

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2025)10-3173-14

论文引用格式: Xiao J, Fan Z H, Li D, Fu X Y and Zha Z J. 2025. Non-causal selective state space model for image restoration. Journal of Image and Graphics, 30(10):3173-3186(肖杰, 范子豪, 李东, 傅雪阳, 查正军. 2025. 面向图像复原的非因果选择性状态空间模型. 中国图象图形学报, 30(10):3173-3186)[DOI:10.11834/jig.240517]

面向图像复原的非因果选择性状态空间模型

肖杰, 范子豪, 李东, 傅雪阳, 查正军*

中国科学技术大学信息科学技术学院, 合肥 230026

摘要: 目的 图像复原是计算机视觉领域的经典研究问题。选择性状态空间模型(selective state space model, selective SSM)因其高效的序列建模能力,广泛应用于各类图像复原任务。另外,非局部图像块之间存在依赖关系,能够辅助提升复原性能。而传统SSM采用确定性的令牌(token)扫描方式,仅能提取令牌序列的单向依赖关系。此时,令牌间的关系建模因在序列中的先后顺序受到因果性制约,这与图像块之间的非因果相互关系形成冲突,限制了复原性能的进一步提升。针对此问题,提出一种面向图像复原的非因果选择性状态空间模型,旨在赋予SSM建模令牌之间非因果依赖关系的能力。**方法** 为解决SSM在因果性建模与图像内容非因果关系之间的矛盾,提出随机扫描策略,突破了传统扫描方式在因果性和空间限制上的局限,实现了令牌序列之间的非因果建模。具体而言,构建了随机重排和逆重排函数,实现了非固定次序下的令牌扫描,有效建模了不同令牌之间的非因果依赖关系。此外,针对图像退化干扰存在空间尺度变化和形态结构复杂的特点,融合多尺度先验构建了具有局部与全局信息互补性的非因果Mamba模型(non-causal Mamba, NCMamba),实现了对于各类图像复原任务的有效适配。**结果** 实验分别在图像去噪、去模糊和去阴影任务上进行,验证了所提非因果建模和局部-全局互补策略的有效性。与现有方法相比,所提模型在图像去阴影数据集SRD(shadow removal dataset)上的峰值信噪比提升0.86 dB。**结论** 面向图像复原任务,构建了非因果选择性状态空间模型,建模了令牌之间的非因果依赖关系,实现了局部与全局信息的有效互补,显著提升了复原性能。所提方法在主客观评价指标上均取得优异性能,为图像复原领域提供了新的解决方案。

关键词: 图像复原;选择性状态空间模型;非因果建模;多尺度建模;图像处理

Non-causal selective state space model for image restoration

Xiao Jie, Fan Zihao, Li Dong, Fu Xueyang, Zha Zhengjun*

School of Information Science and Technology, University of Science and Technology of China, Hefei 230026, China

Abstract: Objective Image restoration plays a crucial role in enhancing the quality of degraded images and boosting the efficacy of downstream applications such as image segmentation, object detection, and autonomous driving. This problem centers on reconstructing a high-quality image from its compromised version, which tackles various distortions such as noise, blur, and compression artifacts. In the era of deep learning, the effectiveness of image restoration techniques has remarkably improved, which is largely due to the development of sophisticated models including convolutional neural networks (CNNs), graph models, and vision Transformers tailored for complex image degradation scenarios. A recently introduced selective state space model, which is known as Mamba, has propelled considerable progress across domains including natural language processing and computer vision. Mamba distinguishes itself with its robust expressive modeling capa-

收稿日期: 2024-09-02; 修回日期: 2024-12-18; 预印本日期: 2024-12-25

* 通信作者: 查正军 zhazj@ustc.edu.cn

基金项目: 国家自然科学基金项目(62225207, 62436008, 62422609, 62276243)

Supported by: National Natural Science Foundation of China (62225207, 62436008, 62422609, 62276243)

bilities, which enable it to adeptly capture and depict intricate data patterns. Moreover, Mamba offers enhanced computational efficiency and a reduced memory footprint, which surpass those of previous models. To utilize Mamba for image processing, tokens must be extracted from images. Following token acquisition, Mamba models the interactions and dependencies along the token sequence. Mamba leverages advanced state space modeling to decode the structure and semantics of an image by analyzing token interactions. In image restoration, tokens represent image patches, and their dependencies are typically considered non-causal, which allows interactions regardless of spatial distance. The existing deterministic token scanning in selective state space models, which imposes a causal relationship and one-way interactions, conflicts with this non-causal approach. This mismatch may constrain the potential of these models in image restoration, which affects their performance. To bridge this gap, we introduce NCMamba, which is a novel model that reconciles the causal bias of selective state space models with the non-causal nature of image patch interactions. Our research validates the effectiveness of NCMamba across diverse image restoration challenges. **Method** To mitigate the contradiction between the inherent causal relationships of the selective state space model and the non-causal relationships among spatial image patches, we introduce a stochastic scanning strategy for tokens. This innovative method deviates from traditional deterministic scanning techniques through effectively eliminating the intrinsic causal constraints. This approach facilitates a more comprehensive and nuanced modeling of the spatial relationships within image patches, which considerably boosts the capability of the model to perform high-quality image restoration. The strategy employs a random shuffle function to disrupt the linear arrangement of tokens. This function ensures that the interaction likelihood of each token pair is equalized, which enables the model to effectively capture non-causal dependencies. An inverse shuffle function complements this approach by reordering the tokens to their original sequence, which aids in precise image reconstruction. Specifically, the random shuffle function stochastically permutes the token sequence. As a result, the concept of “distance” between tokens is eradicated, and all tokens are treated equally post-shuffle. The relationship between every token pair then has the same probability of being processed directly by the selective state space model. Consequently, the inverse shuffle function restores the shuffled tokens to their original order, which facilitates the accurate reconstitution of the data structure of the image. Combining multi-scale, local relationship, and global relationship priors, our proposed model, which is dubbed NCMamba (non-causal Mamba), integrates local-global complementary modeling within a UNet structure, which showcases general modeling capability for diverse image restoration tasks. **Result** NCMamba was tested across multiple datasets, particularly on image denoising, deblurring, and shadow removal. The experiments demonstrate that NCMamba consistently outperformed established models such as CNNs and vision Transformers. Notably, on the SRD dataset for image shadow removal, NCMamba achieved a PSNR improvement of 0.86 dB over the ShadowDiffusion method. Ablation studies further validated the effectiveness of our non-causal modeling and local-global complementary strategies. **Conclusion** This study introduces NCMamba, which is a tailored selective state space model designed for image restoration. Our extensive testing confirms its superiority over traditional models, which establishes it as a robust solution for complex image restoration tasks. The empirical evidence underscores the capability of NCMamba to deliver significant improvements in image quality under various conditions.

Key words: image restoration; selective state space model(selective SSM); non-causal modeling; multi-scale modeling; image processing

0 引言

图像复原技术旨在从退化图像中恢复出高质量的清晰图像,因其复杂的多对一映射关系,通常被视为一个典型的病态逆问题。图像复原不仅对提升图像质量至关重要,而且是许多下游视觉任务的基础。根据不同的退化过程,图像复原可以分为多个具体任务,如图像去阴影、去模糊和去噪等。

深度学习技术的快速发展极大推动了图像复原技术的进步。在众多深度学习模型中,卷积神经网络(convolutional neural network, CNN)凭借其优秀的局部建模能力,在图像超分辨率(Dong等, 2016)、图像去噪(Zhang等, 2017)等各类图像复原任务中展现出卓越性能。随后,研究者们将更多的图像处理领域知识融入模型中,如多尺度特征增强 CNN(Zamir等, 2020b)、多阶段逐步优化的 CNN(Zamir等, 2021),显著提升了CNN的表征与复原能力。然

而, CNN 的局部建模方式导致其难以捕获图像中的长距离依赖关系, 限制了其在图像复原领域的进一步发展。

与 CNN 不同, 视觉 Transformer (Dosovitskiy 等, 2021) 突破了感受野的局限, 能够更好地建模图像中的全局依赖关系。图像处理 Transformer (Chen 等, 2021) 通过在大规模通用数据集上进行预训练, 并在专门的图像复原小数据集上进行微调, 将 Transformer 引入图像复原领域, 在多个图像复原任务中取得了卓越性能。另外, 视觉 Transformer 在高分辨率图像复原中的应用仍面临挑战, 其时间和存储复杂度随令牌数呈二次增长, 给实际应用带来了挑战。针对这一问题, 研究者通过引入局部先验来提高视觉 Transformer 的计算效率, 建立了 Uformer 模型 (Wang 等, 2022) 和 SwinIR 模型 (Liang 等, 2021) 等新模式, 并在各类图像复原任务中展现了出色性能。然而, 局部先验的引入使得 Transformer 丢失了捕捉全局依赖关系的能力。因此, 如何在高效建模长范围依赖关系的同时保持模型表征能力, 是将注意力机制应用于图像复原中的难题。

近年来, 状态空间模型因其在令牌序列建模方面的高效性, 已在自然语言处理 (Gu 和 Dao, 2024) 及计算机视觉 (Zhu 等, 2024) 领域取得了显著进展。如图 1 所示, 状态空间模型通常将图像块 (或像素) 视为令牌, 并采用确定性扫描方式将图像展开成令牌序列。这一建模方法的核心思想是基于前 j 个令牌预测第 i 个令牌, 即建模条件概率 $p(x_i | x_{j < i})$ 。

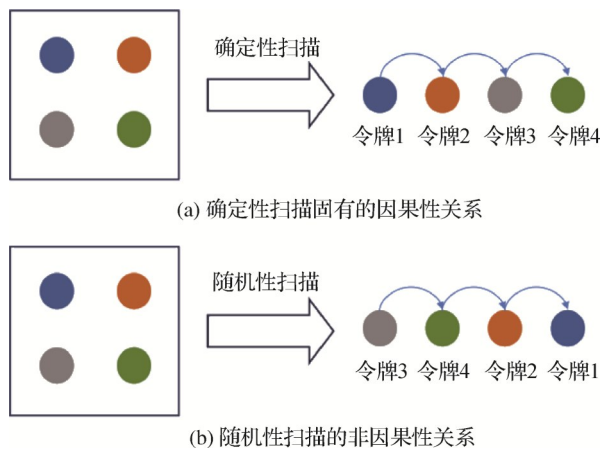


图 1 确定性扫描与随机性扫描对比

Fig. 1 Deterministic scan vs stochastic scan

((a) causal relationship inherent in deterministic scan;

(b) non-causal relationship in stochastic scan)

然而, 这种确定性扫描顺序无法平等地建模图像空间令牌之间的非因果关系。换言之, 图像中任意位置的令牌之间具有相互依赖的关系。例如, 图 1(a) 中, 令牌 1 的扫描结果只能作为其他令牌建模的条件, 而无法反之以其他令牌作为条件来建模令牌 1。为了解决这一问题, 一些研究提出双向扫描策略 (Guo 等, 2025), 以在令牌之间建立任意依赖关系。然而, 这些方法仍然基于确定性扫描, 从而无法对称地建模令牌间的非因果依赖关系, 导致图像复原的性能仍然难以得到进一步提升。

针对上述问题, 本文提出一种非因果的选择性状态空间模型 (selective state space model, selective SSM) 的图像复原方法, 主要贡献如下: 1) 提出一种新的非因果关系扫描方案, 设计了一种随机扫描方法, 以有效替代传统的确定性扫描策略。如图 1(b) 所示, 提出的随机扫描方式能够以等概率建立任意两个令牌之间的依赖关系, 从而突破了确定性扫描方法的因果性限制, 实现了非因果关系建模。2) 融合了图像处理领域的多尺度和局部—全局关系先验知识, 并将其嵌入图像复原网络模型中, 建立了一种具备局部—全局互补特性的非因果选择性状态空间模型 (non-causal Mamba, NCMamba)。3) 实验结果表明, 提出的 NCMamba 模型在图像去模糊、去阴影和去噪等多种图像复原任务中, 均显著超越了传统的 CNN、视觉 Transformer 以及视觉 Mamba 模型, 展现出卓越的性能, 进一步验证了随机扫描策略在建模非因果关系中的有效性, 为复杂依赖关系的高效建模提供了新的解决方案。

1 相关工作

1.1 图像复原

图像复原作为一个经典的病态逆问题, 旨在从退化图像中恢复出其对应的清晰图像, 其典型任务包括图像去噪 (Zhang 等, 2017)、图像超分辨率 (Dong 等, 2016)、图像去模糊 (Pan 等, 2016) 和图像去雨 (Fu 等, 2017; Yang 等, 2017) 等。

近年来, 深度学习技术的飞速发展 (LeCun 等, 2015) 极大地推动了图像复原领域的进步。基于深度学习的图像复原方法主要分为监督学习和无监督学习。无监督学习通过生成模型 (如生成对抗网络 (generative adversarial network, GAN) (Goodfellow 等,

2014)和扩散模型(Ho等,2020))学习数据的先验分布,不依赖于配对的退化-清晰图像数据集,而是依靠单一的清晰数据集(Venkatakrishnan等,2013)或退化数据集(Lehtinen等,2018)。此类方法在推理阶段结合特定的退化过程,例如线性退化场景(Kawar等,2022;Wang等,2022)或非线性退化场景(Pan等,2020),通过持续优化获取最终的图像复原结果。例如,Venkatakrishnan等人(2013)提出一种去噪器模型,可广泛应用于通用图像复原任务。DDNM(denoising diffusion null-space model)(Wang等,2023)通过预训练的扩散模型有效地在特定退化场景下进行图像复原;兰治等人(2023)利用混合双注意力机制生成对抗网络实现图像复原;刘雨轩等人(2023)则通过结构先验信息实现文本图像修复。

监督学习方法则依赖于配对的退化/清晰图像数据,并基于这些数据训练深度学习模型,如卷积神经网络(Dong等,2016)、Transformer(Wang等,2022;陈加保等,2023)、多层感知机(Tu等,2022)、图模型(Li等,2021)和Mamba(Guo等,2025)等,学习退化图像到清晰图像的映射关系。例如,多尺度特征的卷积神经网络(Zamir等,2020b)和多阶段逐步优化的卷积神经网络(Zamir等,2021)通过有效利用图像的多尺度先验知识,显著提升了在图像复原任务中的性能;Uformer模型(Wang等,2022)和SwinIR模型(Liang等,2021)将局部先验信息整合到Transformer中,取得了在图像复原任务中的卓越表现;华夏等人(2024)通过利用局部特征实现了模糊图像的复原;熊巍等人(2023)则实现了基于通道注意力机制的Transformer图像超分辨率;此外,基于恶劣天气条件的图像复原也取得了显著进展(汪文靖等,2024)。

1.2 状态空间模型

选择性状态空间模型(seletive SSM)近年来在自然语言处理和计算机视觉领域受到广泛关注。与Transformer不同,状态空间模型通过固定大小的上下文窗口,以线性复杂度建模全局依赖关系。Gu等人(2022)提出结构化状态空间模型的设计,而Mehta等人(2023)则通过引入门控单元进一步提升了模型的性能。在自适应输入能力方面,Gu和Dao(2024)引入了选择性扫描机制,并提出高效的硬件存储结构实现,进而开发了Mamba模型,该模型在自然语言处理任务中展现了与Transformer相当的性能。

在计算机视觉领域,Mamba模型已成为Trans-

former的有效替代,并在图像分类、物体检测和遥感图像分析等任务中得到了广泛应用(Zhu等,2024;Chen等,2024;Zhao等,2024)。在图像复原领域,Guo等人(2025)引入了基于Mamba模块的通用图像复原框架,而Xiao等人(2024)通过将随机重排技术集成到Transformer中,有效地克服了图像阴影在空间上的非均匀分布难题。

与现有技术方案不同,本文分析了Mamba模型中确定性扫描所引入的因果关系与图像令牌之间的非因果关系之间的矛盾,通过结合局部性、全局性及多尺度等先验知识,设计了一种面向图像复原的高效非因果选择性状态空间模型,提升了复原质量。

2 本文方法

2.1 选择性状态空间模型

状态空间模型是一类深度学习序列模型。状态空间模型通过借助隐状态变量 $\mathbf{h}(t)$ 实现从输入序列 $\mathbf{x}(t)$ 到输出序列 $\mathbf{y}(t)$ 的映射。状态空间模型有5个参数 (Δ, A, B, C, D) ,其定义序列映射包含两个阶段,即

$$\mathbf{h}'(t) = A\mathbf{h}(t) + B\mathbf{x}(t) \quad (1)$$

$$\mathbf{y}(t) = C\mathbf{h}(t) + D\mathbf{x}(t) \quad (2)$$

为了对连续变量进行离散化,首先将连续参数 (Δ, A, B) 采用零阶保持规则(zero-order hold rule)离散化为参数 $(\bar{A}, \bar{B})$,零阶保持规则数学形式为

$$\bar{A} = \exp(\Delta A), \bar{B} = (\Delta A)^{-1}(\exp(\Delta A) - I) \cdot \Delta B \quad (3)$$

经过参数离散化后,连续映射离散化为离散映射,具体为

$$\mathbf{h}_t = \bar{A}\mathbf{h}_{t-1} + \bar{B}\mathbf{x}_t \quad (4)$$

$$\mathbf{y}_t = C\mathbf{h}_t + D\mathbf{x}_t \quad (5)$$

为了增强模型的表征能力,选择性状态空间模型引入了输入自适应的参数设置,即 $B = s_b(\mathbf{x})$, $C = s_c(\mathbf{x})$, $\Delta = \tau_\Delta(P + s_\Delta(\mathbf{x}))$ 。 P 为可学习的参数。

2.2 随机重排函数与逆重排函数

选择性状态空间模型固有的因果性导致令牌之间的交互关系受限于其先后顺序。这与图像块之间非因果性的关系矛盾。针对这个问题,本文提出基于随机重排和逆重排的扫描方法。通过随机重排,引入了扫描的随机性,有效打破了确定性扫描方式带来的令牌间的因果性关系,建模了图像块之间的

非因果交互。设令牌序列为 $X = \{x_i\}_{i=0}^{L-1} \in \mathbf{R}^{L \times c}$, L 为序列长度, 随机重排函数 $S: \mathbf{R}^{L \times c} \rightarrow \mathbf{R}^{L \times c}$ 定义为

$$S(X)_{m'} = X_m \quad (6)$$

式中, m 是索引列表, 如 $[0, 1, \dots, L-1]$, 下标 m' 表示 m 的随机重排, 如 $[3, L-1, \dots, 0]$ 。同时, 逆重排函数 $I_s: \mathbf{R}^{L \times c} \rightarrow \mathbf{R}^{L \times c}$ 为随机重排函数的逆函数, 定义为

$$I_s(S(X)) = X \quad (7)$$

也即, 逆重排函数将随机重排后的令牌序列恢复到原有令牌序列顺序。

2.3 非因果的选择性状态空间模型

在随机重排之后, 令牌序列中的每个令牌都以相同的概率出现在序列的任意位置, 实现了确定性序列顺序到随机性序列顺序的转换。基于随机序列的扫描, 每对令牌之间关系被建模的概率相同, 从而有效实现了令牌间的非因果关系建模。

在选择性状态空间模型处理随机重排的令牌序列后, 令牌的顺序被彻底打乱, 这使得令牌序列无法反映图像的结构信息。因此, 采用逆映射方法将随机重排的令牌序列恢复到原有的顺序。令输入序列为 X , 输出序列为 Y , $SSM()$ 为选择性状态空间模型, 上述过程数学表述为

$$X' = S(X) \quad (8)$$

$$Y' = SSM(X') \quad (9)$$

$$Y = I_s(Y') \quad (10)$$

2.4 局部状态空间模型与全局状态空间模型

为了充分利用图像的局部和全局相关性, 本文提出两种非因果的选择性状态空间模型: 基于局部窗口的局部模型和覆盖整幅图像的全局模型。如图2所示, 选择性状态空间模型可以根据应用区域的不同, 分为局部状态空间模型和全局状态空间模型。

1) 局部状态空间模型。局部状态空间模型建模

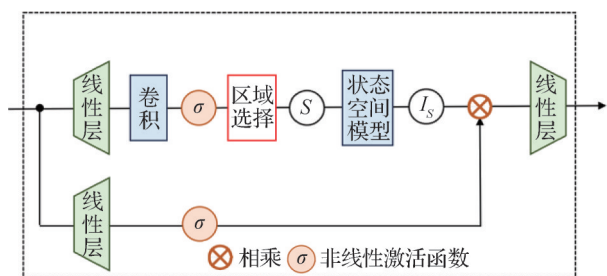


图2 非因果选择性状态空间模型结构

Fig. 2 Structure of non-casual selective state space model

图像局部区域内令牌的交互关系。该模型通过引入尺寸为 $s \times s$ 的局部窗口, 将令牌扫描的范围限制在局部窗口内, 并在不同的局部窗口间并行扫描并共享模型参数。局部窗口设计能够使状态空间模型充分利用图像的局部性先验, 如图3(a)所示, 假设整个图像尺寸为 $H \times W$, 每个局部窗口包含 $s \times s$ 个令牌, 从而在局部窗口内建模令牌之间的相关性。

2) 全局状态空间模型。全局状态空间模型旨在建模整幅图像的令牌间的依赖关系。该模型在整幅图像范围内进行令牌扫描, 建模全局令牌间的相互关系, 以实现状态空间模型对全局依赖关系的有效捕捉, 如图3(b)所示, 全局状态空间模型可以建模整个图像范围 $H \times W$ 内的所有令牌之间的依赖关系。

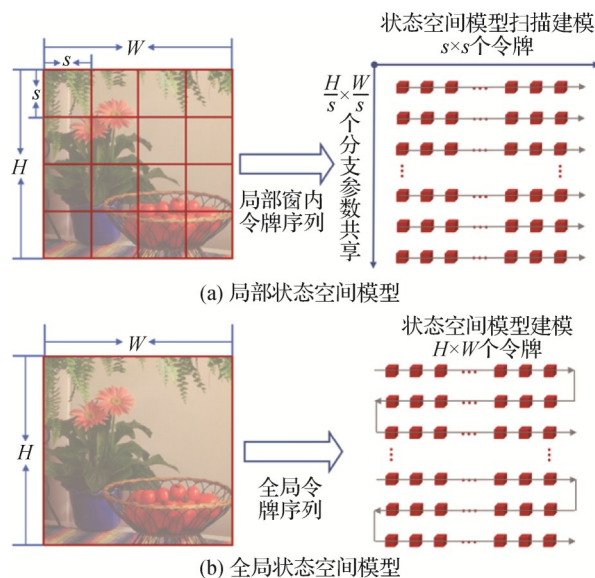


图3 局部状态空间模型与全局状态空间模型示意图

Fig. 3 Illustration of local state space model and global state space model ((a) local state space model; (b) global state space model)

2.5 总体架构

总体架构由非因果的局部和全局选择性状态空间模型交替堆叠构成。模型通过嵌入图像的先验知识, 如多尺度性、局部相关性和全局依赖性, 增强图像处理能力。NCMamba 采用了基于 U-Net 的整体架构。这种架构有效地挖掘图像的多尺度先验, 并能够有效地建模图像的非因果关系。局部非因果选择性状态空间模型—全局非因果选择性状态空间模型的交替堆叠进一步增强了对图像细节和整体内容的分析与复原能力。总体架构如图4所示。

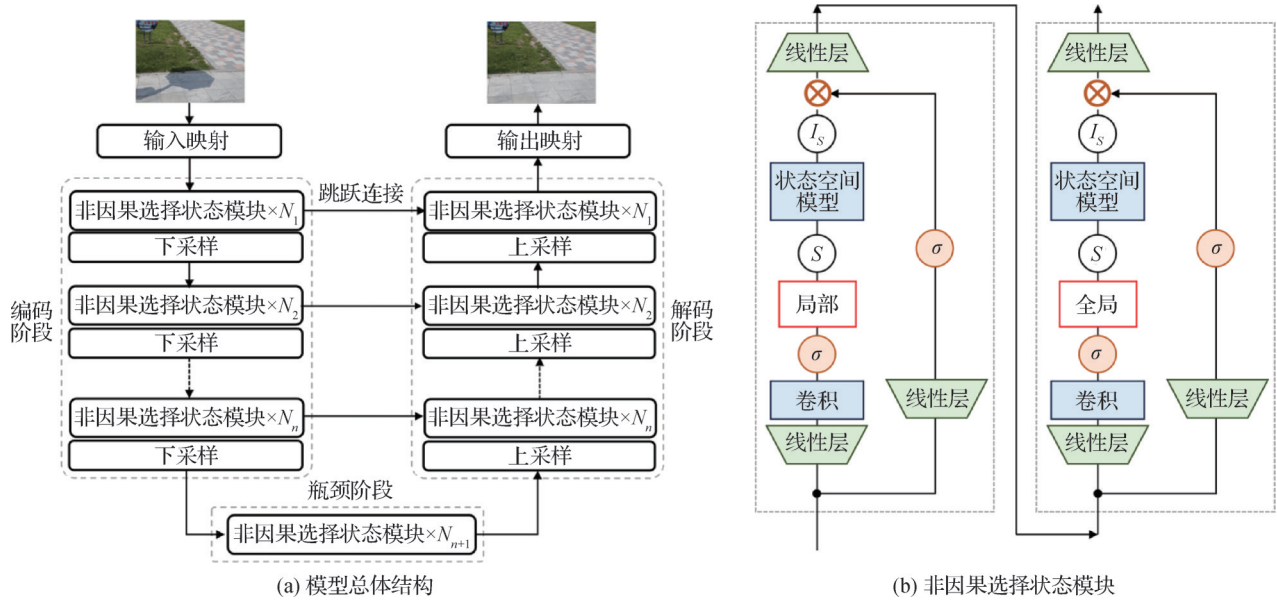


图4 NCMamba 总体架构和核心模块

Fig. 4 Overall architecture of NCMamba ((a) overall architecture; (b) non-casual selective state space model)

2.6 训练与推理

2.6.1 训练阶段

状态空间模型通过引入随机重排突破确定的令牌扫描顺序所带来的因果关系。随机重排使得令牌序列中的每个令牌位置等价,为非因果关系建模奠定了基础。假设特征(或图像)用符号 X, Y 表示,与之对应的随机重排版本用 X', Y' 表示。在训练阶段,非因果状态空间模型经历随机重排映射、状态空间模型建模、逆重排映射3个操作, $NCSSM$ 表示一个非因果状态选择模块,总体过程的数学表达为

$$X' = S(X) \quad (11)$$

$$Y' = SSM(X') \quad (12)$$

$$Y = NCSSM(X; S, I_s) = I_s(Y') \quad (13)$$

通过随机重排,非因果状态空间模型可以以相等的概率建模令牌序列中任意两个令牌之间的相互关系。从期望的意义上,基于随机重排的令牌序列的状态空间模型能够有效地建模非因果关系。同时,训练过程基于随机重排的样本进行,因此这种方法不会增加状态空间模型的训练复杂度。

2.6.2 推理阶段

非因果状态空间模型引入随机重排和逆重排。从贝叶斯推理的角度,图像复原结果应当对随机因子取期望,以得到最优的推理结果, $NCMamba$ 表示非因果 Mamba 模型,其数学表述为

$$NCMamba(X)^{test} = \mathbb{E}_S[NCMamba(X; S)] \quad (14)$$

然而,在模型的深度方向上,每一层的随机重排都是独立的,随机重排的组合可能性随模型深度呈现指数级的增长。直接计算期望值需要枚举所有可能的重排结果,这在实践中成本过高,是不可行的。在实践中,通过蒙特卡洛平均来实现对该期望值的近似,即

$$NCMamba(X)^{test} \approx \frac{1}{M} \sum_{i=1}^M NCMamba(X; S_i) \quad (15)$$

式中, M 为蒙特卡洛采样的样本数。该过程为:将输入 X 以独立随机重排 M 次,经过非因果状态空间模型处理,并计算模型的 M 个输出,然后对输出取平均值以获得期望的蒙特卡洛估计。当 $M \rightarrow \infty$ 时,蒙特卡洛近似值逼近真实的期望值。

为了加速推理,可以利用图形处理器(graphics processing unit, GPU)等硬件加速器并行进行多次前向传播计算。具体操作方法是:将输入数据传输到 GPU,并对输入进行复制 M 次,形成一个由 M 个样本组成的数据批次。同一个数据批次在 GPU 上被并行处理,提高计算效率。在完成一次前向传播后,通过对批次处理结果取平均值,得到期望的蒙特卡洛估计近似。

3 实验结果与分析

3.1 实验设置

为了证明所提出的 NCMamba 模型的有效性,在图像去模糊、去阴影以及去噪多个图像复原任务上进行

了验证。评估方法包括使用定量指标和视觉可视化的方式,与现有方法进行了比较。NCMamba模型包含两种变体:NCMamba ($M = 1$)和NCMamba+ ($M = 8$)。模型的代码实现是基于Python语言和Pytorch深度学习框架,使用4张A100 GPU显卡进行训练和测试。

3.2 实验结果

3.2.1 图像去阴影

图像去阴影任务的评测是基于公开数据集SRD (shadow removal dataset)。为了保持与之前研究的一致性,采用峰值信噪比(peak signal-to-noise ratio, PSNR)、结构相似性(structural similarity index measure, SSIM)和平均绝对误差(mean absolute error, MAE) 3个定量指标来评估方法的性能。所比较的图像去阴影方法包括:Guo 等人(2013)、DeshadowNet(Qu 等,2017)、DSC(direction-aware spatial context)(Hu 等,2020)、DHAN(dual hierarchically aggrega-

tion network)(Cun 等,2020)、Fu 等人(2021)、Jin 等人(2021)、BMNet(bijective mapping network)(Zhu 等,2022)、ShadowFormer(Guo 等,2023a)、Li 等人(2023)、ShadowDiffusion(Guo 等,2023b)和 HomoFormer(Xiao 等,2024)。

表1中的定量指标对比结果表明,NCMamba在PSNR等性能指标上超越了对比的经典方法,其中包括卷积神经网络、Transformer模型和基于扩散模型的方法。具体而言,NCMamba与HomoFormer相比,PSNR提高0.22 dB;与ShadowDiffusion相比,提高0.86 dB;与BMNet相比,提高3.90 dB,综合取得最佳性能。图5中的定性视觉效果对比进一步证实了NCMamba在阴影去除方面以及非阴影区域图像信息的保持方面表现出的优越性,特别是通过红框和蓝框标注的区域可以直观地看出NCMamba具有的视觉效果提升。

表1 基于SRD数据集的图像去阴影定量性能对比
Table 1 Quantitative comparison of image deshadowing on SRD dataset

方法	阴影区域			非阴影区域			全部区域		
	PSNR/dB↑	SSIM↑	MAE↓	PSNR/dB↑	SSIM↑	MAE↓	PSNR/dB↑	SSIM↑	MAE↓
输入图像	18.96	0.871	36.69	31.47	0.975	4.83	18.19	0.829	14.05
Guo 等人(2013)	–	–	29.89	–	–	6.47	–	–	12.60
DeshadowNet	–	–	11.78	–	–	4.84	–	–	6.64
DSC	30.65	0.960	8.62	31.94	0.965	4.41	27.76	0.903	6.71
DHAN	33.67	0.978	8.94	34.79	0.979	4.80	30.51	0.949	5.67
Fu 等人(2021)	32.26	0.966	8.55	31.87	0.945	5.74	28.40	0.893	6.50
Jin 等人(2021)	34.00	0.975	7.70	35.53	0.981	3.65	31.53	0.955	4.65
BMNet	35.05	0.981	6.61	36.02	0.982	3.61	31.69	0.956	4.46
ShadowFormer	36.91	0.989	5.90	36.22	0.989	3.44	32.90	0.958	4.04
Li 等人(2023)	39.33	0.985	6.09	35.61	0.967	2.97	33.17	0.941	3.83
ShadowDiffusion	38.72	<u>0.987</u>	4.98	37.78	0.985	3.44	34.73	<u>0.970</u>	3.63
HomoFormer	38.81	<u>0.987</u>	<u>4.25</u>	39.45	<u>0.988</u>	2.85	35.37	0.972	3.33
NCMamba	39.04	0.983	4.21	<u>39.72</u>	0.982	<u>2.80</u>	<u>35.59</u>	0.958	<u>3.30</u>
NCMamba+	<u>39.06</u>	0.983	4.21	39.75	0.982	2.79	35.62	0.959	3.29

注:加粗、下划线字体分别表示各列最优、次优结果。“–”表示原论文未提供相应数据集评估结果。“↑”和“↓”分别表示值越大越好、值越小越好。

3.2.2 图像去模糊

图像去模糊任务的评测是基于4个图像去模糊标准数据集,其中包括两个合成数据集(GoPro和HIDE(human-aware image deblurring))以及两个真实

世界数据集(RealBlur-R和RealBlur-J)。实验设置与已有的研究一致:NCMamba模型仅在GoPro数据集上进行训练,并在所有4个数据集上测试评估。所比较凸显去模糊方法包括基于卷积神经网络的方

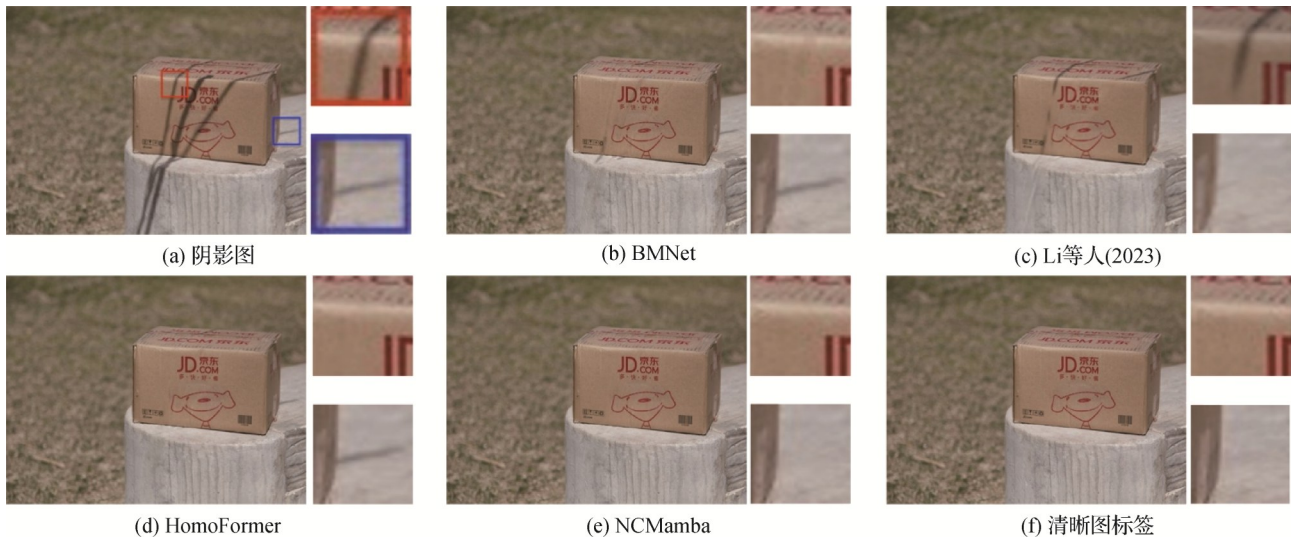


图5 图像去阴影视觉效果对比图

Fig. 5 Visual comparison of image shadow removal

((a) shadowed image; (b) BMNet; (c) Li et al. (2023); (d) HomoFormer; (e) NCMamba; (f) ground truth)

法: Xu 等人(2013)、Nah 等人(2017)、DeblurGAN (Kupyn 等, 2018)、Zhang 等人(2018)、DBGAN (deblur GAN)(Zhang 等, 2020)、MPRNet(multi-stage progressive image restoration)(Zamir 等, 2021); 基于多层感知机的方法: MAXIM(multi-axis multilayer perceptron)(Tu 等, 2022)和基于Transformer的方法: Uformer(Wang 等, 2022)、Restormer(Zamir 等, 2022)。

表2中的定量结果对比显示,NCMamba在PSNR和SSIM指标上超过了对比的基于卷积神经网络的方法,例如MPRNet,并且实现了与最优的Transformer模型(如Uformer)相当的性能。此外,图6中图像去模糊的视觉对比图进一步凸显了NCMamba的优势:特别是图6的车牌数字“2”的复原中,与其他方法相比,NCMamba具有更清晰的图像复原效果。

表2 在GoPro数据集上训练的图像去模糊定量性能对比

Table 2 Quantitative comparison of image deblurring when models is only trained on GoPro dataset

方法	GoPro		HIDE		RealBlur-R		RealBlur-J	
	PSNR/dB↑	SSIM↑	PSNR/dB↑	SSIM↑	PSNR/dB↑	SSIM↑	PSNR/dB↑	SSIM↑
Xu 等人(2013)	21.00	0.741	—	—	34.46	0.937	27.14	0.830
Nah 等人(2017)	29.08	0.914	25.73	0.874	32.51	0.841	27.87	0.827
DeblurGAN	28.70	0.858	24.51	0.871	33.79	0.903	27.97	0.834
Zhang 等人(2018)	29.19	0.931	—	—	35.48	0.947	27.80	0.847
DMPHN	31.20	0.940	29.09	0.924	35.70	0.948	28.42	0.860
DBGAN	31.10	0.942	28.94	0.915	—	—	—	—
MPRNet	32.66	0.959	30.96	0.939	35.99	0.952	28.70	0.873
MAXIM	32.86	<u>0.961</u>	32.83	0.956	35.78	0.947	28.83	0.875
Uformer	32.97	0.967	30.83	<u>0.952</u>	36.22	0.957	29.06	0.884
Restormer	32.92	<u>0.961</u>	<u>31.22</u>	0.942	36.19	0.957	<u>28.96</u>	<u>0.879</u>
NCMamba	<u>33.17</u>	<u>0.961</u>	31.06	0.943	<u>36.20</u>	<u>0.956</u>	28.76	0.873
NCMamba+	33.19	<u>0.961</u>	31.08	0.944	<u>36.20</u>	<u>0.956</u>	<u>28.96</u>	<u>0.879</u>

注:加粗、下划线字体分别表示各列最优、次优结果。“—”表示原论文未提供相应数据集评估结果。



图6 图像去模糊视觉效果对比图

Fig. 6 Visual comparison of image deblurring

((a) blurred images; (b) DBGAN; (c) MPRNet; (d) Uformer; (e) NCMamba; (f) ground truths)

3. 2. 3 图像去噪

图像去噪任务的评测是基于真实噪声场景的图像噪声数据集 SIDD。为了验证 NCMamba 的有效性,选择了 9 种具有代表性的降噪方法进行比较,包括:BM3D (Dabov 等,2007)、RIDNet (Anwar 和 Barnes, 2019)、VDN (Yue 等, 2019)、DANet (Yue 等, 2020)、CycleISP (Zamir 等, 2020a)、MIRNet (Zamir 等, 2020b)、MPRNet (Zamir 等, 2021), Uformer (Wang 等, 2022) 以及 MambaIR (Guo 等, 2025)。

表 3 的定量结果显示,NCMamba 在 PSNR 和 SSIM 两个关键性能指标上,不仅超越了传统的基于卷积神经网络的方法(如 MPRNet),也优于现有的 Transformer 方法(如 Uformer)和最新的 Mamba 方法(MambaIR)。这些结果证明了 NCMamba 在真实噪声环境下具有更优异的图像降噪能力,展示了其在降噪领域的技术领先性和应用潜力。

表 3 基于 SIDD 数据集的图像去噪定量性能对比
Table 3 Quantitative comparison of image denoising on SIDD dataset

方法	SIDD	
	PSNR/dB↑	SSIM↑
BM3D	25.65	0.685
RIDNet	38.71	0.914
VDN	39.28	0.909
DANet	39.47	0.918
CycleISP	39.52	0.957
MIRNet	39.72	<u>0.959</u>
MPRNet	39.71	0.958
Uformer	39.77	<u>0.959</u>
MambaIR	39.89	0.960
NCMamba	<u>39.91</u>	0.960
NCMamba+	39.92	0.960

注:加粗字体、下划线字体分别表示各列最优、次优结果。

3.3 实验分析

3.3.1 消融实验

表4提供了对于所提模型组件的深入消融实验与分析。实验分析基于图像去噪任务(数据集SIDD),目的是评估不同模型组件对最终性能的贡献。所考察的模型组件包括随机重排、双向扫描机制、局部选择性状态空间模型和全局选择性状态空间模型。结果显示,随机重排凭借其独特的非因果建模能力,其性能超过了需要更高计算和存储资源的双向扫描机制。此外,局部选择性状态空间模型和全局选择性状态空间模型的引入都显著提升了图像去噪的效果,PSNR指标分别提升0.04 dB和0.1 dB。

3.3.2 蒙特卡洛近似分析

根据贝叶斯理论,最优的推理方式为模型对随机因子(即随机重排)取数学期望。随机重排的组合的数量随模型深度呈现指数级增长。因此,需使用蒙特卡洛方法通过平均多个独立样本来近似真实的期望值。理论上,当样本数 M 趋于无穷大时,通过蒙特卡洛方法计算出的平均值将逼近期望的真实值。

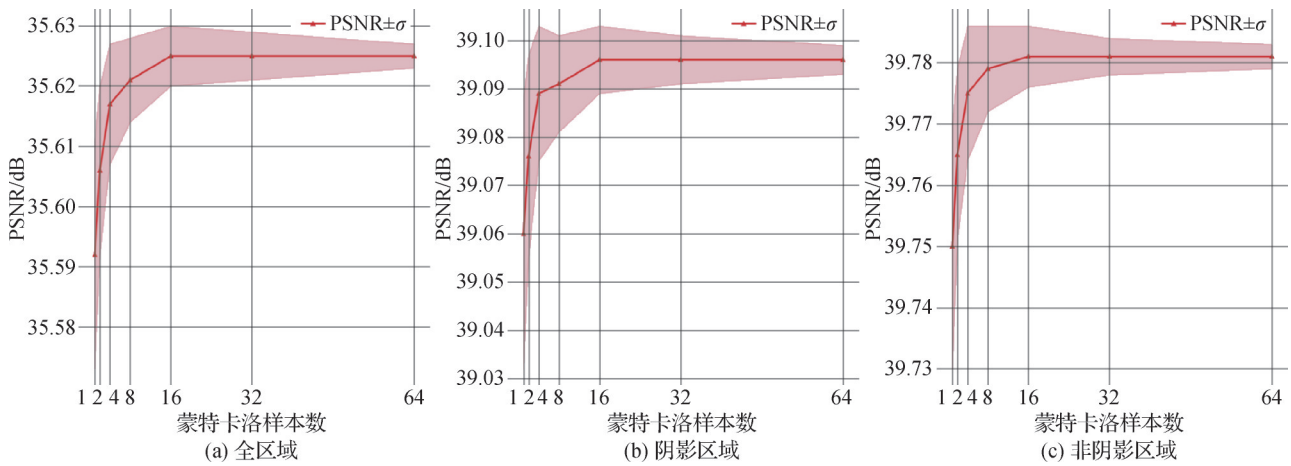


图7 蒙特卡洛样本数对性能(PSNR)和标准差的影响

Fig. 7 Effect of Monte Carlo samples on PSN and standard deviation
((a) all region; (b) shadowed region; (c) non-shadowed region)

3.3.3 模型效率分析

模型效率方面,表5提供了NCMamba与其他方法在参数量和浮点运算次数(floating point operations per second,FLOPs)方面的对比。FLOPs是模型处理 256×256 像素的输入图像所需的运算量。结果表明,NCMamba在保持合理的参数和计算量的同

表4 消融实验的定量结果

模型变体	PSNR/dB
w/o 随机重排	39.80
双向扫描	39.82
w/o 局部	39.87
w/o 全局	39.81
NCMamba	39.91

蒙特卡洛样本数 M 对性能有怎样的影响呢?本小节基于图像去阴影的数据集SRD进行实验探究。图7显示,随着样本数量 M 的增加,图像复原的性能持续提升,主要体现在更高的评价指标和更低的标准差上。这个性能提升现象说明,样本数的增加有助于更准确地估计期望值,进而改进复原的质量。所采用的默认样本数 $M=8$ 能在尽可能节省计算资源的同时,得到优异的结果。这种计算—性能的平衡在实际应用中是至关重要的。它允许在有限的计算预算下达到较好的复原效果。这个发现验证了在资源受限的环境下使用蒙特卡洛方法的可行性和有效性。

时,相比于已有的Transformer模型(如HomoFormer)和卷积神经网络模型(如DHAN),展现出更高的性能。具体而言,与HomoFormer相比,NCMamba在PSNR上提升0.22 dB;与DHAN相比,提升5.08 dB。

这些结果表明,NCMamba不仅在图像复原任务中表现出色,而且也具有优秀的效率表现。通过在

表 5 模型效率对比
Table 5 Comparison of model efficiency

方法	参数量/M	FLOPs/G	PSNR/dB
DHAN	22	113.9	30.51
Fu 等人(2021)	142	104.8	28.40
Jin 等人(2021)	21	113.9	31.53
HomoFormer	18	38.4	35.37
NCMamba	26	49.0	35.59
NCMamba+	26	392.0	35.62

参数和计算量上的合理分配,NCMamba能够在不牺牲性能的前提下,提供更加经济的解决方案。这使得 NCMamba 在需要处理大量数据或在计算资源受限的场景下有特别的价值。

3.3.4 随机重排分析

基于随机重排的随机性扫描使得 Mamba 模型能建模非因果关系。本小节研究重排程度对性能的影响,重排程度具体定义为模型随机重排的层数占总层数的比例。随机重排层沿模型深度方向是均匀选取的,实验分析是基于图像去噪的 SIDD 数据集,不同重排比例对应实验结果如表 6 所示。可以发现,随着重排比例的增加,模型性能整体呈现上升趋势,最佳打乱比例仍为 100%。理论上,随机重排消除 Mamba 建模的因果性,使 Mamba 更加适配图像复原任务,因此完全重排取得最优的结果。

表 6 重排程度的图像去噪性能的影响
Table 6 Effect of degree of random shuffle on image denoising

打乱比例/%	PSNR/dB
0	39.80
25	39.85
50	39.89
75	39.88
100	39.91

4 讨 论

4.1 随机重排和因果建模之间的联系

状态空间模型通常将图像块(或像素)视为令牌,并采用确定性扫描方式将图像展开成令牌序列。

状态空间模型的建模逻辑是基于前 $i-1$ 个令牌预测第 i 个令牌(假设令牌总数为 L),即建模依赖关系 $p(x_i|x_j, j < i \leq L)$,也即只建模了 $x_{j < i} \rightarrow x_i, i \leq L$ 的依赖关系,忽略了 $x_i \rightarrow x_{j < i}, i \leq L$ 的依赖关系。因此,确定性扫描方式会将图像展开成确定排列的令牌序列,这种确定排列的令牌序列强制有 $x_{j < i} \rightarrow x_i, i \leq L$ 的因果性。为了解决这一难点,NCMamba 引入随机重排函数,与确定性扫描方式不同,随机重排函数会将图像展开成乱序排列的令牌序列,状态空间模型基于重排后的序列建模了 $x_j \rightarrow x_i, i \neq j$ 且 $i, j \leq L$ 的关系,消除了对 $j \leq i$ 的因果关系的强制性要求。因此,引入随机重排函数后,状态空间模型可以建模令牌之间非因果关系。

4.2 随机重排的机理与先验

令牌随机重排机制蕴含的图像先验是图像的全局性先验。根据图像的全局性先验,取自图像的任意两个令牌 x_j, x_i , 其中, $i \neq j$ 且 $i, j \leq L$ 之间具有依赖关系,也即 $x_j \rightarrow x_i, i \neq j$ 且 $i, j \leq L$ 。如前所述,令牌序列的随机重排允许建模任意 i, j 令牌 $x_j \rightarrow x_i, i \neq j$ 且 $i, j \leq L$ 之间的关系。因此,图像依赖关系的全局性先验保证了无论令牌顺序如何变化,其来自于同一幅图像的不同令牌的内容总是有关联性的。即,经过随机重排后,虽然令牌是无序的,NCMamba 仍然可以有效地建模令牌内容的依赖关系。

5 结 论

为解决状态空间模型中确定性扫描带来的强制因果建模与图像复原任务中的非因果关系之间的矛盾,本文引入了随机重排函数和逆重排函数,并提出基于随机性扫描的非因果选择性状态空间模型。通过融合多尺度先验、局部关系先验和全局关系先验,本文提出适配的图像复原的 NCMamba 模型。实验结果表明,NCMamba 在图像去阴影、图像去模糊以及图像去噪任务上,相比传统的卷积神经网络和 Transformer 模型,实现了显著的性能提升,验证了所有组件设计的有效性。

尽管取得了一系列成果,NCMamba 仍存在一些不足。例如,NCMamba 虽然建模了令牌之间的非因果关系,但未能充分考虑令牌的提取过程。未来的

研究将综合考虑图像令牌的提取过程及其令牌序列间的非因果关系,以进一步提升模型的泛化能力。

参考文献(References)

- Anwar S and Barnes N. 2019. Real image denoising with feature attention//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 3155-3164 [DOI: 10.1109/ICCV.2019.00325]
- Chen H T, Wang Y H, Guo T Y, Xu C, Deng Y P, Liu Z H, Ma S W, Xu C J, Xu C and Gao W. 2021. Pre-trained image processing transformer//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 12294-12305 [DOI: 10.1109/CVPR46437.2021.01212]
- Chen J B, Xiong B S, Kuang F and Zhang Z Z. 2023. Motion deblurring based on deep feature fusion attention and double-scale. *Journal of Image and Graphics*, 28(12): 3731-3743 (陈加保, 熊邦书, 况发, 章照中. 2023. 深度特征融合注意力与双尺度的运动去模糊. *中国图象图形学报*, 28(12): 3731-3743) [DOI: 10.11834/jig.220931]
- Chen T X, Ye Z, Tan Z T, Gong T, Wu Y, Chu Q, Liu B, Yu N H and Ye J P. 2024. MiM-ISTD: Mamba-in-Mamba for efficient infrared small-target detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62: #5007613 [DOI: 10.1109/TGRS. 2024. 3485721]
- Cun X, Pun C M and Shi C. 2020. Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting GAN//Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York, USA: AAAI: 10680-10687 [DOI: 10.1609/aaai.v34i07.6695]
- Dabov K, Foi A, Katkovnik V and Egiazarian K. 2007. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8): 2080-2095
- Dong C, Loy C C, He K M and Tang X O. 2016. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2): 295-307 [DOI: 10.1109/TPAMI.2015.2439281]
- Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X H, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J and Housby N. 2021. An image is worth 16×16 words: transformers for image recognition at scale//Proceedings of the 9th International Conference on Learning Representations. [s.l.]: ICLR: 1-21
- Fu L, Zhou C Q, Guo Q, Juefei-Xu F, Yu H K, Feng W, Liu Y and Wang S. 2021. Auto-exposure fusion for single-image shadow removal//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 10566-10575 [DOI: 10.1109/CVPR46437.2021.01043]
- Fu X Y, Huang J B, Ding X H, Liao Y H and Paisley J. 2017. Clearing the skies: a deep network architecture for single-image rain removal. *IEEE Transactions on Image Processing*, 26(6): 2944-2956. [DOI: 10.1109/TIP.2017.2691802]
- Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y. 2014. Generative adversarial networks. *Communications of the ACM*, 63(11): 139-144 [DOI: 10.1145/3422622]
- Gu A and Dao T. 2024. Mamba: linear-time sequence modeling with selective state spaces [EB/OL]. [2024-08-07]. <https://arxiv.org/pdf/2312.00752.pdf>
- Gu A, Goel K and Ré C. 2022. Efficiently modeling long sequences with structured state spaces//Proceedings of the 10th International Conference on Learning Representations. [s.l.]: ICLR: 1-27
- Guo H, Li J M, Dai T, Ouyang Z H, Ren X D and Xia S T. 2025. MambaIR: a simple baseline for image restoration with state-space model//Proceedings of the 18th European Conference on Computer Vision. Milan, Italy: Springer: 222-241 [DOI: 10.1007/978-3-031-72649-1_13]
- Guo L Q, Huang S Y, Liu D, Cheng H and Wen B H. 2023a. ShadowFormer: global context helps shadow removal//Proceedings of the 37th AAAI Conference on Artificial Intelligence. Washington, USA: AAAI: 710-718 [DOI: 10.1609/aaai.v37i1.25148]
- Guo L Q, Wang C, Yang W H, Huang S Y, Wang Y F, Pfister H and Wen B H. 2023b. ShadowDiffusion: when degradation prior meets diffusion model for shadow removal//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 14049-14058 [DOI: 10.1109/CVPR52729. 2023.01350]
- Guo R Q, Dai Q Y and Hoiem D. 2013. Paired regions for shadow detection and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(12): 2956-2967 [DOI: 10.1109/TPAMI. 2012.214]
- Ho J, Jain A and Abbeel P. 2020. Denoising diffusion probabilistic models//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada: ACM: #574
- Hu X W, Fu C W, Zhu L, Qin J and Heng P A. 2020. Direction-aware spatial context features for shadow detection and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(11): 2795-2808 [DOI: 10.1109/TPAMI.2019.2919616]
- Hua X, Shu T, Li M X, Shi Y and Hong H Y. 2024. Nonlocal feature representation-embedded blurred image restoration. *Journal of Image and Graphics*, 29(10): 3033-3046 (华夏, 舒婷, 李明欣, 时愈, 洪汉玉. 2024. 融合非局部特征表示的模糊图像复原. *中国图象图形学报*, 29(10): 3033-3046) [DOI: 10.11834/jig. 230735]
- Jin Y Y, Sharma A and Tan R T. 2021. DC-ShadowNet: single-image hard and soft shadow removal using unsupervised domain-classifier guided network//Proceedings of 2021 IEEE/CVF International Con-

- ference on Computer Vision. Montreal, Canada: IEEE: 5007-5016 [DOI: 10.1109/ICCV48922.2021.00498]
- Kawar B, Elad M, Ermon S and Song J M. 2022. Denoising diffusion restoration models//Proceedings of the 36th International Conference on Neural Information Processing Systems. New Orleans, USA: ACM: #1714
- Kupyn O, Budzan V, Mykhailych M, Mishkin D and Matas J. 2018. DeblurGAN: blind motion deblurring using conditional adversarial networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 8183-8192 [DOI: 10.1109/CVPR.2018.00854]
- Lan Z, Yan C P, Li H and Zheng Y D. 2023. HDA-GAN: hybrid dual attention generative adversarial network for image inpainting. *Journal of Image and Graphics*, 28(11): 3440-3452 (兰治, 严彩萍, 李红, 郑雅丹. 2023. 混合双注意力机制生成对抗网络的图像修复模型. *中国图象图形学报*, 28(11): 3440-3452) [DOI: 10.11834/jig.220919]
- LeCun Y, Bengio Y and Hinton G. 2015. Deep learning. *Nature*, 521(7553): 436-444 [DOI: 10.1038/nature14539]
- Lehtinen J, Munkberg J, Hasselgren J, Laine S, Karras T, Aittala M and Aila T. 2018. Noise2Noise: learning image restoration without clean data//Proceedings of the 35th International Conference on Machine Learning. Stockholm, Sweden: ICML: 2965-2974
- Li X G, Guo Q, Abdelfattah R, Lin D, Feng W, Tsang I and Wang S. 2023. Leveraging inpainting for single-image shadow removal//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 13009-13018 [DOI: 10.1109/ICCV51070.2023.01200]
- Li Y, Fu X Y and Zha Z J. 2021. Cross-patch graph convolutional network for image denoising//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 4631-4640 [DOI: 10.1109/ICCV48922.2021.00461]
- Liang J Y, Cao J Z, Sun G L, Zhang K, Van Gool L and Timofte R. 2021. SwinIR: image restoration using swin transformer//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision Workshops. Montreal, Canada: IEEE: 1833-1844 [DOI: 10.1109/ICCVW54120.2021.00210]
- Liu Y X, Zhao Q J, Pan F, Gao D G and Pubu D Z. 2023. Structure prior guided text image inpainting model. *Journal of Image and Graphics*, 28(12): 3699-3712 (刘雨轩, 赵启军, 潘帆, 高定国, 普布旦增. 2023. 结构先验指导的文本图像修复模型. *中国图象图形学报*, 28(12): 3699-3712) [DOI: 10.11834/jig.220960]
- Mehta H, Gupta A, Cutkosky A and Neyshabur B. 2023. Long range language modeling via gated state spaces//Proceedings of the 11th International Conference on Learning Representations. Kigali, Rwanda: ICLR: 1-20
- Nah S, Kim T H and Lee K M. 2017. Deep multi-scale convolutional neural network for dynamic scene deblurring//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 257-265 [DOI: 10.1109/CVPR.2017.35]
- Pan J S, Sun D Q, Pfister H and Yang M H. 2016. Blind image deblurring using dark channel prior//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 1628-1636 [DOI: 10.1109/CVPR.2016.180]
- Pan X G, Zhan X H, Dai B, Lin D H, Loy C C and Luo P. 2020. Exploiting deep generative prior for versatile image restoration and manipulation//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 262-277 [DOI: 10.1007/978-3-030-58536-5_16]
- Qu L Q, Tian J D, He S F, Tang Y D and Lau R W H. 2017. Deshad-owNet: a multi-context embedding deep network for shadow removal//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 2308-2316 [DOI: 10.1109/CVPR.2017.248]
- Tu Z Z, Talebi H, Zhang H, Yang F, Milanfar P, Bovik A and Li Y X. 2022. MAXIM: multi-axis MLP for image processing//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 5759-5770 [DOI: 10.1109/CVPR52688.2022.00568]
- Venkatakrishnan S V, Bouman C A and Wohlberg B. 2013. Plug-and-play priors for model based reconstruction//Proceedings of 2013 IEEE Global Conference on Signal and Information Processing. Austin, USA: IEEE: 945-948 [DOI: 10.1109/GlobalSIP.2013.6737048]
- Wang W J, Yang W H, Fang Y M, Huang H and Liu J Y. 2024. Visual perception and understanding in degraded scenarios. *Journal of Image and Graphics*, 29(6): 1667-1684 (汪文靖, 杨文瀚, 方玉明, 黄华, 刘家瑛. 2024. 恶劣场景下视觉感知与理解综述. *中国图象图形学报*, 29(6): 1667-1684) [DOI: 10.11834/jig.240041]
- Wang Y H, Yu J W and Zhang J. 2023. Zero-shot image restoration using denoising diffusion null-space model//Proceedings of the 11th International Conference on Learning Representations. Kigali, Rwanda: ICLR: 1-31
- Wang Z D, Cun X, Bao J M, Zhou W G, Liu J Z and Li H Q. 2022. Uformer: a general U-shaped transformer for image restoration//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 17662-17672 [DOI: 10.1109/CVPR52688.2022.01716]
- Xiao J, Fu X Y, Zhu Y R, Li D, Huang J, Zhu K and Zha Z J. 2024. HomoFormer: homogenized transformer for image shadow removal//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 25617-25626 [DOI: 10.1109/CVPR52733.2024.02420]
- Xiong W, Xiong C Y, Gao Z R, Chen W Q, Zheng R H and Tian J W. 2023. Image super-resolution with channel-attention-embedded transformer. *Journal of Image and Graphics*, 28(12): 3744-3757

- (熊巍, 熊承义, 高志荣, 陈文旗, 郑瑞华, 田金文. 2023. 通道注意力嵌入的Transformer图像超分辨率重构. 中国图象图形学报, 28(12): 3744-3757 [DOI: 10.11834/jig.221033]
- Xu L, Zheng S C and Jia J Y. 2013. Unnatural L0 sparse representation for natural image deblurring//Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, USA: IEEE: 1107-1114 [DOI: 10.1109/CVPR.2013.147]
- Yang W H, Tan R T, Feng J S, Liu J Y, Guo Z M and Yan S C. 2017. Deep joint rain detection and removal from a single image//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 1685-1694 [DOI: 10.1109/CVPR.2017.183]
- Yue Z S, Yong H W, Zhao Q, Zhang L and Meng D Y. 2019. Variational denoising network: toward blind noise modeling and removal//Proceedings of the 33rd International Conference on Neural Information Processing Systems. [s.l.]: ACM: #151
- Yue Z S, Zhao Q, Zhang L and Meng D Y. 2020. Dual adversarial network: toward real-world noise removal and noise generation//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 41-58 [DOI: 10.1007/978-3-030-58607-2_3]
- Zamir S W, Arora A, Khan S, Hayat M, Khan F S and Yang M H. 2022. Restormer: efficient transformer for high-resolution image restoration//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 5718-5729 [DOI: 10.1109/CVPR52688.2022.00564]
- Zamir S W, Arora A, Khan S, Hayat M, Khan F S, Yang M H and Shao L. 2020a. CycleISP: real image restoration via improved data synthesis//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 2693-2702 [DOI: 10.1109/CVPR42600.2020.00277]
- Zamir S W, Arora A, Khan S, Hayat M, Khan F S, Yang M H and Shao L. 2020b. Learning enriched features for real image restoration and enhancement//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 492-511 [DOI: 10.1007/978-3-030-58595-2_30]
- Zamir S W, Arora A, Khan S, Hayat M, Khan F S, Yang M H and Shao L. 2021. Multi-stage progressive image restoration//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 14816-14826 [DOI: 10.1109/CVPR46437.2021.01458]
- Zhang J W, Pan J S, Ren J, Song Y B, Bao L C, Lau R W H and Yang M H. 2018. Dynamic scene deblurring using spatially variant recurrent neural networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 2521-2529 [DOI: 10.1109/CVPR.2018.00267]
- Zhang K, Zuo W M, Chen Y J, Meng D Y and Zhang L. 2017. Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising. IEEE Transactions on Image Processing, 26(7): 3142-3155 [DOI: 10.1109/TIP.2017.2662206]
- Zhang K H, Luo W H, Zhong Y R, Ma L, Stenger B, Liu W and Li H D. 2020. Deblurring by realistic blurring//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 2734-2743 [DOI: 10.1109/CVPR42600.2020.00281]
- Zhao S J, Chen H, Zhang X L, Xiao P F, Bai L and Ouyang W L. 2024. RS-Mamba for large remote sensing image dense prediction. IEEE Transactions on Geoscience and Remote Sensing, 62: #5633314 [DOI: 10.1109/TGRS.2024.3425540]
- Zhu L H, Liao B C, Zhang Q, Wang X L, Liu W Y and Wang X G. 2024. Vision Mamba: efficient visual representation learning with bidirectional state space model//Proceedings of the 41st International Conference on Machine Learning. Vienna, Austria: ACM: #2584
- Zhu Y R, Huang J, Fu X Y, Zhao F, Sun Q B and Zha Z J. 2022. Bijective mapping network for shadow removal//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 5617-5626 [DOI: 10.1109/CVPR52688.2022.00554]

作者简介

肖杰,男,博士研究生,主要研究方向为图像复原。

E-mail: ustchbxj@mail.ustc.edu.cn

查正军,通信作者,男,教授,主要研究方向为人工智能、计算机视觉和模式识别。E-mail: zhazj@ustc.edu.cn

傅雪阳,男,副教授,主要研究方向为计算机视觉、图像处理 and 人工智能。E-mail: xyfu@ustc.edu.cn

范子豪,男,本科生,主要研究方向为图像复原。

E-mail: zh211050@mail.ustc.edu.cn

李东,男,博士研究生,主要研究方向为图像复原和图像鉴别。E-mail: dongli6@mail.ustc.edu.cn