

MambaIR: 状态空间模型图像恢复的简单基线

Hang Guo^{1,4,*}, Jinmin Li^{1,*}, Tao Dai^{2,†},
Zhihao Ouyang^{3,4}, Xudong Ren¹, and Shu-Tao Xia^{1,5}

1 清华大学深圳国际研究生院
2 深圳大学计算机科学与软件工程学院
3 字节跳动公司
4 Aitist.ai
5 鹏程实验室
[{cshguo, daitao.edu}@gmail.com,](mailto:{cshguo, daitao.edu}@gmail.com)
{ljm22, rxd21}@mails.tsinghua.edu.cn,
xiaast@sz.tsinghua.edu.cn

抽象的。近年来，图像恢复领域取得了重大进展，这很大程度上归功于现代深度神经网络（例如 CNN 和 Transformer）的发展。然而，现有的恢复骨干经常面临全局感受野和高效计算之间的困境，阻碍了其在实践中的应用。近年来，选择性结构化状态空间模型，特别是改进版本的Mamba，在线性复杂度的远程依赖建模方面表现出了巨大的潜力，为解决上述困境提供了一种方法。然而，标准的Mamba在低级视觉方面仍然面临一定的挑战，例如如局部像素遗忘和通道冗余。在这项工作中，我们引入了一个简单但有效的基线，名为 MambaIR，它引入了局部增强和通道注意力来改进普通 Mamba。通过这种方式，我们的 MambaIR 利用了局部像素相似性并减少了通道冗余。大量实验证明了我们方法的优越性，例如，MambaIR 在图像 SR 上的性能比 SwinIR 高出 0.45dB，使用相似的计算成本但具有全局感受野。代码可在 <https://github.com/csguoh/MambaIR> 获取。

关键词：图像恢复·状态空间模型·Mamba

1 简介

图像恢复旨在从给定的低质量输入中重建高质量图像，是计算机视觉中长期存在的问题，并且还具有广泛的子问题，例如超分辨率、图像去噪等。CNN [13,16,42,81,89] 和 Transformers [8,10,12,40,41] 等现代深度学习模型的最新性能在过去几年中不断刷新。

* 等贡献 † 通讯作者：戴涛
(daitao.edu@gmail.com)

†Corresponding author: Tao Dai (daitao.edu@gmail.com)

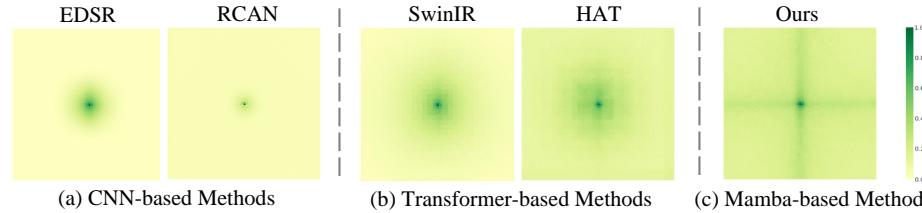


图 1: EDSR [42]、RCAN [88]、SwinIR [41]、HAT [10] 和建议的 MambaIR 的有效感受野 (ERF) 可视化 [14, 46]。ERF 越大，暗区分布越广泛。所提出的 MambaIR 实现了显着的全局有效感受野。

在某种程度上，深度恢复模型性能的提高很大程度上源于网络感受野的增加。首先，大的感受野允许网络从更广泛的区域捕获信息，使其能够参考更多的像素以促进锚像素的重建。其次，具有更大的感受野，恢复网络可以提取图像中更高级别的模式和结构，这对于图像去噪等一些结构保留任务至关重要。最后，基于 Transformer 的恢复方法具有更大的感受野，在实验上优于基于 CNN 的方法，并且最近的工作 [10] 也指出激活更多的像素通常会带来更好的恢复结果。尽管具有许多吸引人的特性，但似乎存在当前图像恢复主干的全局感受野和高效计算之间存在固有的选择困境。对于基于 CNN 的恢复网络 [42, 89]，虽然有效感受野有限（如图 1 (a) 所示），但由于卷积并行运算的良好效率，它适合资源受限的设备部署。相比之下，基于 Transformer 的图像恢复方法通常将 token 的数量设置为图像分辨率 [8, 10, 41]，因此，尽管有全局感受野，直接使用标准 Transformer [65] 将会得到不可接受的结果。能够实现二次计算复杂度。此外，采用一些有效的注意力技术，例如用于图像恢复的移位窗口注意力 [45]，通常会以牺牲全局有效的感受野为代价（如图 1 (b) 所示），并且本质上不会逃脱交易。最近，结构化状态空间序列模型 (S4)，特别是改进版本的 Mamba，已成为构建深度网络的高效骨干 [18, 22, 24, 52, 62]。这一进展暗示了平衡图像恢复中的全局感受野和计算效率的潜在解决方案。具体来说，Mamba 中的离散状态空间方程可以形式化为递归形式，并且在配备专门设计的结构化重新参数化时可以对非常长范围的依赖性进行建模 [23]。这意味着基于 Mamba 的恢复网络可以自然地激活更多

This means that Mamba-based restoration networks can naturally activate more

像素，从而提高重建质量。此外，并行扫描算法 [22] 使 Mamba 以并行方式处理每个令牌，有助于在 GPU 等现代硬件上进行高效训练。上述有前景的特性促使我们探索 Mamba 实现高效的图像恢复网络远程建模的潜力。然而，为 NLP 中的一维序列数据设计的标准 Mamba [22] 并不适合图像恢复场景。首先，由于 Mamba 以递归方式处理扁平化的一维图像序列，因此可能会导致在扁平化序列中非常遥远的位置找到空间上接近的像素，从而导致局部像素遗忘的问题。其次，由于需要记住长序列依赖关系，状态空间方程中的隐藏状态数量通常很大，这可能导致通道冗余，从而阻碍关键通道表示的学习。为了解决上述挑战，我们引入了 MambaIR，一个简单但非常有效的基准模型，用于适应 Mamba 进行图像恢复。MambaIR 分为三个主要阶段。具体来说，1) 浅层特征提取阶段采用简单的卷积层来提取浅层特征。然后2) 深度特征提取阶段使用几个堆叠的残余状态空间块 (RSSB) 来执行。作为我们 MambaIR 的核心组件，RSSB 设计有局部卷积，以减轻将普通 Mamba 应用于 2D 图像时的局部像素遗忘，并且还配备了通道注意功能，以减少因隐藏状态数过多而导致的通道冗余。我们还使用可学习因子来控制每个 RSSB 内的跳跃连接。最后，3) 高质量图像重建阶段聚合浅层和深层特征以产生高质量的输出图像。通过拥有全局有效感受野和线性计算复杂性，我们的 MambaIR 成为图像恢复主干的新替代方案。简而言之，我们的主要贡献可以总结如下：

- 我们是第一个通过大量实验制定 MambaIR 来调整状态空间模型以进行低级图像恢复的工作，它是基于 CNN 和 Transformer 的方法的简单但有效的替代方案。
- 我们提出了残差状态空间块 (RSSB)，它可以通过局部增强和通道冗余减少来增强标准 Mamba 的功能。
- 对各种任务的大量实验表明，我们的 MambaIR 优于其他强大的基线，为图像恢复提供强大且有前途的骨干解决方案。

2 相关工作

2.1 图像恢复

自从一些开创性的工作引入深度学习以来，图像恢复已经取得了显着的进步，例如用于图像超级的 SRCNN [16] learning by several pioneering works, such as SRCNN [16] for image super-

分辨率，用于图像去噪的 DnCNN [81]，用于减少 JPEG 压缩伪影的 ARCNN [15] 等。早期的尝试通常使用残差连接 [6, 34]、密集连接 [68, 89] 等技术来详细阐述 CNN。等人[13,19,36,70]提高模型表示能力。尽管取得了成功，但基于 CNN 的恢复方法通常在有效建模全局依赖性方面面临挑战。由于 Transformer 已证明其在多个任务中的有效性，例如时间序列 [43]、3D 云 [75, 77] 和多模态 [4,20,21, 86]，因此使用 Transformer 进行图像恢复似乎很有前景。尽管具有全局感受野，Transformer 仍然面临着来自自注意力二次计算复杂性的特定挑战[65]。为了解决这个问题，IPT [8] 将一个图像分成几个小块，并通过自注意力独立处理每个块。SwinIR [41] 进一步引入了转移窗口注意力 [45] 以提高性能。此外，在设计有效的恢复注意力方面不断取得进展[9-12,26,38,63,72,78,85]。然而，高效的注意力设计通常是以牺牲全局感受野为代价的，高效计算与全局建模之间的权衡困境并未得到本质解决。

2.2 状态空间模型

状态空间模型（SSM）[24,25,62]源于经典控制理论[33]，最近被引入深度学习作为状态空间转换的竞争支柱。长程依赖建模中随序列长度线性缩放的有前景的特性引起了搜索者的极大兴趣。例如，结构化状态空间序列模型（S4）[24]是深度状态空间模型在建模远程依赖方面的开创性工作。后来，S5层[62]在S4的基础上提出，并引入了MIMO SSM和高效并行扫描。此外，H3 [18] 取得了有希望的结果，几乎填补了自然语言中 SSM 和 Transformers 之间的性能差距。[52]进一步改进S4与门控单元以获得门控状态空间层以提升能力。最近，Mamba [22]，一种具有选择性机制和高效硬件设计的数据依赖型 SSM，在自然语言方面优于 Transformer，并享有输入长度的线性缩放。此外，还有采用 Mamba 来完成图像分类 [44, 92]、视频理解 [37, 66]、生物医学图像分割 [48,71] 等视觉任务的开创性工作 [28,31,56,59,76]。在这项工作中，我们通过特定于修复的设计探索 Mamba 在图像修复方面的潜力，以作为未来工作的简单但有效的基线。

3 方法论

3.1 Preliminaries

结构化状态空间序列模型（S4）类的最新进展很大程度上受到连续线性时不变（LTI）系统的启发，
(S4) are largely inspired by the continuous linear time-invariant (LTI) systems,

它通过隐式潜在状态 $h(t) \in \mathbb{R}^N$ 映射一维函数或序列 $x(t) \in \mathbb{R} \rightarrow y(t) \in \mathbb{R}$ 。形式上，该系统可以表示为线性常微分方程 (ODE):

$$\begin{aligned} h'(t) &= \mathbf{A}h(t) + \mathbf{B}x(t), \\ y(t) &= \mathbf{C}h(t) + \mathbf{D}x(t), \end{aligned} \quad (1)$$

其中 N 是状态大小， $\mathbf{A} \in \mathbb{R}^{N \times N}$, $\mathbf{B} \in \mathbb{R}^{N \times 1}$, $\mathbf{C} \in \mathbb{R}^{1 \times N}$, $\mathbf{D} \in \mathbb{R}$ 。之后，通常采用离散化过程来积分方程。(1) 融入实用的深度学习算法。具体来说，将 Delta 表示为时间尺度参数，将连续参数 \mathbf{A} 、 \mathbf{B} 转换为离散参数 $\bar{\mathbf{A}}$ 、 $\bar{\mathbf{B}}$ 。常用的离散化方法是零阶保持 (ZOH) 规则，其定义如下：

$$\begin{aligned} \bar{\mathbf{A}} &= \exp(\Delta \mathbf{A}), \\ \bar{\mathbf{B}} &= (\Delta \mathbf{A})^{-1}(\exp(\mathbf{A}) - \mathbf{I}) \cdot \Delta \mathbf{B}. \end{aligned} \quad (2)$$

离散化后，得到方程的离散化版本。(1) 步长为 Delta 的式子可以重写为以下 RNN 形式：

$$\begin{aligned} h_k &= \bar{\mathbf{A}}h_{k-1} + \bar{\mathbf{B}}x_k, \\ y_k &= \mathbf{C}h_k + \mathbf{D}x_k. \end{aligned} \quad (3)$$

此外，方程。(3) 也可以在数学上等价地转化为如下 CNN 形式：

$$\begin{aligned} \bar{\mathbf{K}} &\triangleq (\mathbf{C}\bar{\mathbf{B}}, \mathbf{C}\bar{\mathbf{A}}\bar{\mathbf{B}}, \dots, \mathbf{C}\bar{\mathbf{A}}^{L-1}\bar{\mathbf{B}}), \\ \mathbf{y} &= \mathbf{x} \circledast \bar{\mathbf{K}}, \end{aligned} \quad (4)$$

其中 L 是输入序列的长度， \circledast 表示卷积运算， $\mathbf{K} \in \mathbb{R}^{L \times \text{输入通道数}}$ 是结构化卷积核。最近的先进状态空间模型 Mamba [22] 进一步改进了 \mathbf{B} 、 \mathbf{C} 和 Delta，使其与输入相关，从而允许动态特征表示。Mamba 对于图像修复的直觉在于它是在 S4 模型的优点上发展起来的。具体来说，Mamba 具有与方程式相同的递归形式。(3)，这使得模型能够记住超长序列，从而可以激活更多像素来帮助恢复。同时，并行扫描算法 [22] 使 Mamba 能够享受与等式 1 相同的并行处理优势。

(4)、从而促进高效的培训。

3.2 整体架构

如图 2 所示，我们的 MambaIR 由三个阶段组成：浅层特征提取、深层特征提取和高质量重建。给定低质量 (LQ) 输入图像 $I_{LQ} \in \mathbb{R}^{H \times W \times 3}$ ，我们首先使用浅层特征提取中的 3×3 卷积层来生成浅层特征

tion layer from the shallow feature extraction to generate the shallow feature

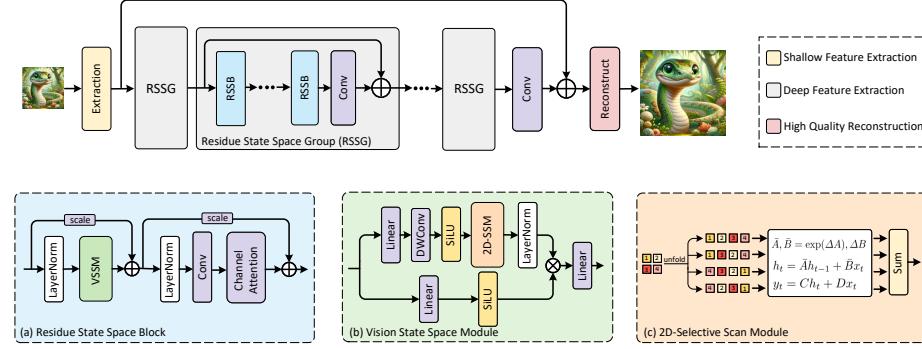


图2：我们的 MambaIR 的整体网络架构，以及 (a) 残余状态空间模块 (RSSB)、(b) 视觉状态空间模块 (VSSM) 和 (c) 2D 选择性扫描模块 (2D-SSM)。

$F_S \in R^{H \times W \times C}$ ，其中 H 和 W 表示输入图像的高度和宽度， C 是通道数。随后，浅层特征 F_S 经过深层特征提取阶段，获得第 l 层的深层特征 $F_{LD}^l \in R^{H \times W \times C}$, $l \in \{1, 2, \dots, L\}$ 。该阶段由多个剩余状态空间组 (RSSG) 堆叠而成，每个 RSSG 包含多个剩余状态空间块 (RSSB)。此外，在每组末尾引入一个额外的卷积层来细化从 RSSB 中提取的特征。最后，我们使用逐元素求和来获得高质量重建阶段的输入 $F_R = F_{LD} + F_S$ ，其中用于重建高质量 (HQ) 输出图像 I_{HQ} 。

3.3 剩余状态空间块

之前基于 Transformer 的恢复网络 [10, 12, 41, 78] 中的块设计主要遵循 $\text{Norm} \rightarrow \text{Attention} \rightarrow \text{Norm} \rightarrow \text{MLP}$ 流程。虽然 Attention 和 SSM 都可以建模全局依赖关系，但是，我们发现这两个模块的行为不同（更多细节请参阅补充材料），简单地用 SSM 替换 Attention 只能获得次优结果。因此，有希望定制一个全新的模型基于 Mamba 的恢复网络的块结构。为此，我们提出了剩余状态空间块 (RSSB) 来适应 SSM 块进行恢复。如图2 (a) 所示，给定输入深度特征 $F_{LD}^l \in R^{H \times W \times C}$ ，我们首先使用 LayerNorm (LN)，然后使用 Vision State-Space Module (VSSM) [44] 来捕获空间长。此外，我们还使用可学习的比例因子 $s \in R^C$ 来控制跳过连接的信息：

$$Z^l = \text{VSSM}(\text{LN}(F_{LD}^l)) + s \cdot F_{LD}^l. \quad (5)$$

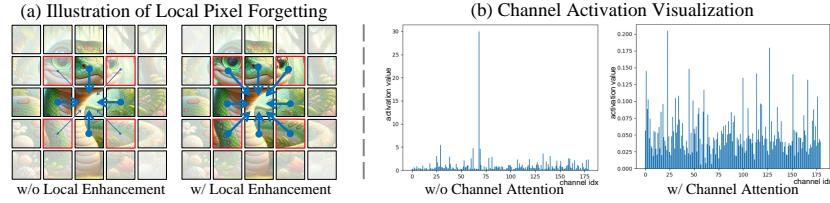


图3: (a) 不使用局部增强会导致空间上接近的像素（红色框中的区域）由于距离较长而在展平的一维序列中被遗忘。 (b) 我们对最后一层的 VSSM 输出使用 RELU 和全局平均池化来获取通道激活值。当不使用通道注意力时，大多数通道不会被激活（即通道冗余）。

此外，由于 SSM 将扁平化特征图处理为一维标记序列，因此序列中邻域像素的数量很大程度上受到扁平化策略的影响。例如，当采用[44]的四方向展开策略时，只有四个最近邻可用于锚像素（见图3(a)），即 2D 特征图中一些空间上接近的像素是在-相反，在一维标记序列中彼此远离，这种距离过大可能会导致局部像素遗忘。为此，我们在 VSSM 之后引入一个额外的局部卷积来帮助恢复邻域相似性。具体来说，我们使用 LayerNorm 首先对 Z^l 进行归一化，然后使用卷积层来补偿局部特征。为了保持效率，卷积层采用瓶颈结构，即首先对通道进行因子 γ 压缩以获得形状为 $R \times H \times W \times C_\gamma$ 的特征，然后进行通道扩展以恢复原始形状。此外，SSM 通常引入大量的隐藏状态来记忆非常长范围的依赖关系，我们在图 3(b) 中可视化不同通道的激活结果，并发现显着的通道冗余。为了增强不同通道的表达能力，我们将通道注意力 (CA) [27] 引入 RSSB。通过这种方式，SSM 可以专注于学习不同的通道表示，然后通过后续通道注意力来选择关键通道，从而避免通道冗余。最后，在残差连接中使用另一个可调比例因子 $s' \cdot \epsilon \cdot R \cdot C$ 得到最终输出 F^{l+1}

RSSB 的。上述过程可以表述为：

$$F_D^{l+1} = CA(\text{Conv}(\text{LN}(Z^l))) + s' \cdot Z^l. \quad (6)$$

3.4 视觉状态空间模块

为了保持效率，基于 Transformer 的恢复网络通常将输入划分为小块 [8] 或采用移位窗口注意力 [41]，从而阻碍了整个图像级别的交互。受 Mamba 在具有线性复杂性的远程建模中取得成功的启发，我们将 Vision State-SpaceModule 引入图像恢复。

Module to image restoration.

视觉状态空间模块 (VSSM) 可以利用状态空间方程捕获长程依赖性, VSSM的架构如图1所示。2(b)。按照[44], 输入特征 $X \in \mathbb{R}^{H \times W \times C}$ 将经过两个并行分支。在第一个分支中, 特征通道通过线性层扩展至 λC , 其中 λ 是预定义的通道扩展因子, 随后是深度卷积、SiLU [61]激活函数, 以及 2D-SSM 层和 LayerNorm。在第二个分支中, 特征通道也通过线性层扩展至 λC , 后跟 SiLU 激活函数。之后, 来自两个分支的特征与 Hadamard 积进行聚合。最后, 通道号被投影回 C 以生成与输入具有相同形状的输出 X :

$$\begin{aligned} X_1 &= \text{LN}(2\text{D-SSM}(\text{SiLU}(\text{DWConv}(\text{Linear}(X)))), \\ X_2 &= \text{SiLU}(\text{Linear}(X)), \\ X_{out} &= \text{Linear}(X_1 \odot X_2), \end{aligned} \quad (7)$$

其中DWConv表示深度卷积, \odot 表示Hadamard积。

3.5 2D选择性扫描模块

标准 Mamba [22] 对输入数据进行因果处理, 因此只能捕获数据扫描部分内的信息。此属性非常适合涉及顺序性质的 NLP 任务, 但在转移到图像等非因果数据时会带来重大挑战。为了更好地利用二维空间信息, 我们按照[44]引入了二维选择性扫描模块 (2D-SSM)。如图2 (c) 所示, 2D图像特征被展平为1D序列, 并沿四个不同方向扫描: 从左上到右下, 从右下到左上, 从右上到左下, 然后从左下到右上。然后根据离散状态空间方程捕获每个序列的长程依赖性。最后, 使用求和来合并所有序列, 然后进行重塑操作以恢复二维结构。

3.6 损失函数

为了与之前的作品[41,78,89]进行公平的比较, 我们用 L 1 损失优化我们的 MambaIR 图像 SR, 可以表示为:

$$\mathcal{L} = \|I_{HQ} - I_{LQ}\|_1, \quad (8)$$

其中 $\|\cdot\|_1$ 表示L 1 范数。对于图像去噪, 我们利用 Charbonnierloss [7], 其中 $\epsilon = 10^{-3}$:

$$\mathcal{L} = \sqrt{\|I_{HQ} - I_{LQ}\|^2 + \epsilon^2}. \quad (9)$$

4 经验

4.1 实验设置

数据集和评估。遵循之前的工作[41, 78]中的设置，我们对各种图像恢复任务进行了实验，包括图像超分辨率（即经典SR、轻量级SR、真实SR）和图像去噪（即高斯彩色图像去噪和真实去噪）世界去噪）和JPEG压缩伪影减少（JPEG CAR）。我们使用DIV2K [64]和Flickr2K [42]来训练经典的SR模型，并使用DIV2K仅训练轻量级SR模型。此外，我们使用Set5 [5], Set14 [74], B100 [50], Urban100 [29], 和 Manga109 [51]评估不同SR方法的有效性。对于高斯彩色图像去噪，我们利用 DIV2K [64]、Flickr2K [42]、BSD500 [3] 和 WED [49] 作为训练数据集。我们的高斯彩色图像去噪测试数据集包括 BSD68 [50]、Kodak24 [17]、McMaster [84] 和 Urban100 [29]。对于真正的图像去噪，我们使用来自 SIDD [1] 数据集的 320 个高分辨率图像来训练我们的模型，并使用 SIDD 测试集和 DND [58] 数据集进行测试。按照 [41,89]，当 self 时，我们将模型表示为 MambaIR+ -ensemble strategy [42] 用于测试。使用 PSNR 和 SSIM 在 YCbCr 颜色空间的 Y 通道上评估性能。由于页数限制，JPEG CAR 的结果显示在补充材料《训练详情》中。根据之前的工作[10,41,78]，我们通过应用水平翻转和90°、180°和270°随机旋转来执行数据增强。此外，我们在训练期间将原始图像裁剪为 64×64 的图像块（用于 imageSR）和 128×128 的块（用于图像去噪）。对于图像SR，我们使用 $\times 2$ 模型的预训练权重来初始化 $\times 3$ 和 $\times 4$ 的权重，并将学习率和总训练迭代次数减半以减少训练时间[42]。为了确保公平比较，我们将图像 SR 的训练批量大小调整为 32，将图像去噪的训练批量大小调整为 16。我们使用 Adam [35] 作为优化器来训练我们的 MambaIR， $\beta_1 = 0.9$, $\beta_2 = 0.999$ 。初始学习率设置为 2×10^{-4} ，并且当训练迭代达到特定里程碑时减半。我们的 MambaIR 模型使用 8 个 NVIDIA V100 GPU 进行训练。

4.2 消融研究

RSSB 不同设计的效果。作为核心组件，RSSB 可以通过特定于恢复的先验来改进 Mamba。在本节中，我们将消除 RSSB 的不同组件。结果显示在表中。如图1所示，表明 (1) 在平坦图像上应用一维扫描会导致局部像素遗忘，而利用简单的卷积层可以有效增强局部交互。 (2) 在不使用额外的卷积和通道注意力的情况下，即直接使用现成的Mamba进行恢复，只能获得次优结果，这也支持了我们之前的分析。 (3) 将Conv +ChanelAttention替换为MLP，其结构会类似于Transformer，也会导致不利的结果，这表明虽然SSM和Attention都具有全局建模能力，但两者的行为

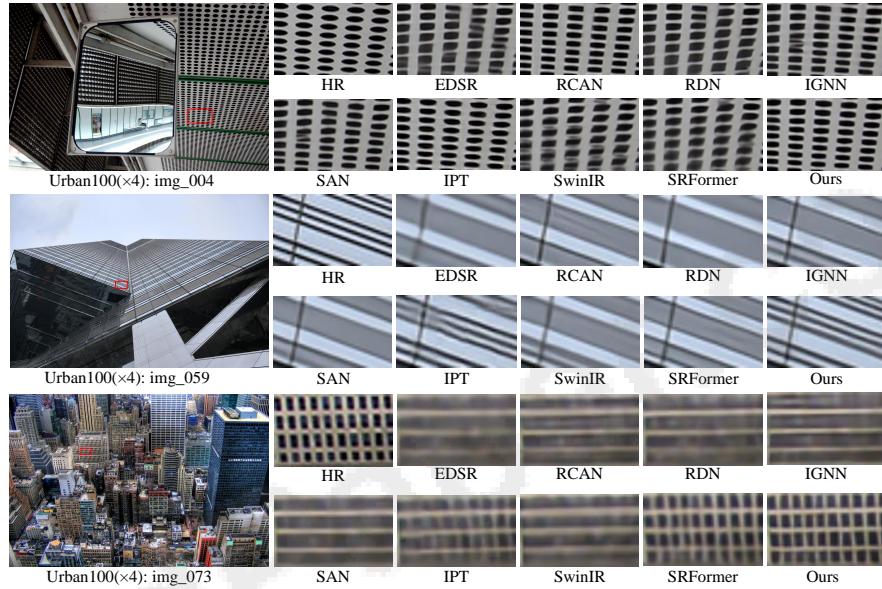
SSMs and Attention have the global modeling ability, the behavior of these two

表 1: RSSB 不同设计选择的消融实验。

| settings | Set5 | Set14 | Urban100 |
|---------------------|-------|-------|----------|
| (1)remove Conv | 38.48 | 34.54 | 34.04 |
| (2)remove Conv+CA | 38.55 | 34.64 | 34.06 |
| (3)replace with MLP | 38.55 | 34.68 | 34.22 |

表 2: VSSM 中不同扫描模式的烧蚀实验。

| scan mode | Set5 | Set14 | Urban100 |
|---------------|-------|-------|----------|
| one-direction | 38.53 | 34.63 | 34.06 |
| two-direction | 38.56 | 34.60 | 33.96 |
| baseline | 38.57 | 34.67 | 34.15 |

图 4: 我们的 MambaIR 与基于 CNN 和 Transformer 的方法在尺度为 $\times 4$ 的经典图像 SR 上的定性比较。

模块不同，因此应考虑惯用的块结构以进行进一步改进。

VSSM 中不同扫描模式的影响。为了允许 Mamba 处理二维图像，在通过状态空间方程迭代之前需要展平特征图。因此，展开策略就显得尤为重要。在这项工作中，我们遵循[44]，它使用四个不同方向的扫描来生成扫描序列。这里，我们烧蚀不同的扫描模式来研究其效果，结果如表1所示。2. 与单向（从左上到右下）和双向（从左上到右下、从右下到左上）相比，使用四个方向扫描可以让锚像素感知更广的范围的社区，从而取得更好的成果。我们还包括其他烧蚀实验，例如 RSSB 的层数，请参阅补充材料以进行更多分析。

material for more analysis.

表3：经典图像超分辨率与最先进方法的定量比较。最好和第二好的结果是红色和蓝色。

| Method | scale | Set5 | | Set14 | | BSDS100 | | Urban100 | | Manga109 | |
|---------------|------------|--------------|---------------|--------------|---------------|--------------|---------------|--------------|---------------|--------------|---------------|
| | | PSNR | SSIM |
| EDSR [42] | $\times 2$ | 38.11 | 0.9602 | 33.92 | 0.9195 | 32.32 | 0.9013 | 32.93 | 0.9351 | 39.10 | 0.9773 |
| RCAN [88] | $\times 2$ | 38.27 | 0.9614 | 34.12 | 0.9216 | 32.41 | 0.9027 | 33.34 | 0.9384 | 39.44 | 0.9786 |
| SAN [13] | $\times 2$ | 38.31 | 0.9620 | 34.07 | 0.9213 | 32.42 | 0.9028 | 33.10 | 0.9370 | 39.32 | 0.9792 |
| HAN [57] | $\times 2$ | 38.27 | 0.9614 | 34.16 | 0.9217 | 32.41 | 0.9027 | 33.35 | 0.9385 | 39.46 | 0.9785 |
| IGNN [90] | $\times 2$ | 38.24 | 0.9613 | 34.07 | 0.9217 | 32.41 | 0.9025 | 33.23 | 0.9383 | 39.35 | 0.9786 |
| CSNLN [54] | $\times 2$ | 38.28 | 0.9616 | 34.12 | 0.9223 | 32.40 | 0.9024 | 33.25 | 0.9386 | 39.37 | 0.9785 |
| NLSA [53] | $\times 2$ | 38.34 | 0.9618 | 34.08 | 0.9231 | 32.43 | 0.9027 | 33.42 | 0.9394 | 39.59 | 0.9789 |
| ELAN [87] | $\times 2$ | 38.36 | 0.9620 | 34.20 | 0.9228 | 32.45 | 0.9030 | 33.44 | 0.9391 | 39.62 | 0.9793 |
| IPT [8] | $\times 2$ | 38.37 | - | 34.43 | - | 32.48 | - | 33.76 | - | - | - |
| SwinIR [41] | $\times 2$ | 38.42 | 0.9623 | 34.46 | 0.9250 | 32.53 | 0.9041 | 33.81 | 0.9427 | 39.92 | 0.9797 |
| SRFormer [91] | $\times 2$ | 38.51 | 0.9627 | 34.44 | 0.9253 | 32.57 | 0.9046 | 34.09 | 0.9449 | 40.07 | 0.9802 |
| MambaIR | $\times 2$ | 38.57 | 0.9627 | 34.67 | 0.9261 | 32.58 | 0.9048 | 34.15 | 0.9446 | 40.28 | 0.9806 |
| MambaIR+ | $\times 2$ | 38.60 | 0.9628 | 34.69 | 0.9260 | 32.60 | 0.9048 | 34.17 | 0.9443 | 40.33 | 0.9806 |
| EDSR [42] | $\times 3$ | 34.65 | 0.9280 | 30.52 | 0.8462 | 29.25 | 0.8093 | 28.80 | 0.8653 | 34.17 | 0.9476 |
| RCAN [88] | $\times 3$ | 34.74 | 0.9299 | 30.65 | 0.8482 | 29.32 | 0.8111 | 29.09 | 0.8702 | 34.44 | 0.9499 |
| SAN [13] | $\times 3$ | 34.75 | 0.9300 | 30.59 | 0.8476 | 29.33 | 0.8112 | 28.93 | 0.8671 | 34.30 | 0.9494 |
| HAN [57] | $\times 3$ | 34.75 | 0.9299 | 30.67 | 0.8483 | 29.32 | 0.8110 | 29.10 | 0.8705 | 34.48 | 0.9500 |
| IGNN [90] | $\times 3$ | 34.72 | 0.9298 | 30.66 | 0.8484 | 29.31 | 0.8105 | 29.03 | 0.8696 | 34.39 | 0.9496 |
| CSNLN [54] | $\times 3$ | 34.74 | 0.9300 | 30.66 | 0.8482 | 29.33 | 0.8105 | 29.13 | 0.8712 | 34.45 | 0.9502 |
| NLSA [53] | $\times 3$ | 34.85 | 0.9306 | 30.70 | 0.8485 | 29.34 | 0.8117 | 29.25 | 0.8726 | 34.57 | 0.9508 |
| ELAN [87] | $\times 3$ | 34.90 | 0.9313 | 30.80 | 0.8504 | 29.38 | 0.8124 | 29.32 | 0.8745 | 34.73 | 0.9517 |
| IPT [8] | $\times 3$ | 34.81 | - | 30.85 | - | 29.38 | - | 29.49 | - | - | - |
| SwinIR [41] | $\times 3$ | 34.97 | 0.9318 | 30.93 | 0.8534 | 29.46 | 0.8145 | 29.75 | 0.8826 | 35.12 | 0.9537 |
| SRformer [91] | $\times 3$ | 35.02 | 0.9323 | 30.94 | 0.8540 | 29.48 | 0.8156 | 30.04 | 0.8865 | 35.26 | 0.9543 |
| MambaIR | $\times 3$ | 35.08 | 0.9323 | 30.99 | 0.8536 | 29.51 | 0.8157 | 29.93 | 0.8841 | 35.43 | 0.9546 |
| MambaIR+ | $\times 3$ | 35.13 | 0.9326 | 31.06 | 0.8541 | 29.53 | 0.8162 | 29.98 | 0.8838 | 35.55 | 0.9549 |
| EDSR [42] | $\times 4$ | 32.46 | 0.8968 | 28.80 | 0.7876 | 27.71 | 0.7420 | 26.64 | 0.8033 | 31.02 | 0.9148 |
| RCAN [88] | $\times 4$ | 32.63 | 0.9002 | 28.87 | 0.7889 | 27.77 | 0.7436 | 26.82 | 0.8087 | 31.22 | 0.9173 |
| SAN [13] | $\times 4$ | 32.64 | 0.9003 | 28.92 | 0.7888 | 27.78 | 0.7436 | 26.79 | 0.8068 | 31.18 | 0.9169 |
| HAN [57] | $\times 4$ | 32.64 | 0.9002 | 28.90 | 0.7890 | 27.80 | 0.7442 | 26.85 | 0.8094 | 31.42 | 0.9177 |
| IGNN [90] | $\times 4$ | 32.57 | 0.8998 | 28.85 | 0.7891 | 27.77 | 0.7434 | 26.84 | 0.8090 | 31.28 | 0.9182 |
| CSNLN [54] | $\times 4$ | 32.68 | 0.9004 | 28.95 | 0.7888 | 27.80 | 0.7439 | 27.22 | 0.8168 | 31.43 | 0.9201 |
| NLSA [53] | $\times 4$ | 32.59 | 0.9000 | 28.87 | 0.7891 | 27.78 | 0.7444 | 26.96 | 0.8109 | 31.27 | 0.9184 |
| ELAN [87] | $\times 4$ | 32.75 | 0.9022 | 28.96 | 0.7914 | 27.83 | 0.7459 | 27.13 | 0.8167 | 31.68 | 0.9226 |
| IPT [8] | $\times 4$ | 32.64 | - | 29.01 | - | 27.82 | - | 27.26 | - | - | - |
| SwinIR [41] | $\times 4$ | 32.92 | 0.9044 | 29.09 | 0.7950 | 27.92 | 0.7489 | 27.45 | 0.8254 | 32.03 | 0.9260 |
| SRFormer [91] | $\times 4$ | 32.93 | 0.9041 | 29.08 | 0.7953 | 27.94 | 0.7502 | 27.68 | 0.8311 | 32.21 | 0.9271 |
| MambaIR | $\times 4$ | 33.03 | 0.9046 | 29.20 | 0.7961 | 27.98 | 0.7503 | 27.68 | 0.8287 | 32.32 | 0.9272 |
| MambaIR+ | $\times 4$ | 33.13 | 0.9054 | 29.25 | 0.7971 | 28.01 | 0.7510 | 27.80 | 0.8303 | 32.48 | 0.9281 |

4.3 图像超分辨率对比

经典图像超分辨率。标签。图3显示了MambaIR和最先进的超分辨率方法之间的定量结果。由于具有显着的全局感受野，我们提出的MambaIR实现了最佳性能
significant global receptive field, our proposed MambaIR achieves the best per-

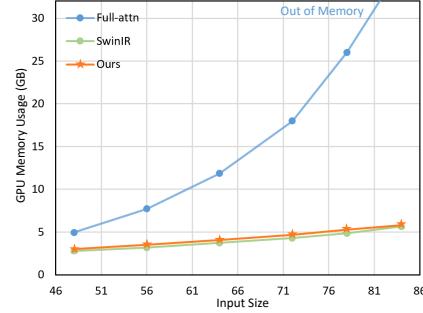


图 5：不同输入尺度的计算复杂度比较。我们将具有全局感受野的标准注意力 [65] 设置为基线，并将其表示为“Full-attn”。我们调整模型以确保一开始的 GPU 使用率大致相似，然后将输入分辨率从 48×48 缩放到 84×84 。

几乎所有五个基准数据集上所有比例因子的性能。例如，我们基于 Mamba 的基线在 Manga109 上的 $\times 2$ 尺度上优于基于 Transformer 的基准模型 SwinIR 0.41dB，展示了 Mamba 在图像恢复方面的前景。我们还在图4中给出了视觉比较，我们的方法可以促进锐利边缘和自然纹理的重建。

模型复杂性比较。我们在图 5 中给出了不同输入大小的计算复杂度的比较结果。正如我们所看到的，我们的方法比全注意力基线 [65] 更有效，并且表现出与输入分辨率的线性复杂度，类似于有效的注意力技术，例如 SwinIR。上述观察结果表明，MambaIR 具有与转移窗口注意力相似的尺度特性，同时拥有与标准全注意力相似的全局感受野。

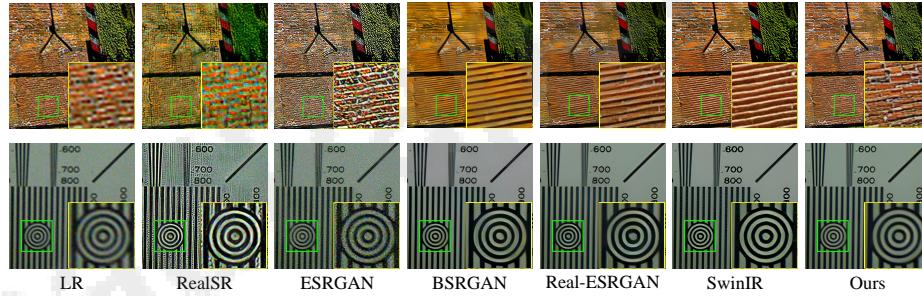
轻量级图像超分辨率。为了证明我们方法的可扩展性，我们训练了 MambaIR-light 模型并将其与最先进的图像方法进行比较。继之前的工作 [47, 91] 之后，我们还报告了参数的数量 (#param) 和 MAC（将低分辨率图像升级到 1280×720 分辨率）。图4显示了结果。可以看出，在具有相似参数和 MAC 的 $\times 4$ 尺度 Manga109 数据集上，我们的 MambaIR-light 比 SwinIR-light [41] 的 PSNR 提高了 0.34dB。性能结果证明了我们方法的可扩展性和效率。

真实世界图像超分辨率。我们还研究了网络在现实世界图像恢复方面的性能。我们遵循 [10] 中的训练协议来训练我们的 MambaIR 真实模型。由于该任务没有真实图像，因此图 6 中仅给出了视觉比较。与其他方法相比，我们的 MambaIR 在解析精细细节和纹理保留方面表现出显着的进步，证明了我们方法的稳健性。

details and texture preservation, demonstrating the robustness of our method.

表4：轻量级图像超分辨率与最先进方法的定量比较。

| Method | scale | #param | MACs | Set5 | | Set14 | | BSDS100 | | Urban100 | | Manga109 | |
|---------------------|------------|--------|--------|--------------|---------------|--------------|---------------|--------------|---------------|--------------|---------------|--------------|---------------|
| | | | | PSNR | SSIM |
| CARN [2] | $\times 2$ | 1,592K | 222.8G | 37.76 | 0.9590 | 33.52 | 0.9166 | 32.09 | 0.8978 | 31.92 | 0.9256 | 38.36 | 0.9765 |
| IMDN [30] | $\times 2$ | 694K | 158.8G | 38.00 | 0.9605 | 33.63 | 0.9177 | 32.19 | 0.8996 | 32.17 | 0.9283 | 38.88 | 0.9774 |
| LAPAR-A [39] | $\times 2$ | 548K | 171.0G | 38.01 | 0.9605 | 33.62 | 0.9183 | 32.19 | 0.8999 | 32.10 | 0.9283 | 38.67 | 0.9772 |
| LatticeNet [47] | $\times 2$ | 756K | 169.5G | 38.15 | 0.9610 | 33.78 | 0.9193 | 32.25 | 0.9005 | 32.43 | 0.9302 | - | - |
| SwinIR-light [41] | $\times 2$ | 878K | 195.6G | 38.14 | 0.9611 | 33.86 | 0.9206 | 32.31 | 0.9012 | 32.76 | 0.9340 | 39.12 | 0.9783 |
| SRFormer-light [91] | $\times 2$ | 853K | 236G | 38.23 | 0.9613 | 33.94 | 0.9209 | 32.36 | 0.9019 | 32.91 | 0.9353 | 39.28 | 0.9785 |
| Ours | $\times 2$ | 859K | 198.1G | 38.16 | 0.9610 | 34.00 | 0.9212 | 32.34 | 0.9017 | 32.92 | 0.9356 | 39.31 | 0.9779 |
| CARN [2] | $\times 3$ | 1,592K | 118.8G | 34.29 | 0.9255 | 30.29 | 0.8407 | 29.06 | 0.8034 | 28.06 | 0.8493 | 33.50 | 0.9440 |
| IMDN [30] | $\times 3$ | 703K | 71.5G | 34.36 | 0.9270 | 30.32 | 0.8417 | 29.09 | 0.8046 | 28.17 | 0.8519 | 33.61 | 0.9445 |
| LAPAR-A [39] | $\times 3$ | 544K | 114.0G | 34.36 | 0.9267 | 30.34 | 0.8421 | 29.11 | 0.8054 | 28.15 | 0.8523 | 33.51 | 0.9441 |
| LatticeNet [47] | $\times 3$ | 765K | 76.3G | 34.53 | 0.9281 | 30.39 | 0.8424 | 29.15 | 0.8059 | 28.33 | 0.8538 | - | - |
| SwinIR-light [41] | $\times 3$ | 886K | 87.2G | 34.62 | 0.9289 | 30.54 | 0.8463 | 29.20 | 0.8082 | 28.66 | 0.8624 | 33.98 | 0.9478 |
| SRFormer-light [91] | $\times 3$ | 861K | 105G | 34.67 | 0.9296 | 30.57 | 0.8469 | 29.26 | 0.8099 | 28.81 | 0.8655 | 34.19 | 0.9489 |
| Ours | $\times 3$ | 867K | 88.7G | 34.72 | 0.9296 | 30.63 | 0.8475 | 29.29 | 0.8099 | 29.00 | 0.8689 | 34.39 | 0.9495 |
| CARN [2] | $\times 4$ | 1,592K | 90.9G | 32.13 | 0.8937 | 28.60 | 0.7806 | 27.58 | 0.7349 | 26.07 | 0.7837 | 30.47 | 0.9084 |
| IMDN [30] | $\times 4$ | 715K | 40.9G | 32.21 | 0.8948 | 28.58 | 0.7811 | 27.56 | 0.7353 | 26.04 | 0.7838 | 30.45 | 0.9075 |
| LAPAR-A [39] | $\times 4$ | 659K | 94.0G | 32.15 | 0.8944 | 28.61 | 0.7818 | 27.61 | 0.7366 | 26.14 | 0.7871 | 30.42 | 0.9074 |
| LatticeNet [47] | $\times 4$ | 777K | 43.6G | 32.30 | 0.8962 | 28.68 | 0.7830 | 27.62 | 0.7367 | 26.25 | 0.7873 | - | - |
| SwinIR-light [41] | $\times 4$ | 897K | 49.6G | 32.44 | 0.8976 | 28.77 | 0.7858 | 27.69 | 0.7406 | 26.47 | 0.7980 | 30.92 | 0.9151 |
| SRFormer-light [91] | $\times 4$ | 873K | 62.8G | 32.51 | 0.8988 | 28.82 | 0.7872 | 27.73 | 0.7422 | 26.67 | 0.8032 | 31.17 | 0.9165 |
| Ours | $\times 4$ | 879K | 50.6G | 32.51 | 0.8993 | 28.85 | 0.7876 | 27.75 | 0.7423 | 26.75 | 0.8051 | 31.26 | 0.9175 |

图6：与RealSR [32]、ESRGAN [68]、BSRGAN [80]、Real-ESRGAN [67]和SwinIR [41]在尺度 $\times 4$ 的真实图像超分辨率上的定性比较。

4.4 图像去噪比较

高斯彩色图像去噪。高斯彩色图像去噪结果如表2所示。5. 根据 [79,81]，比较的噪声级别包括15、25和50。可以看出，我们的模型在大多数数据集上实现了最佳性能。特别是，它在 Urban100 数据集上 $\sigma=50$ 时甚至超过了 SwinIR [41] 0.48dB。我们还在图7中给出了视觉比较。由于全局感受野，我们的MambaIR可以实现更好的结构保留，从而导致更清晰的边缘和自然的形状。真实图像去噪。我们进一步转向真实图像去噪任务来评估 MambaIR 在面对现实世界退化时的鲁棒性。按照[72]，我们采用渐进式训练策略进行公平比较。这

lowing [72]，we adopt the progressive training strategy for fair comparison. The

表 5：与最先进方法的高斯彩色图像去噪的定量比较。

| Method | BSD68 | | | Kodak24 | | | McMaster | | | Urban100 | | |
|----------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | $\sigma=15$ | $\sigma=25$ | $\sigma=50$ |
| IRCNN [82] | 33.86 | 31.16 | 27.86 | 34.69 | 32.18 | 28.93 | 34.58 | 32.18 | 28.91 | 33.78 | 31.20 | 27.70 |
| FFDNet [83] | 33.87 | 31.21 | 27.96 | 34.63 | 32.13 | 28.98 | 34.66 | 32.35 | 29.18 | 33.83 | 31.40 | 28.05 |
| DnCNN [81] | 33.90 | 31.24 | 27.95 | 34.60 | 32.14 | 28.95 | 33.45 | 31.52 | 28.62 | 32.98 | 30.81 | 27.59 |
| DRUNet [79] | 34.30 | 31.69 | 28.51 | 35.31 | 32.89 | 29.86 | 35.40 | 33.14 | 30.08 | 34.81 | 32.60 | 29.61 |
| SwinIR [41] | 34.42 | 31.78 | 28.56 | 35.34 | 32.89 | 29.79 | 35.61 | 33.20 | 30.22 | 35.13 | 32.90 | 29.82 |
| Restormer [72] | 34.40 | 31.79 | 28.60 | 35.47 | 33.04 | 30.01 | 35.61 | 33.34 | 30.30 | 35.13 | 32.96 | 30.02 |
| MambaIR | 34.48 | 32.24 | 28.66 | 35.42 | 32.99 | 29.92 | 35.70 | 33.43 | 30.35 | 35.37 | 33.21 | 30.30 |

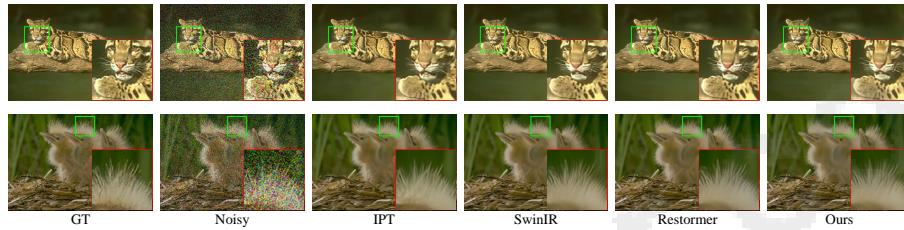
Fig. 7: Qualitative comparison of our MambaIR with other methods on color image denoising task with noise level level $\sigma=50$.

表 6：真实图像去噪任务的定量比较。

| Dataset | DeamNet [60] | | MPRNet [73] | | DAGL [55] | | Uformer [69] | | Restormer [72] | | Ours | |
|---------|--------------|-------|-------------|-------|-----------|-------|--------------|-------|----------------|--------------|--------------|--------------|
| | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| SIDD | 39.47 | 0.957 | 39.71 | 0.958 | 38.94 | 0.953 | 39.77 | 0.959 | 40.02 | 0.960 | 39.89 | 0.960 |
| DND | 39.63 | 0.953 | 39.80 | 0.954 | 39.77 | 0.956 | 39.96 | 0.956 | 40.03 | 0.956 | 40.04 | 0.956 |

如表所示。如图 6 所示，表明我们的方法在 SIDD 数据集上实现了与现有最先进模型 Restormer [69] 相当的性能，并且比 Uformer [69] 等其他方法的性能高出 0.12dB PSNR，这表明了我们方法的能力在真实图像去噪中。

5 结论

在这项工作中，我们首次探索了最近先进的状态空间模型（即 Mamba）在图像恢复方面的强大功能，以帮助解决高效计算和全局有效感受野之间的权衡困境。具体来说，我们引入了局部增强以减轻扁平化策略中的邻域像素遗忘问题，并提出通道注意以减少通道冗余。对多个恢复任务的广泛实验表明，我们的 MambaIR 是一种简单但有效的图像恢复状态空间模型。

model for image restoration.

Acknowledgements

这项工作得到了国家自然科学基金项目（62302309、62171248）、深圳市科技计划（JCYJ20220818101014030、JCYJ20220818101012025）和PCNLKEY项目（PCL2023AS6-1）的部分支持。

References

1. Abdelhamed, A., Lin, S., Brown, M.S.: 智能手机摄像头的高质量去噪数据集。见：IEEE 计算机视觉和模式识别会议论文集。第 1692–1700 页 (2018)2。Ahn, N., Kang, B., Sohn, K.A.: 使用级联残差网络实现快速、准确且轻量级的超分辨率。见：欧洲计算机视觉会议 (ECCV) 会议记录。第 252–268 页 (2018)3。Arbelaez, P., Maire, M., Fowlkes, C., Malik, J.: 轮廓检测和分层图像分割。IEEE 模式分析和机器智能汇刊 33(5), 898–916 (2010)4。Bai, J., Gao, K., Min, S., Xia, S.T., Li, Z., Liu, W.: Badclip: 针对剪辑后门攻击的触发感知提示学习。见：CVPR (2024)5。Bevilacqua, M., Roomy, A., Guillemot, C., Alberi-Morel, M.L.: 基于非负邻域嵌入的低复杂度单图像超分辨率 (2012)6。Cavigelli, L., Hager, P., Benini, L.: Cas-cnn: 用于图像压缩伪影抑制的深度卷积神经网络。见：2017 年国际神经网络联合会议 (IJCNN)。第 752–759 页。IEEE (2017) 7。
- Charbonnier, P., Blanc-Feraud, L., Aubert, G., Barlaud, M.: 用于计算成像的两种确定性半二次正则化算法。见：第一届图像处理国际会议论文集。卷。2, 第 168–172 页。IEEE (1994) 8。陈红、王云、郭天、徐成、邓勇、刘志、马胜、徐成、徐成、高文：预先训练的图像处理变压器。见：IEEE/CVF 计算机视觉和模式识别会议论文集。第 12299–12310 页 (2021)9。陈L., 褚X., 张X., 孙J.: 图像恢复的简单基线。见：欧洲计算机视觉会议。第 17–33 页。施普林格 (2022) 10。陈X., 王X., 周J., 乔Y., 董C.: 在图像超分辨率转换器中激活更多像素。见：IEEE/CVF 计算机视觉和模式识别会议论文集。第 22367–22377 页 (2023)11。Chen, Z., Zhang, Y., Gu, J., Kong, L., Yang, X.: 图像超分辨率的递归泛化变换器。arXiv 预印本 arXiv: 2303.06373 (2023)12。Chen, Z., Zhang, Y., Gu, J., Kong, L., Yang, X., Yu, F.: 图像超分辨率的双聚合变换器。见：IEEE/CVF 计算机视觉国际会议论文集。第 12312–12321 页 (2023)13。戴 T., 蔡 J., 张 Y., 夏 S.T., 张 L.: 单图像超分辨率的二阶注意力网络。见：IEEE/CVF 计算机视觉和模式识别会议论文集。第 11065–11074 页 (2019)14。Ding, X., Zhang, X., Han, J., Ding, G.: 将内核扩展到 31x31：重新审视 CNN 中的大型内核设计。见：IEEE/CVF 计算机视觉和模式识别会议论文集。第 11963–11975 页 (2022)15。Dong, C., Deng, Y., Loy, C.C., Tang, X.: 通过深度卷积网络减少压缩伪影。见：IEEE 计算机视觉国际会议论文集。第 576–584 页 (2015)

on computer vision. pp. 576–584 (2015)

16. Dong, C., Loy, C.C., He, K., Tang, X.: 学习图像超分辨率的深度卷积网络。见: 计算机视觉 - ECCV 2014: 第 13 届欧洲会议, 瑞士苏黎世, 2014 年 9 月 6-12 日, 会议记录, 第 IV 部分 13。第 184-199 页。
- 施普林格 (2014) 17。Franzen, R.: 柯达无损真彩色图像套件 (2021), <http://r0k.us/graphics/kodak/18>。Fu, D.Y., Dao, T., Saab, K.K., Thomas, A.W., Rudra, A., Ré, C.: 饥饿的饥饿河马: 利用状态空间模型进行语言建模。arXiv 预印本 arXiv: 2212.14052 (2022)19。Fu, X., Zha, Z.J., Wu, F., Ding, X., Paisley, J.: 通过深度卷积稀疏编码减少 Jpeg 伪影。见: IEEE/CVF 国际计算机视觉会议论文集。第 2501-2510 页 (2019)20。高, K., 白, Y., 顾, J., 夏, S.T., 托尔, P., 李, Z., 刘, W.: 用详细图像诱导大型视觉语言模型的高能量延迟。见: ICLR (2024)21。Gau, K., Gu, J., Bai, Y., Xia, S.T., Torr, P., Liu, W., Li, Z.: 通过详细样本对多模态大语言模型进行能量延迟操作。arXiv 预印本 arXiv: 2404.16557 (2024)22。Gu, A., Dao, T.: Mamba: 具有选择性状态空间的线性时间序列建模。arXiv 预印本 arXiv:2312.00752 (2023)23。Gu, A., Dao, T., Ermon, S., Rudra, A., Ré, C.: Hippo: 具有最佳多项式投影的循环记忆。神经信息处理系统的进展 33, 1474-1487 (2020)24。Gu, A., Goel, K., Ré, C.: 利用结构化状态空间对长序列进行高效建模。arXiv 预印本 arXiv:2111.00396 (2021)25。Gu, A., Johnson, I., Goel, K., Saab, K., Dao, T., Rudra, A., Ré, C.: 将循环、卷积和连续时间模型与线性状态空间层相结合。神经信息处理系统的进展 34, 572-585 (2021)26。郭华、戴涛、白云、陈波、夏思涛、朱Z.: Adaptir: 预训练图像恢复模型的参数高效多任务自适应。arXiv preprint arXiv:2312.08881 (2023)27。Hu, J., Shen, L., Sun, G.: 挤压激励网络。见: IEEE 计算机视觉和模式识别会议论文集。第 7132-7141 页 (2018)28。Hu, V.T., Baumann, S.A., Gui, M., Grebenkova, O., Ma, P., Fischer, J., Ommer, B.: Zigma: 一种点状锯齿曼巴扩散模型。见: ECCV (2024)29。Huang, J.B., Singh, A., Ahuja, N.: 来自变换自我范例的单图像超分辨率。见: IEEE 计算机视觉和模式识别会议论文集。第 5197-5206 页 (2015)30。Hui Z., Gao, X., Yang, Y., Wang, X.: 基于信息多重蒸馏网络的轻量级图像超分辨率。见: 第 27 届 ACM 国际多媒体会议论文集。第 2024-2032 页 (2019)31。Islam, M.M., Hasan, M., Athrey, K.S., Braskich, T., Bertasius, G.: 使用状态空间转换器进行高效电影场景检测。见: IEEE/CVF 计算机视觉和模式识别会议论文集。第 18749-18758 页 (2023)32。Ji, X., Cao, Y., Tai, Y., Wang, C., Li, J., Huang, F.: 现实世界的超分辨率通过核估计和噪声注入。见: IEEE/CVF 计算机视觉和模式识别研讨会会议记录。第 466-467 页 (2020)33。Kalman, R.E.: 线性滤波和预测问题的新方法 (1960) 34。Kim, J., Lee, J.K., Lee, K.M.: 使用非常深的卷积网络实现精确的图像超分辨率。见: IEEE 计算机视觉和模式识别会议论文集。第 1646-1654 页 (2016)

and pattern recognition. pp. 1646-1654 (2016)

35. Kingma, D.P., Ba, J.: Adam: 一种随机优化方法。arXiv preprint arXiv: 1412.6980 (2014)36。Lai, W.S.、Huang, J.B.、Ahuja, N.、Yang, M.H.: 用于快速准确超分辨率的深层拉普拉斯金字塔网络。见: IEEE 计算机视觉和模式识别会议论文集。第 624–632 页 (2017)37。
- Li, K., Li, X., Wang, Y., He, Y., Wang, Y., Wang, L., Qiao, Y.: Videomamba: 用于高效视频理解的状态空间模型。arXiv 预印本 arXiv: 2403.06977(2024)38。Li, W., Lu, X., Qian, S., Lu, J., Zhang, X., Jia, J.: 基于 Transformer 的高效低级视觉图像预训练。arXiv 预印本 arXiv: 2112.10175 (2021)39。Li, W., Zhou, K., Qi, L., Jiang, N., Lu, J., Jia, J.: Lapar: 用于单图像超分辨率及以上的线性组装像素自适应回归网络。神经信息处理系统 33, 20343–20355 (2020)40。Li, Y., Fan, Y., Xiang, X., Demandolx, D., Ranjan, R., Timofte, R., Van Gool, L.: 用于图像恢复的图像层次结构的高效且明确的建模。见: IEEE/CVF 计算机视觉和模式识别会议论文集。第 18278–18289 页 (2023)41。Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: Swinir: 使用 swin 变压器进行图像恢复。见: IEEE/CVF 计算机视觉国际会议论文集。第 1833–1844 页 (2021)42。Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K.: 用于单图像超分辨率的增强型深度残差网络。见: IEEE 计算机视觉和模式识别研讨会会议记录。第 136–144 页 (2017)43。Liu, P., Guo, H., Dai, T., Li, N., Bao, J., Ren, X., Jiang, Y., Xia, S.T.: 通过交叉训练用于广义时间序列预测的预训练 llms -模态知识蒸馏。arXiv 预印本 arXiv: 2403.07300 (2024)44。刘Y., 田Y., 赵Y., 于H., 谢L., 王Y., 叶Q., 刘Y.: Vmamba: 视觉状态空间模型。arXiv 预印本 arXiv: 2401.10166 (2024)45。Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swintransformer: 使用移动窗口的分层视觉变换器。见: IEEE/CVF 计算机视觉国际会议论文集。第 10012–10022(2021)46 页。Luo, W., Li, Y., Urtasun, R., Zemel, R.: 了解深度卷积神经网络中的有效感受野。神经信息处理系统的进展 29 (2016) 47。罗X., 谢Y., 张Y., 曲Y., 李成., 付Y.: Latticenet: 利用晶格块实现轻量级图像超分辨率。见: 计算机视觉 - ECCV 2020: 第 16 届欧洲会议, 英国格拉斯哥, 2020 年 8 月 23 日至 28 日, 会议记录, 第 XXII 部分 16。第 272–289 页。施普林格 (2020) 48。Ma, J., Li, F., Wang, B.: U-mamba: 增强生物医学图像分割的远程依赖性。arXiv 预印本 arXiv: 2401.04722 (2024)49。Ma, K., Duanmu, Z., Wu, Q., Wang, Z., Yong, H., Li, H., Zhang, L.: 滑铁卢探索数据库: 图像质量评估模型的新挑战。IEEE 图像处理交易 26(2), 1004–1016 (2016)50。Martin, D., Fowlkes, C., Tal, D., Malik, J.: 人类分割自然图像数据库及其在评估分割算法和测量生态统计方面的应用。见: 第八届 IEEE 国际计算机视觉会议论文集。ICCV 2001. 卷。2, 第 416–423 页。IEEE (2001) 51。Matsui, Y., Ito, K., Aramaki, Y., Fujimoto, A., Okawa, T., Yamasaki, T., Aizawa, K.: 使用 manga109 数据集进行基于草图的漫画检索。多媒体工具和应用 76, 21811–21838 (2017)

52. Mehta, H., Gupta, A., Cutkosky, A., Neyshabur, B.: 通过门控状态空间进行长程语言建模。arXiv 预印本 arXiv: 2206.13947 (2022)53. Mei, Y., Fan, Y., Zhou, Y.: 具有非局部稀疏注意力的图像超分辨率。见: IEEE/CVF 计算机视觉和模式识别会议论文集。第 3517–3526 页 (2021)54. Mei, Y., Fan, Y., Zhou, Y., Huang, L., Huang, T.S., Shi, H.: 图像超分辨率, 跨尺度非局部注意力和详尽的自范例挖掘。见: IEEE/CVF 计算机视觉和模式识别会议论文集.pp。5690–5699 (2020)55. Mou, C., Zhang, J., Wu, Z.: 用于图像恢复的动态注意力图学习。见: IEEE/CVF 计算机视觉国际会议论文集。第 4328–4337 页 (2021)56. Nguyen, E., Goel, K., Gu, A., Downs, G., Shah, P., Dao, T., Baccus, S., Ré, C.: S4nd: 将图像和视频建模为多维信号状态空间。神经信息处理系统的进展 35, 2846–2861 (2022)57. 牛B.、文W.、任W.、张X.、杨L.、王S.、张K.、曹X.、沉H.: 单图像超分辨率通过整体注意力网络。见: 计算机视觉 - ECCV 2020: 第 16 届欧洲会议, 英国格拉斯哥, 2020 年 8 月 23-28 日, 会议记录, 第 XII 部分 16。第 191-207 页。施普林格 (2020) 58. Plotz, T., Roth, S.: 用真实照片对去噪算法进行基准测试。见: IEEE 计算机视觉和模式识别会议论文集.pp。1586–1595 (2017) 59. 秦S.、王J.、周Y.、陈B.、罗T.、安B.、戴T.、夏S.、王Y.: Mambavc: 学习视觉压缩具有选择性状态空间。arXiv preprint arXiv:2405.15413 (2024)60. 任春、何新、王春、赵志: 基于自适应一致性先验的图像去噪深度网络。见: IEEE/CVF 计算机视觉和模式识别会议论文集。第 8596–8606 页 (2021)61. Shazeer, N.: Glu 变体改进了变压器。arXiv 预印本 arXiv: 2002.05202(2020)62. Smith, J.T., Warrington, A., Linderman, S.W.: 用于序列建模的简化状态空间层。arXiv 预印本 arXiv: 2208.04933 (2022)63. Sun, L., Dong, J., Tang, J., Pan, J.: 用于高效图像超分辨率的空间自适应特征调制。见: ICCV (2023)64. Timofte, R., Agustsson, E., Van Gool, L., Yang, M.H., Zhang, L.: Ntire 2017 单图像超分辨率挑战: 方法和结果。见: IEEE 计算机视觉和模式识别研讨会会议记录。第 114–125 页 (2017)65. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: 注意力就是您所需要的。神经信息处理系统的进展 30 (2017) 66. Wang, J., Zhu, W., Wang, P., Yu, X., Liu, L., Omar, M., Hamid, R.: 用于长格式视频理解的选择性结构化状态空间。见: IEEE/CVF 计算机视觉和模式识别会议论文集。第 6387–6397 页 (2023)67. Wang, X., Xie, L., Dong, C., Shan, Y.: Real-esrgan: 用纯合成数据训练真实世界的盲超分辨率。见: IEEE/CVF 计算机视觉国际会议论文集。第 1905–1914 页 (2021)68. Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., Loy, C.C.: Esr-gan: 增强型超分辨率生成对抗网络。见: 欧洲计算机视觉研讨会 (ECCVW) (2018 年 9 月)

69. Wang, Z., Cun, X., Bao, J., Zhou, W., Liu, J., Li, H.: Uformer: 一种用于图像恢复的通用 U 形变压器。见: IEEE/CVF 计算机视觉和模式识别 (CVPR) 会议论文集。第 17683–17693 页 (2022 年 6 月) 70。Wei, Y., Gu, S., Li, Y., Timofte, R., Jin, L., Song, H.: 通过域距离感知训练实现无监督现实世界图像超分辨率。见: IEEE/CVF 计算机视觉和模式识别会议论文集。第 13385–13394 页 (2021)71。Xing, Z., Ye, T., Yang, Y., Liu, G., Zhu, L.: Segmamba: 用于 3D 医学图像分割的远程顺序建模曼巴。arXiv 预印本 arXiv: 2401.13560(2024)72。Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H.: Restormer: 用于高分辨率图像恢复的高效变压器。见: IEEE/CVF 计算机视觉和模式识别会议论文集。第 5728–5739(2022)73 页。Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: 多阶段渐进图像恢复。见: IEEE/CVF 计算机视觉和模式识别会议论文集。第 14821–14831 (2021)74 页。Zeyde, R., Elad, M., Protter, M.: 使用稀疏表示进行单图像放大。见: 曲线与曲面: 第七届国际会议, 法国阿维尼翁, 2010 年 6 月 24-30 日, 修订后的精选论文 7。第 711–730 页。施普林格 (2012) 75。Zha, Y., Ji, H., Li, J., Li, R., Dai, T., Chen, B., Wang, Z., Xia, S.T.: 通过点特征增强掩模自动编码器实现紧凑的 3d 表示。见: AAAI 人工智能会议论文集。卷。38, 第 6962–6970 页 (2024)76。查 Y.、李 N.、王 Y.、戴 T.、郭 H.、陈 B.、王 Z.、欧阳 Z.、夏 S.T.: Lcm: 局部约束紧致点用于掩蔽点建模的云模型。arXiv 预印本 arXiv: 2405.17149 (2024)77。Zha, Y., Wang, J., Dai, T., Chen, B., Wang, Z., Xia, S.T.: 预训练点云模型的实例感知动态提示调整。见: IEEE/CVF 国际计算机视觉会议论文集。第 14161–14170 (2023)78 页。张 J., 张 Y., 顾 J., 张 Y., 孔 L., 袁 X.: 使用注意可伸缩变压器的精确图像恢复。见: ICLR (2023)79。张, K., 李, Y., 左, W., 张, L., 范古尔, L., 蒂莫特, R.: 使用深度降噪器先验进行即插即用图像恢复。IEEE 模式分析和机器智能汇刊 44(10), 6360–6376 (2021)80。张, K., 梁, J., 范古尔, L., 蒂莫特, R.: 设计深度盲图像超分辨率的实用退化模型。见: IEEE/CVF 国际计算机视觉会议论文集。第 4791–4800 页 (2021)81。张 K., 左 W., 陈 Y., 孟德., 张 L.: 超越高斯去噪器: 深度 CNN 的残差学习用于图像去噪。IEEE 图像处理交易 26(7), 3142–3155 (2017)82。张 K., 左 W., 顾 S., 张 L.: 学习深度 CNN 去噪器先验以进行图像恢复。见: IEEE 计算机视觉和模式识别会议论文集。第 3929–3938 页 (2017)83。张 K., 左 W., 张 L.: Ffdnet: 针对基于 CNN 的图像去噪的快速灵活的解决方案。IEEE 图像处理汇刊 27(9), 4608–4622(2018)84。张 L., 吴 X., 布阿德斯 A., 李 X.: 通过局部定向插值和非局部自适应阈值进行颜色去马赛克。电子成像杂志 20(2), 023016–023016 (2011)

- 85.Zhang, T., Bai, J., Lu, Z., Lian, D., Wang, G., Wang, X., Xia, S.T.: 视觉变压器的参数高效和内存高效调整: 解开方法。arXiv 预印本 arXiv: 2407.06964 (2024)86。张, T., 何, S., 戴, T., 王, Z., 陈, B., 夏, S.T.: 视觉语言预训练与对象对比学习用于 3D 场景理解。见: AAAI 人工智能会议论文集。卷。38, 第 7296–7304 页 (2024)87。张X., 曾红, 郭S., 张L.: 用于图像超分辨率的高效远程注意力网络。见: 欧洲计算机视觉会议。第 649–667 页。施普林格 (2022) 88。张, Y., 李, K., 李, K., 王, L., 钟, B., 付, Y.: 使用非常深的残差通道注意网络的图像超分辨率。见: 欧洲计算机视觉会议 (ECCV) 会议记录。第 286–301 页 (2018)89。张Y., 田Y., 孔Y., 钟B., 付Y.: 用于图像超分辨率的残差密集网络。见: IEEE 计算机视觉和模式识别会议论文集。第 2472–2481 页 (2018)90。Zhou, S., Zhang, J., Zuo, W., Loy, C.C.: 用于图像超分辨率的跨尺度内部图神经网络。见: 神经信息处理系统的进展 (2020) 91。Zhou, Y., Li, Z., Guo, C.L., Bai, S., Cheng, M.M., Hou, Q.: Srformer: 单图像超分辨率的置换自注意力。arXiv 预印本 arXiv: 2303.09735(2023)92。朱L., 廖B., 张Q., 王X., 刘W., 王X.: 视觉曼巴: 利用双向状态空间模型的高效视觉表示学习。arXiv preprint arXiv:2401.09417 (2024)
- arXiv:2401.09417 (2024)