

## HomoFormer: 用于图像阴影去除的均质化 Transformer

Jie Xiao<sup>1</sup> Xueyang Fu<sup>1†</sup> Yurui Zhu<sup>1</sup> Dong Li<sup>1</sup> Jie Huang<sup>1</sup> Kai Zhu<sup>2</sup> Zheng-Jun Zha<sup>1</sup>  
<sup>1</sup>University of Science and Technology of China <sup>2</sup>Alibaba Group

ustchbxj@mail.ustc.edu.cn xyfu@ustc.edu.cn

### Abstract

阴影退化的空间不均匀性和多样化模式与主导模型的权重共享方式相冲突，这可能导致不令人满意的折衷方案。为了解决这个问题，我们从阴影变换的角度提出了一种新的策略：直接均匀化阴影退化的空间分布。我们的关键设计是随机洗牌操作及其相应的逆操作。具体来说，随机洗牌操作随机地重新排列空间空间上的像素，而逆操作则恢复原始顺序。随机洗牌后，阴影在整个图像中扩散，并且退化以均匀的方式出现，可以通过局部自注意力层进行有效处理。此外，我们进一步设计了一种新的前馈网络和位置建模来利用图像结构信息。基于这些元素，我们构建了最终的基于局部窗口的变压器，名为 HomoFormer，用于图像阴影去除。我们的 HomoFormer 可以享受局部变压器的线性复杂性，同时绕过阴影的不均匀性和多样性的挑战。我们进行了大量的实验来验证我们的 HomoFormer 在公共数据集上的优越性。代码可在 <https://github.com/jiexiaou/HomoFormer> 获取。

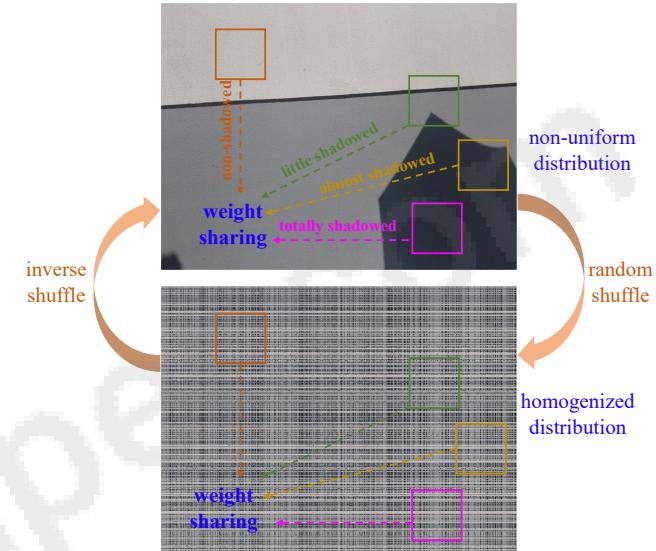


图 1. 阴影不均匀分布带来的挑战示意图。不均匀性对权重共享模型施加了限制，它们难以在不同退化程度的区域之间寻求妥协。随机洗牌创建了均匀分布，为权重共享局部自注意力奠定了基础。

### 一、简介

当光源被部分或完全遮挡时，在自然场景下捕获的图像中，阴影无处不在。然而，阴影不仅损害图像的视觉质量，而且对后续的各种下游视觉任务（例如对象跟踪[40]和深度学习）施加了严重限制。-保护[36]、人脸识别[56]等。因此，从阴影图像中恢复干净的图像具有重要意义。图像去阴影的主要障碍之一是阴影退化的空间分布不均匀。影子的图案多种多样。这些人物——

这些特性使得阴影很难通过主流模型（例如卷积神经网络（CNN）和基于窗口的 Transformer [34]）进行建模。这是由这些模型固有的权重共享属性引起的。内置权重共享属性决定了他们必须使用一组参数来覆盖复杂退化程度的影子情况，这可能会导致不令人满意的折衷。图 1 示意性地说明了这一想法。为了克服这一挑战，一个简单的解决方案是选择能够对空间异质性相互作用进行建模的高级模型。换句话说，期望的模型预计将根据阴影的具体内容采取自适应行动。一个有竞争力的候选者是普通视觉变压器 [4, 10]，它通过利用全局自注意力，能够处理：

† 通讯作者。

自适应地处理图像。尽管普通视觉变换器可以在一定程度上解决这种异质性问题，但它们的应用仍然受到输入分辨率的二次复杂度的限制，对于该任务来说，输入分辨率通常很高（例如  $840 \times 640$ ）。基于局部窗口的变换器[34]可以以线性复杂度有效地处理图像，但在处理非均匀阴影退化时它们会受到权重共享的影响。在这项工作中，我们转而探索硬币的另一面：是否有可能使非均匀分布均匀化，而不是被动地选择更复杂的适应模型？

为了实现这个目标，我们希望通过一些函数直接变换原始的非均匀阴影，该函数由专用操作对  $S(\cdot)$  和  $IS(\cdot)$  实现。 $S(\cdot)$  用于将非均匀分布投影到均匀空间，而  $IS(\cdot)$  是  $S(\cdot)$  的精确逆运算，将其投影回来。我们提供了一个潜在的解决方案， $S(\cdot)$  是随机洗牌操作， $IS(\cdot)$  是相应的逆洗牌操作。具体来说，通过跨空间空间的随机重排，每个像素都会以相等的概率分配到任意位置，达到均匀化非均匀分布的目的。另一方面，像素的随机重新排列彻底破坏了语义信息（见图1，洗牌后图像语义丢失）。因此，在均匀化空间中经过一系列计算后，需要反向投影以与原始空间对齐。幸运的是，随机洗牌操作可以完全反转。我们可以轻松地反转随机洗牌，通过恢复像素之间的原始相对位置来重建图像语义。值得注意的是，随机洗牌操作及其逆操作的实现成本非常低，无需引入额外的参数或 FLOP。

通过随机洗牌操作及其逆操作，我们评估了一个没有任何信息丢失的均质空间，消除了对具有权重共享属性的模型的约束（见图1）。现在，我们进入考虑图像去阴影的具体模型的阶段。一个微妙的问题是，由于随机洗牌破坏了相对位置关系，因此在同质空间中工作的所需层不应依赖于基于位置的信息来建模关系。考虑到这些因素，所需的层被实现为局部自注意力[34]，而不使用位置编码。我们可以将结构信息建模的责任转移到后续的前馈网络（FFN）层[43]。因此，我们引入了一个 localwindow Transformer，称为 HomoFormer，作为整体模型。HomoFormer 不仅具有输入分辨率的线性复杂性、变压器的强大表示能力，而且还绕过了对非均匀分布的阴影退化建模的挑战。我们进行前

紧张而全面的实验来验证我们的 HomoFormer 的优越性（第 4.2 节和 4.3 节）并解释其行为（第 4.4 节）。总而言之，这项工作的主要贡献包括：· 我们分析了建模的挑战我们设计了随机洗牌和逆洗牌，这是一个互补的操作对，没有任何信息丢失。· 我们精心设计了一种名为 Homo-Former 的局部窗口 Transformer，它可以处理具有线性复杂度的图像，同时避免建模不均匀和多样化的阴影。· 在基准数据集上进行了广泛而全面的实验来验证和进一步解释我们人类的优越性。

## 2. 相关作品

### 2.1. 图像阴影去除

用于阴影去除的经典方法 [11, 12, 42, 52] 通常利用各种手工先验，例如照明 [48, 55]、区域 [17]、密度 [1] 或用户交互 [14]。近年来，随着深度学习的蓬勃发展，基于学习的方法在图像阴影去除方面也取得了辉煌的进展。例如，De-shadowNet [38] 通过融合多级特征来聚合上下文信息来预测用于阴影去除的阴影遮罩。胡等人。[19, 20] 利用方向感知的空间上下文来检测和消除阴影。Mask-shadowGAN [21] 提出了一个框架，该框架从输入阴影图像中估计阴影掩模，并随后利用掩模作为阴影生成的指导来建立循环一致性约束。存等人。[8] 通过堆叠扩张卷积来利用上下文特征。陈等人。[6] 尝试通过将上下文信息从非阴影区域转移到阴影区域来去除阴影。傅等人。[13] 将阴影去除任务表述为多重曝光图像融合问题。DC-ShadowNet [23] 集成了域分类器和基于物理的损失来实现不成对的阴影去除。G2R [35] 和 BMNet [57] 都引入了阴影生成过程来提高阴影去除的性能。一些作品 [9, 33] 还采用生成对抗网络来增强阴影去除结果或不配对数据训练的真实性。SP+M-Net [26] 和 EMDN [58] 都试图提出用于阴影去除的合理阴影照明模型。[15] 利用通道注意力来利用阴影和非阴影区域之间的全局上下文相关性。万等人。[44] 提出一种风格指导

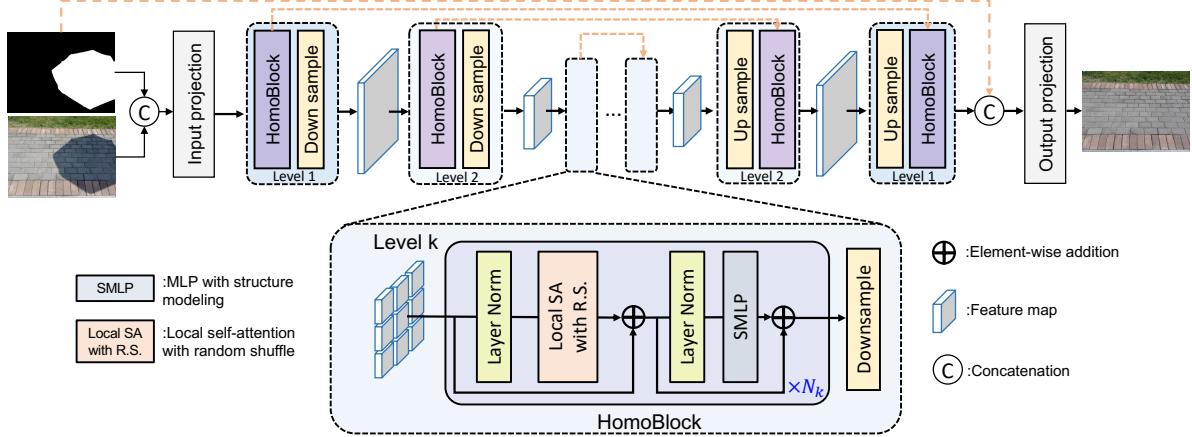


图2. 我们提出的 HomoFormer 的整体架构。HomoFormer 的核心是使用随机洗牌来均质化原始图像空间，并利用局部自注意力来建模均质化空间中的交互。

阴影去除网络可在阴影去除后获得更好的图像风格一致性。

## 2.2. 视觉变压器

视觉变形金刚[10,32,34,37,47]在视觉界取得了辉煌的成就。ViT [10]首先将图像块视为标记序列，并对其应用普通变换器进行图像分类。SwinTransformer [34]在自注意力之前引入了局部性和层次结构，并采用移位窗口自注意力为各种任务建立了有效的架构，包括图像分类、目标检测和语义分割。一些Transformer [4,31,46,49,50,53]也出现用于各种低级视觉任务。然而，它们中的大多数采用普通自注意力或移位窗口自注意力，要么面临巨大的复杂性，要么建模非均匀分布的阴影退化。对于图像阴影去除，ShadowFormer [15]采用通道注意力来聚合全局上下文而不是空间注意力，以避免昂贵的复杂性。与ShadowFormer相比，我们的动机主要是关注阴影的不均匀问题，我们的HomoFormer仍然采用视觉Transformer的“自注意力→MLP”的经典范式。

## 2.3. 图像洗牌策略

洗牌像素是计算机视觉中的一种常见技术，但我们将其去除阴影的潜力有了新的认识。Kang等人。[24]建议将局部补丁中的像素打乱作为训练正则化。除了明显的动机外，HomoFormer将洗牌范围扩大到整个图像。像素洗牌还在上采样/下采样以重塑特征中发挥作用 [29, 41]。它用于在通道和空间之间交换信息，

不随机破坏像素的空间重新排列。最近，为了捕获非局部交互，Xiao等人。[51]提出随机洗牌来替代Swin Transformer的移位窗口策略。与[51]不同，HomoFormer的动机是创建一个同质化的空间，这与权重共享机制兼容。此外，我们在整个网络中采用随机洗牌而不是取代移位窗口策略，并开发单独的SMLP模块来建模结构信息。

## 3. 方法

我们首先在第二节中介绍专用随机洗牌操作和逆洗牌操作。3.1. 然后在秒。3.2，我们称之为局部自注意力的表述，然后将这两个洗牌操作与局部自注意力相结合，在同质化空间上建立一个精细的层计算。具有结构建模的FFN在第二节中被提出。3.3. 最后，我们在第2节中介绍了用于图像阴影去除的整体Transformer模型。3.4.

### 3.1. 随机洗牌和逆洗牌

阴影退化在整个空间中分布不均匀，这对于具有权重共享属性的主导模型来说是不希望的。为了解决非均匀性问题，我们提出了两个关键操作来均匀化分布：随机洗牌操作  $S(\cdot)$  和相应的逆洗牌操作  $I_S(\cdot)$ 。随机洗牌负责随机排列输入的元素，而逆洗牌对应于恢复原始顺序。形式上，假设  $X$  是输入， $m$  是  $X$  的索引列表， $m$  是  $m$  的随机排列，我们有随机洗牌操作的定义：

$$S(X)_m = X_m. \quad (1)$$

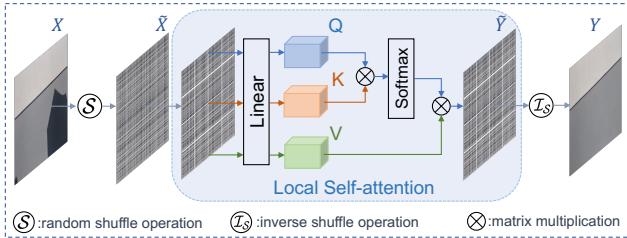


图 3. 所提出的随机洗牌局部自注意力的计算图。

经过随机洗牌后，每个像素出现在任意位置的概率相等。因此，随机洗牌可以在均匀化非均匀分布的阴影退化方面发挥关键作用，消除权值共享模型的约束。另一方面，随机洗牌彻底破坏了与像素顺序密切相关的图像语义。因此，在均匀化空间中拥有后，需要反转洗牌投影以与原始特征对齐。受此启发，我们详细阐述了逆洗牌操作，其定义为：

$$\mathcal{I}_S(\mathcal{S}(X)) = X. \quad (2)$$

如上所述，逆洗牌操作能够抵消随机洗牌操作的随机重新排序。此外，由于仅涉及元素的重新排列，洗牌操作对在现代加速器上实现起来非常高效，不会产生额外的参数或 FLOP。

### 3.2. 具有随机洗牌的局部自注意力

自注意力[43]可以概括为将查询和一组键值对映射到输出，其中查询、键和值是从输入的线性投影获得的。自注意力的公式表示为

$$\left( \frac{XW^Q(XW^K)^T}{\sqrt{d_k}} \right) XW^V. \quad (3)$$

$W^Q$ 、 $W^K$  和  $W^V$  分别是查询、键和值的参数矩阵。自注意力通常用于模拟远程交互。然而，它的时间和内存复杂度与令牌数（视觉任务的输入图像的分辨率）成二次方。当将自注意力直接应用于图像阴影时，二次复杂度会在计算和内存成本方面带来巨大的负担。去除，因为输入分辨率通常很高。Swin Transformer [34]采用局部自注意力代替全局自注意力，这显着降低了输入分辨率的线性复杂度。具体来说，局部注意力首先对输入进行分区

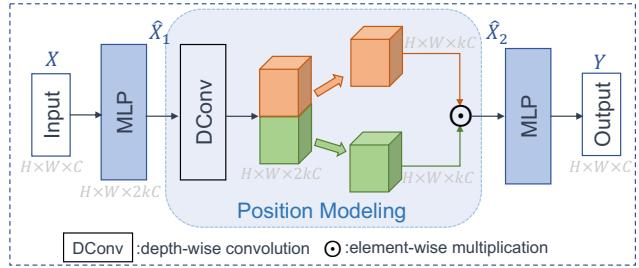


图 4. 所提出的 SMLP 的计算图。

进入非重叠窗口，并将自注意力的计算限制在本地窗口内。用于处理那些非重叠窗口的权重是共享的。局部自注意力的数学表述是

$$\text{LSA}(X) = \text{SA}(\text{Par}(X)), \quad (4)$$

其中  $\text{Par}(\cdot)$  表示配分函数。自注意力可以作为多头版本来实现，以提高其表达能力[10,34,43]。为了简单起见，我们在这里以单头情况为例，但不失一般性。尽管局部自注意力的效率很高，但由于阴影的空间不均匀性和多样性，其权重共享特性使其不利于图像去阴影。考虑到随机洗牌和逆洗牌操作，我们可以通过将洗牌对与局部自注意力相结合来克服这个困难。图3给出了详细的计算图。具体来说，输入  $X$  首先通过随机洗牌操作  $S(\cdot)$  随机重新排列，产生由  $X$  表示的输入的同质化版本。均质化输入被馈送到局部自注意力，产生均质化输出  $\tilde{Y}$ 。最后，通过逆洗牌操作  $I_S(\cdot)$  将均质化输出恢复到最终输出  $Y$  的原始顺序。上述程序的制定是

$$\tilde{X} = \mathcal{S}(X), \tilde{Y} = \text{LSA}(\tilde{X}), Y = \mathcal{I}_S(\tilde{Y}). \quad (5)$$

一个潜在的担忧可能是随机洗牌带来的像素位置的随机性是否会影响局部自注意力的训练稳定性。幸运的是，自注意力的一个重要属性是它与重新排序是等变的[7]，也就是说，无论输入标记如何打乱，它都会给出相同的输出。因此，自注意力可以作为处理同质化特征的所需层，而不受随机洗牌的随机性干扰。与原始局部自注意力 [34] 的区别在于，相对位置编码被丢弃，留下未探索的结构信息 [2]。原因很简单：随机洗牌后，位置不再可靠。但是这个问题可以通过将基于结构的交互建模的功能转移到后续的前馈网络来轻松解决。

### 3.3. FFN 与结构建模

遵循 Transformer 的经典范例，我们希望通过采用这种定制的自注意力和前馈网络（FFN）来设计基本的 Transformer 构建块。考虑到我们定制的自注意力无法利用基于结构的信息来探索结构信息，我们将建模基于结构的交互的责任转移到了 FFN。通常，Transformer 中的 FFN 由两个多层感知器（MLP）实现。我们在这里采用 FFN 来建模基于结构的交互。提醒一下，CNN 根据相对位置分配权重，从而充当一个简单的层来建模基于结构的交互。因此，我们通过在 MLP 之间插入深度卷积来设计一个简单的 FFN，我们称之为 SMLP。图 4 显示了该过程。形式上，SMLP 计算为

$$\begin{aligned}\hat{X}_1 &= \text{MLP}(X), \\ \hat{X}_2 &= \text{DConv}_1(\hat{X}_1) \odot \text{DConv}_2(\hat{X}_1), \\ Y &= \text{MLP}(\hat{X}_2),\end{aligned}\quad (6)$$

其中  $\text{DConv}_{1,2}(\cdot)$  表示深度卷积， $\odot$  表示元素乘法。请注意，我们放弃了 GELU 激活，因为它在处理元素乘法时是多余的 [5]，这也通过我们的消融研究得到了验证。

### 3.4. HomoFormer

考虑到具有随机洗牌和结构 MLP 的局部自注意力的基本块，我们现在准备通过将基本构建块集成到广泛使用的 UNet [22, 39] 架构中来构建最终的 Transformer（即 HomoFormer）。这个过程很简单，整体架构如图 2 所示。训练 HomoFormer 时采用的损失函数是单一 Charbonnier 损失 [3]，而不是复杂的混合损失 [13, 57]，其数学表达式为

$$L(I', I) = \sqrt{\|I' - I\|^2 + \epsilon^2}, \quad (7)$$

其中  $I'$  和  $I$  分别是输出图像和无阴影图像。为了数值稳定性，常数  $\epsilon$  根据经验设置为  $10^{-3}$ 。

## 4. 实验

### 4.1. 实验设置

随机因素。由于随机洗牌操作引入了随机性，因此我们运行评估五次，计算定量得分，并报告平均得分 1，其中

<sup>1</sup>实际上，标准差远小于报告平均值的精度。因此，我们忽略了标准差。

是默认配置（由“HomoFormer”表表示）。此外，考虑到这些随机因素，模型期望将其输出平均作为最终预测，即从贝叶斯角度边缘化随机因素。然而，随机洗牌操作对于每个自注意力层来说是独立的，从而导致组合的指数数量。因此，我们使用蒙特卡罗平均来近似预期预测。蒙特卡罗样本数设置为 8，表中用“HomoFormer+”标记。我们将进一步研究消融研究中样本数量的影响。

数据集。在两个代表性基准数据集上进行了阴影去除实验。(i) AdjustedISTD (ISTD+) 数据集[26]由 1870 个图像三元组（阴影图像、无阴影图像和阴影掩模）组成，分为 1330 个训练三元组和 540 个测试三元组。与 ISTD 数据集[45]相比，ISTD+ 数据集[26]通过图像处理算法减少了 ISTD 的阴影和无阴影图像之间的光照不一致。由于数据重复，我们将 ISTD 数据集的评估结果移至补充材料；(ii) SRD 数据集[38]由 2680 个训练对和 408 个测试对组成。由于 SRD 中缺乏 groundtruth shadowmask，我们直接利用 DHAN[8] 提供的公共 SRD shadowmask 进行训练和测试阶段。

评估指标。为了与其他方法进行定量比较，按照之前的方法[13, 18, 23, 45]，我们利用估计图像和真实无阴影图像之间的 LAB 颜色空间中的均方根误差 (MAE)。对于 MAE 度量，较低的值意味着更忠实的恢复，从而更好的结果。此外，我们还采用经典的峰值信噪比 (PSNR) 和结构相似性 (SSIM) 标准来评估各种方法在 RGB 空间中的性能。对于 PSNR 和 SSIM 指标，值越高表示结果越好。为了保持一致的比较，我们将估计的无阴影图像的尺寸调整为  $256 \times 256$  的分辨率以获得定量结果。

### 4.2. ISTD+ 数据集上的比较

我们在选项卡中报告了 ISTD+ 数据集 [26] 上的 MAE 分数与其他最先进的方法的比较。2. 包括 12 种以前的 SOTA 方法，范围从传统的阴影去除方法：Guo 等人。[18]，最近基于深度学习的方法：DeshadowNet [38]、ST-CGAN [54]、ShadowGAN [21]、SP+M-Net [26]、Param+M+D-Net [27]、G2R [35]，Fu 等人。[13]，金等人。[23]、BMNet [57]、SG-ShadowNet [44] 和 Shad-owFormer [15]。为了保证公平比较，这些比较方法的结果由作者提供或

表 1. 在 SRD 数据集上与 SOTA 方法的定量比较 [38]。最好和第二的结果分别用粗体和下划线表示。

Method	Shadow Region			Non-Shadow Region			All Region		
	PSNR ↑	SSIM ↑	MAE ↓	PSNR ↑	SSIM ↑	MAE ↓	PSNR ↑	SSIM ↑	MAE ↓
Input images	18.96	0.871	36.69	31.47	0.975	4.83	18.19	0.829	14.05
Guo <i>et al.</i> [18] (TPAMI'12)	-	-	29.89	-	-	6.47	-	-	12.60
DeshadowNet [38] (CVPR'17)	-	-	11.78	-	-	4.84	-	-	6.64
DSC [20] (TPAMI'19)	30.65	0.960	8.62	31.94	0.965	4.41	27.76	0.903	5.71
DHAN [8] (AAAI'20)	33.67	0.978	8.94	34.79	0.979	4.80	30.51	0.949	5.67
Fu <i>et al.</i> [13] (CVPR'21)	32.26	0.966	8.55	31.87	0.945	5.74	28.40	0.893	6.50
Jin <i>et al.</i> [23] (ICCV'21)	34.00	0.975	7.70	35.53	0.981	3.65	31.53	0.955	4.65
BMNet [57] (CVPR'22)	35.05	0.981	6.61	36.02	0.982	3.61	31.69	0.956	4.46
SG-ShadowNet [44] (ECCV'22)	-	-	7.53	-	-	2.97	-	-	4.23
ShadowFormer [15] (AAAI'23)	36.91	<b>0.989</b>	5.90	36.22	<b>0.989</b>	3.44	32.90	0.958	4.04
ShadowDiffusion [16] (CVPR'23)	38.72	<u>0.987</u>	4.98	37.78	0.985	3.44	34.73	0.970	3.63
Li <i>et al.</i> [30] (ICCV'23)	<b>39.33</b>	0.985	6.09	35.61	0.967	2.97	33.17	0.941	3.83
HomoFormer(ours)	<u>38.81</u>	<u>0.987</u>	<b>4.25</b>	<u>39.45</u>	<u>0.988</u>	<u>2.85</u>	<u>35.37</u>	<u>0.972</u>	<u>3.33</u>
HomoFormer+(ours)	38.64	0.987	<u>4.33</u>	<b>40.04</b>	<b>0.989</b>	<b>2.76</b>	<b>35.50</b>	<b>0.972</b>	<b>3.29</b>

表 2. 在 ISTD+ 数据集上与 SOTA 方法的定量比较。最好和第二的结果分别用粗体和下划线表示。

Method	Region	Shadow	Non-Shadow	All
		MAE↓	MAE↓	MAE↓
Input images		40.2	2.6	8.5
Guo <i>et al.</i> [18] (TPAMI'12)		22.0	3.1	6.1
DeshadowNet [38] (CVPR'17)		15.9	6.0	7.6
ST-CGAN [54] (CVPR'18)		13.4	7.7	8.7
ShadowGAN [21] (ICCV'19)		12.4	4.0	5.3
SP+M-Net [26] (ICCV'19)		7.9	3.1	3.9
Param+M+D-Net [27] (ECCV'20)		9.7	3.0	4.0
G2R [35] (CVPR'21)		7.3	2.9	3.6
Fu <i>et al.</i> [13] (CVPR'21)		6.5	3.8	4.2
Jin <i>et al.</i> [23] (ICCV'21)		10.3	3.5	4.6
BMNet [57] (CVPR'22)		5.6	2.5	3.0
SG-ShadowNet [44] (ECCV'22)		5.9	2.9	3.4
ShadowFormer [15] (AAAI'23)		5.2	2.3	2.8
ShadowDiffusion [16] (CVPR'23)		4.9	2.3	2.7
Li <i>et al.</i> [30] (ICCV'23)		5.9	2.9	3.3
HomoFormer(ours)		<b>5.0</b>	<u>2.3</u>	<u>2.7</u>
HomoFormer+(ours)		5.0	<u>2.2</u>	<u>2.6</u>

表 3. ISTD+ 数据集的消融研究。

Method	Region	Shadow	Non-Shadow	All
		MAE↓	MAE↓	MAE↓
w/o random shuffle		6.4	2.5	3.0
w/o structure		5.6	2.5	2.9
Ours (default)		5.0	2.3	2.7

从原始论文中获得。如表所示。2，我们的 HomoFormer 比之前所有的 SOTA 方法获得了更低的 MAE 分数，例如与 BMNet [57] 相比，增益为 0.4，这表明我们的方法能够更忠实地恢复干净的图像。此外，通过蒙特卡罗平均，我们的方法 (HomoFormer+) ob-

在阴影和非阴影区域都获得了更好的结果。图 5 显示，与其他方法相比，我们的 Ho-moFormer 产生的结果具有较少的边界伪影。补充材料提供了 SBU 数据集 [28] 上的扩展结果，以进一步支持其泛化。

#### 4.3. SRD数据集比较

在表选项卡中。如图 1 所示，我们报告了 SRD 数据集上的 PSNR/SSIM/MAE 与其他 SOTA 方法的定量比较 [38]，包括 Guo 等人。[18]、De-shadowNet [38]、DSC [20]、DHAN [8]、Fu 等人。[13]，金等人。[23]、BMNet [57]、SG-ShadowNet [44]、Shadow-Former [15]、Li 等人。[30] 和阴影扩散 [16]。我们的方法还以最低的 MAE 和最高的 PSNR/SSIM 值实现了最佳的去阴影性能。与 ShadowFormer [15] 相比，我们的方法的 PSNR 值从 32.90 dB 提高到 35.37 dB。此外，我们还在图 6 中提供了视觉比较。我们可以观察到 HomoFormer 可以去除阴影，留下的伪影更少。

#### 4.4. 消融研究与分析

为了验证我们的选择并进一步促进对我们方法的理解，我们基于 ISTD+ 数据集进行消融实验和分析 [26]。

随机洗牌的效果。为了研究 randomshuffle 的效果，我们通过删除 randomshuffle 和 inverse shuffle 设计了一个模型变体。如表所示。如图 3 所示，MAE 在阴影 (+1.4)、非阴影 (+0.2) 和所有区域 (+0.3) 上增加，这表明随机洗牌有助于清楚地消除阴影退化并忠实地维护非阴影区域。根本原因是随机洗牌使非均匀分布均匀化。

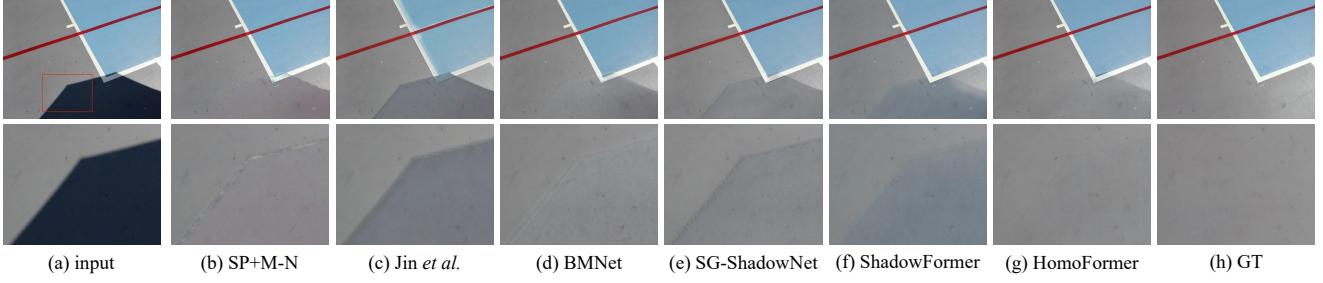


图 5. 在 ISTD+ 数据集上与最先进方法的视觉比较 [26]。第一行：全尺寸评估。第二行：放大区域。我们的 HomoFormer 能够以更少的伪影获得视觉上令人愉悦的结果。

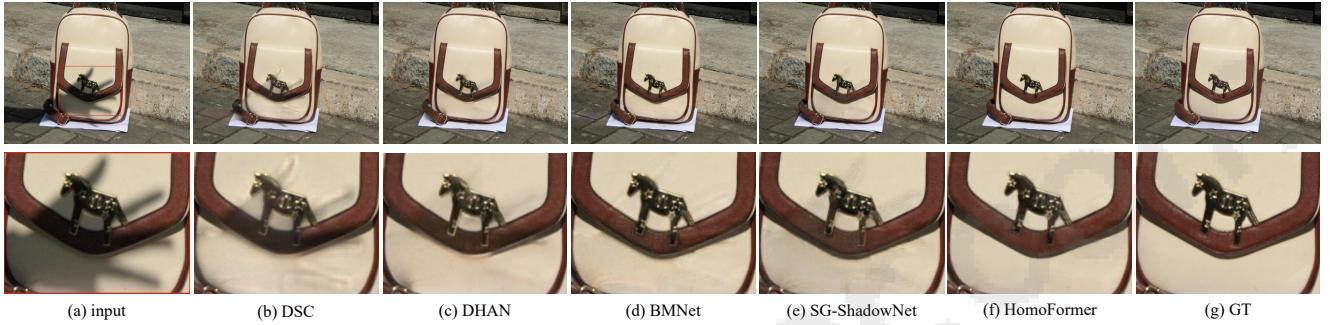


图 6. 在 SRD 数据集上与最先进方法的视觉比较 [38]。

退化的影响，这有利于权重共享局部自注意力。

为了进一步理解随机洗牌的行为，我们将默认的 HomoFormer 和没有随机洗牌的变体的中间特征可视化。图7示出了第一编码层和最后解码层的一些代表性示例。一般来说，在没有随机洗牌的情况下，局部自注意力必须利用单个参数集来在不同阴影程度的区域之间达成折衷（见图 1）。我们可以从实验中观察到这种效果：对于第一个编码层（图 7 中的第 2-3 列）的特征，突出显示的区域在没有随机洗牌的情况下丢弃了几乎所有纹理，这对于恢复忠实的结果是有害的。对于最后一个解码层的特征，我们希望它们尽可能干净，即它们不应该包含退化带来的伪影，因为它们将用于构造最终的干净输出。观察最后一个解码层的特征，我们发现具有随机洗牌的 HomoFormer 会产生更加视觉上令人愉悦的结果。这些观察结果共同得出这样的结论：随机洗牌有利于提取更有效的特征（在我们的例子中更多的纹理或更少的伪影）。不确定性可以预测错误。不确定性对于计算机视觉可以发挥重要作用[25]。对于阴影重新

移动、不确定性可以估计预测的置信度。由于其固有的随机洗牌行为，所提出的 HomoFormer 提提供了一种自然的方法来估计其不确定性。例如，我们可以多次评估图像并计算标准偏差作为不确定性。图9表明，不依赖真实图像的不确定性计算可以预测哪里容易发生错误，这对于现实场景具有实际意义。FFN 中结构建模的效果。由于我们丢弃了局部自注意力中的结构信息，因此我们将建模结构信息的责任转移到 FFN。我们通过删除 FFN 中的深度卷积来验证有效性。标签。图 3 表明位置建模的缺失会显着损害性能。蒙特卡罗样本数量的影响。为了边缘化随机因素，我们利用蒙特卡罗平均来近似期望。理论上，随着采样数趋于无穷大，平均值逐渐接近真实期望。我们在 SRDdataset 上研究了这个属性。图8表明，随着样本数量的增加，性能先提升后收敛。我们采用的默认采样数 8 可以产生有希望的结果，同时尽可能节省计算量。

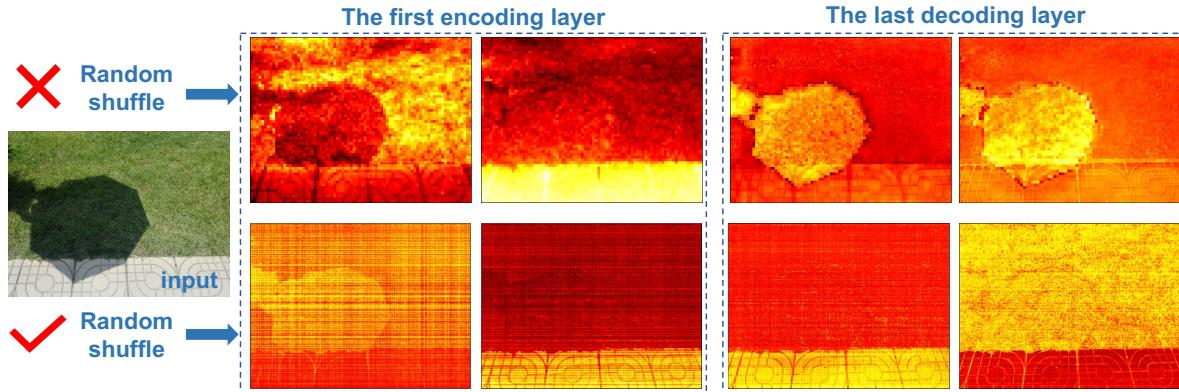


图 7.第一个编码层和最后一个解码层的特征可视化。第一行的特征取自没有随机洗牌的变体模型，第二行对应于默认 HomoFormer 模型的特征。与没有随机洗牌的变体相比，默认的 HomoFormer 的特征在浅层（第 2-3 列）中包含更详细的纹理，在深层（第 4-5 列）中包含更忠实的内容。

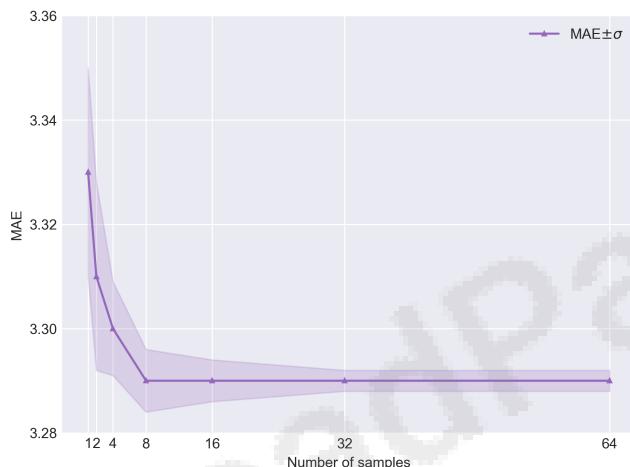


图 8. 蒙特卡洛样本数量的影响。

## 5. 讨论和未来的工作

在本文中，我们将可逆洗牌操作与局部自注意力相结合，为建模具有非均匀分布和多样化模式的复杂阴影退化的挑战提供了新的视角。期待与权重共享兼容的动机，HomoFormer的有效性可以从全局交互的角度来解释。我们假设图像内的像素是相关的。随机洗牌可以将两个像素拉入单个窗口，无论它们的距离如何。因此，混淆图像上的局部 SA 相当于从该图像中捕获稀疏的全局交互，为模型学习提供丰富的信息。此外，由于非均匀性和多样性并不是阴影退化所独有的，因此我们也对更一般的图像恢复任务中同质化的未来感到兴奋，例如图像修复，这也是

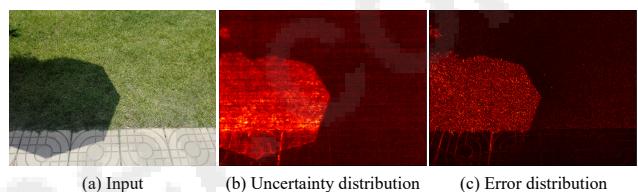


图 9. 源自随机洗牌的不确定性可用于预测评估的干净图像  $I'$  和无阴影图像  $I$  之间的误差分布（即  $|I - I'|$ ）。

我们未来的工作。

## 六，结论

在本文中，我们提供了一个新的视角来解决具有不均匀分布和多样化模式的复杂阴影退化建模问题。通过精心设计的随机洗牌和逆洗牌操作对，使非均匀分布均匀化，为利用权重共享模型有效地建模复杂退化奠定了基础。在此基础上，我们建立了一种新颖的基于局部窗口的变压器，名为 Homo-Former，用于图像阴影去除。我们的 HomoFormer 可以享受输入分辨率的高效线性复杂性，并克服建模非均匀分布阴影退化的挑战。我们进行了广泛而全面的实验来验证和理解所提出的方法。

## Acknowledgement

这项工作得到了国家自然科学基金委 (NSFC) 的资助，编号为 62225207 和 62276243。

## References

[1] Masashi Baba、Masayuki Mukunoki 和 Naoki Asada。基于阴影密度的真实图像中的阴影去除。在 ACM SIGGRAPH 2004 海报中。2004年.[2]伊尔万·贝洛。Lambda networks：无需注意即可对远程交互进行建模。ICLR，2021 年。[3]皮埃尔·夏博尼耶、劳尔·布兰克-费罗、吉尔·奥贝尔和米歇尔·巴洛。用于计算成像的两种确定性半二次正则化算法。ICIP，1994 年。[4]陈汉庭、王云鹤、郭天宇、徐昌、邓一平、刘振华、马思维、徐春静、徐超和高文。预先训练的图像处理变压器。在 CVPR，2021 年。[5]陈良宇，褚晓杰，张翔宇，孙健。图像恢复的简单基线。在 ECCV 中，2022 年。[6]陈子培、龙成江、张岭、肖春霞。Canet：用于阴影去除的上下文感知网络。In ICCV，2021。[7]让-巴蒂斯特·科多尼埃、安德烈亚斯·卢卡斯和马丁·贾吉。关于自注意力和卷积层之间的关系。ICLR，2020 年。[8]村晓东，潘志文，石成。通过双层次聚合网络和阴影抠图实现无重影去除。在 AAAI，2020 年。[9] Bin Ding, Chengjian Long, Ling Chang, and Chunxia Shaw。Argan：用于阴影检测和去除的注意力循环生成对抗网络。在 ICCV，2019 年。[10] Alexey Dosovitskiy、Lucas Beyer、Alexander Kolesnikov、Dirk Weissenborn、Zhai Hua、Thomas Unterthiner、Mostafa Dehghani、Matthias Minderer、Georg Heigold、Syl-vain Gelly、Jakob Uszkoreit 和 Neil Houlsby。一张图像相当于 16x16 个单词：用于大规模图像识别的 Transformers。ICLR，2021 年。[11]格雷厄姆·D·芬利森、史蒂文·D·霍德利、程璐和马克·S·德鲁。关于图像中阴影的去除。TPAMI，2005。[12]格雷厄姆·D·芬利森、马克·S·德鲁和程璐。用于阴影去除的熵最小化。国际 JCV，2009。[13]付兰，周长青，郭庆，徐觉飞，于洪凯，冯伟，刘阳，王松。用于单图像阴影去除的自动曝光融合。CVPR，2021。[14]韩工和达伦·科斯克。针对困难的阴影场景进行交互式去除和地面实况。乔萨 A，2016。[15]郭兰清、黄思雨、刘丁、程浩、文碧涵。Shadowformer：全局上下文有助于图像阴影去除。在 AAAI，2023 年。[16]郭兰清、王冲、杨文涵、黄思雨、王雨菲、汉斯彼得·菲斯特和文碧涵。Shadowd-iffusion：当退化先验满足用于阴影去除的扩散模型时。在 CVPR，2023 年。[17]郭瑞琪、戴切云和德里克·霍伊姆。用于阴影检测和去除的配对区域。TPAMI，2012。[18]郭瑞琪、戴切云和德里克·霍伊姆。用于阴影检测和去除的配对区域。TPAMI，2013。

[19] 胡晓伟，朱雷，傅志荣，秦静，恒鹏安。用于阴影检测的方向感知空间上下文特征。CVPR，2018。[20]胡晓伟、傅志荣、朱磊、秦静和恒鹏安。用于阴影检测和去除的方向感知空间上下文功能。TPAMI，2019。[21]胡晓伟、姜一桐、傅志荣和彭安恒。Mask-shadowgan：学习从不配对的数据中删除阴影。在 ICCV，2019 年。[22]菲利普·伊索拉、朱俊彦、周廷辉和阿列克谢·埃夫罗斯。使用条件对抗网络进行图像到图像的翻译。CVPR，2017。[23] Yeying Jin、Aashish Sharma 和 Robby T Tan。DC-shadownet：使用无监督域分类器引导网络进行单图像硬阴影和软阴影去除。在 ICCV 中，2021 年。[24]康国良，董宣仪，梁正，易阳。Patchshuffle 正则化。arXiv preprint arXiv:1707.07103, 2017。[25]亚历克斯·肯德尔和亚林·加尔。计算机视觉的贝叶斯深度学习需要哪些不确定性？见 NeurIPS，2017。[26] Hieu Le 和迪米特里斯·萨马拉斯。通过阴影图像分解去除阴影。在 ICCV，2019 年。[27] Hieu Le 和迪米特里斯·萨马拉斯。从阴影分割到阴影去除。在 ECCV，2020 年。[28] Hieu Le 和迪米特里斯·萨马拉斯。基于物理的阴影图像分解用于阴影去除。TPAMI，2022。[29]李宇石、孙相贤、李景武。Ap-bsn：通过非对称 pd 和盲点网络对真实世界图像进行自监督去噪。CVPR，2022 年。[30]李晓光、郭清、Rabab Abdelfattah、林迪、伟峰、Ivor Tsang 和王松。利用修复来消除单图像阴影。2023。[31]梁景云、曹杰章、孙国磊、张凯、LucVan Gool 和 Radu Timofte。Swinir：使用 swin 变压器进行图像恢复。ICCV 研讨会，2021 年。[32]林静、蔡元浩、胡小万、王浩谦、严友亮、邹学一、丁恒辉、张玉伦、RaduTimofte 和 Luc Van Gool。用于视频去模糊的流引导稀疏变压器。在 ICML，2022 年。[33]刘大全、龙成江、张红攀、于汉宁、董新志、肖春霞。Arshadowgan：用于单光场景中增强现实的阴影生成对抗网络。在 CVPR，2020 年。[34]刘泽、林雨桐、曹悦、胡涵、魏艺轩、张正、林史蒂芬和郭百宁。Swin 变压器：使用移动窗口的分层视觉变压器。在 ICCV 中，2021 年。[35]刘志浩、尹慧、吴欣怡、吴振耀、杨幂和王松。从阴影生成到阴影消除。在 CVPR，2021 年。[36] S.纳迪米和B.巴努。视频中移动阴影和对象检测的物理模型。TPAMI，2004。[37] Tam Minh Nguyen、Tan Minh Nguyen、Dung DD Le、Duy Khuong Nguyen、Viet-Anh Tran、Richard Baraniuk、Nhat Ho 和 Stanley Osher。使用概率注意键改进变压器。在 ICML，2022 年。

[38]曲良琼, 田建东, 何胜峰, 唐彦东, 刘伟华。Deshadownet: 用于阴影去除的多上下文嵌入深度网络。CVPR, 2017。[39]奥拉夫·罗纳伯格、菲利普·费舍尔和托马斯·布洛克斯。U-net: 用于生物医学图像分割的卷积网络。在 MICCAI, 2015 年。[40]安德烈斯·萨宁、康拉德·桑德森和布莱恩·C·洛弗尔。改进了阴影去除功能, 可在监控场景中实现稳健的人员跟踪。ICPR, 2010 年。[41]史文哲、Jose Caballero、Ferenc Huszár、Johannes Totz、Andrew P Aitken、Rob Bishop、Daniel Rueckert 和 Zehan Wang。使用高效的子像素卷积神经网络进行实时单图像和视频超分辨率。在 CVPR, 2016 年。[42]雅埃尔·肖尔和丹尼·利钦斯基。阴影遇到遮罩: 基于金字塔的阴影去除。在计算机图形学论坛中。Wiley 在线图书馆, 2008 年。[43]Ashish Vaswani、Noam Shazeer、Niki Parmar、Jakob Uszkoreit、Llion Jones、Aidan N Gomez、Łukasz Kaiser 和 Illia Polosukhin。您所需要的就是关注。在 NeurIPS, 2017 年。[44]金万, 尹慧, 吴振耀, 吴欣怡, 刘彦婷, 王松。风格引导的阴影去除。在 ECCV, 2022 年。[45]王继峰, 李翔, 杨健。堆叠条件生成对抗网络, 用于联合学习阴影检测和阴影去除。CVPR, 2018。[46]王振东、村晓东、鲍建民、刘建壮。Uformer: 用于图像恢复的通用 U 形变压器。在 CVPR, 2022 年。[47]吴海旭、吴家龙、徐杰辉、王建民和龙明胜。Flowformer: 使用守恒流对变压器进行线性化。在 ICML, 2022 年。[48]肖春霞、余瑞云、肖东林和马关六。使用自适应多尺度照明传输快速去除阴影。在 CGF, 2013 年。[49]肖杰, 付雪阳, 吴峰, 查正军。用于图像恢复的随机窗口变换器。InNeurIPS, 2022。[50]肖杰、付雪阳、刘爱萍、吴凤和查正军。图像除雨变压器。帕米, 2023。[51]肖杰、付雪阳、周满、刘洪建和查正军。用于图像恢复的随机洗牌变压器。ICML, 2023。[52]Qingxiong Yang、Kar-Han Tan 和 Narendra Ahuja。使用双边滤波去除阴影。提示, 2012。[53]Syed Waqas Zamir、Aditya Arora、Salman Khan、Mu-nawar Hayat、Fahad Shahbaz Khan 和 Ming-Hsuan Yang。Restormer: 用于高分辨率图像恢复的高效转换器。CVPR, 2022 年。[54]张宏光、戴玉超、李宏东、Piotr Ko-niusz。用于图像去模糊的深堆叠分层多补丁网络。在 CVPR, 2019 年。[55]张凌、张清、肖春霞。Shadow remover: 基于光照恢复优化的图像阴影去除。提示, 2015。[56]张武明、赵曦、Jean-Marie Morvan 和陈黎明。改善照明鲁棒人脸识别的阴影抑制。TPAMI, 2018。

- [57] Yurui Zhu, Jie Huang, Xueyang Fu, Feng Zhao, Qibin Sun, and Zheng-Jun Zha. Bijective mapping network for shadow removal. In CVPR, 2022.
- [58] Yurui Zhu, Zeyu Xiao, Yanchi Fang, Xueyang Fu, Zhiwei Xiong, and Zheng-Jun Zha. Efficient model-driven network for shadow removal. In AAAI, 2022.