-Haze，分别包含了 45 对和 55 对图像，将每个数据集的最后 5 幅图像作为测试集，其余的图像作为训练集。



图 4 合成雾霾图像的视觉对比

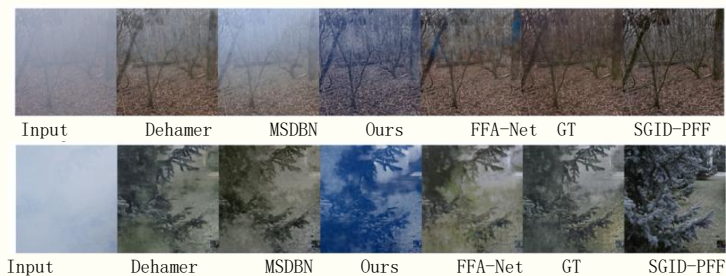


图 5 真实雾霾图像的视觉对比

4.2 SOTS-Indoor 数据对比
各种方法对去马赛克基准的定量比较如表 1 所示。将 TF-Net 和 SOTA 方法在合成数据集上的性能进行了对比，基本模型 TF-Net 在 SOTS-Indoor 数据集上取得了 40.71dB 的峰值信噪比 (PSNR) 和 0.994 的结构相似性指数 (SSIM)。与之前的最先进方法 SGID-PFF 相比，模型的 PSNR 提高了 2.19dB，参数数量仅为 SGID-PFF 的 10%，计算成本仅为 SGID-PFF 的 14%。模型 TF-Net 比 SGID-PFF 提高了 4.12dB，复杂度仅为 SGID-PFF 的一半左右，在 SOTS-Outdoor 上，比 Dehamer 方法高出 291dB。此外，比较了 TF-Net 和其他 SOTA 去雾方法的视觉效果，得出合成雾霾图像的视觉对比如图 4 所示。TF-Net 生成的图像在阴影和高频区域呈现出更自然的效果，还能有效去除雾霾，恢复阴影部分的细节，并且与地面实际情况更为相符，尤其是在高频区域。

真实雾霾图像的视觉对比如图5所示，第一行图像来自 O-HAZE 数据集，第二行来自 Dense-Haze 数据集。本文的 MB-TaylorFormer 模型能够生成色彩逼真、无人工痕迹的无雾霾图像。

4.3 Real Hazy 数据对比

对真实数据集 O-HAZE 和 Dense Haze 进一步评估，比较 TF-Net 和 SOTA 方法的性能，得出定量比较结果，并得出以下结论。

(1) 在 O-HAZE 和 Dense-Haze 数据集上，TF-Net 的 PSNR 值分别比之前的方法高出了 0.20dB 和 0.04dB。

(2) 通常情况下，当在小尺寸数据集上进行训练时，视觉变换器的非凸损失会导致性能不佳。

相较于其他基于 CNN 的模型，TF-Net 在小尺寸数据集上仍然取得了最佳的 PSNR 和 SSIM 分数。同时，将 TF-Net 与之前的方法进行了对比，FFA-Net 生成的图像存在颗粒感和细节损失，SGID-PFF 生成的图像存在色彩失真，Dehamer 生成的图像存在伪影和纹理损失。相比之下，本文生成的无雾图像更加清晰。

多尺度斑块嵌入和多分支结构的消融研究如表 2 所示，“-S”表示两个串联的卷积层，“-P”表示两个并联的卷积层，核大小同为 3。

4.4 消融实验

使用从 SOTS-Indoor 数据集中裁剪的 256×256 像素的图像块对所有的 MB-TaylorFormer 模型进行训练，训练历时设置为 500 个 Epoch。基于 MB-TaylorFormer-B 的基础，对框架中不同模块的有效性进行分析，使用 MB-TaylorFormer-L 探索分支维度、通道维度和深度对模型性能的影响。

表 2 多尺度斑块嵌入和多分支结构的消融研究

Branch	Type of Conv	PSNR	SSIM	#Params	MACs
Single	Conv	38.28	0.992	2.656M	33.64G
Double	Conv-P	38.43	0.992	2.653M	37.90G
	Dilated Conv-P	38.78	0.992	2.653M	37.90G
	Conv-S	39.05	0.993	2.653M	37.90G
	DSDCN-S	40.72	0.993	2.678M	38.52G

表 3 近似误差分析 (Softmax-attention 的近似误差越小，性能越好)

Methods	PSNR	SSIM	#Params	MACs
Swin	36.60	0.989	2.518M	36.39G
Swin+T-MSA-2nd	36.51	0.989	2.518M	36.39G
Swin+T-MSA-1st	36.38	0.988	2.518M	36.39G

表 2 中，进行了关于补丁嵌入和不同分支数量的差异研究。具体来说，以基于单尺度标准卷积的单分支模型作为基准，并对其进行以下修改。(1) 在单尺度模型和多分支 (Conv-P) 模型上进行补丁嵌入的设计。(2) 使用平行扩张卷积层 (DF=1,2) 嵌入补丁 (Dilated Conv-P)。(3) 使用标准卷积替代扩张卷积嵌入补丁，并采用串联两个卷积层的方法 (Conv-S)。(4) 使用 DSDCN 替代标准卷积 (DSDCN-S)。

实验结果表明，性能从优到劣依次为 DSDCN-S, Conv-S, Dilated Conv-S, ConvP 和 Conv。表明多尺度补丁嵌入技术可以灵活地嵌入补丁，并用另一个线性自我注意模块取代了模型中的 T-MSA 模块。

鉴于二阶 T-MSA (T-MSA-2nd) 不适用于关联律，会导致计算负担大，TF-Net 对 Swin 进行了一阶和二阶泰勒展开。近似误差分析 (Softmax-attention 的近似误差越小，性能越好) 如表 3 所示，T-MSA 能有效逼近 Softmax-attention，且 T-MSA-2nd 的性能非常接近 Softmax-attention。但 T-MSA-2nd 和 Softmax-attention 的计算复杂度随着图像分辨率的增加而呈二次方增加，在实际应用中难以对远距离像素关系进行建模。因此，最终选择了 T-MSA-1st 方法。

结语

本文介绍了一种名为 TF-Net 的多分支线性化变换器网络，包括多尺度斑块嵌入能有效嵌入各种视觉标记。使用矩阵乘法关联定律进行泰勒扩展，降低了计算复杂度。此外，通过门控注意力校正自我注意力的输出，使 TF-Net 同时执行远程关注和局部校正。经过在多个数据集上的实验，TF-Net 的有效性、轻便性和通用性得到了证明，在去雾任务中取得了显著的成绩。

引用

[1] Li, Z.,Liu,F.,Yang,W.,et al.A Survey of Convolutional Neural Networks: Analysis,Applications,and Prospects[J].IEEE Transactions on Neural Networks and Learning Systems,2021.

[2] Sinha,A.,&Dolz,J.Multi-scale Self-guided Attention for Medical Image Segmentation[J].IEEE Journal of Biomedical and Health Informatics,2020, 25(1):121-130.