

文章编号: 1007-2780(2025)06-0881-14

基于 Mamba-2 的视频快照压缩成像重构方法

石敦攀^{1,2,3,4}, 徐伟^{1,3,4*}, 朴永杰^{1,3,4}, 方应红^{1,3,4}, 籍浩林^{1,3,4}, 李鹏飞^{1,2,3,4}

(1. 中国科学院 长春光学精密机械与物理研究所, 吉林 长春 130033;

2. 中国科学院大学, 北京 100049;

3. 中国科学院 天基动态快速光学成像技术重点实验室, 吉林 长春 130033;

4. 吉林省航天先进光学成像技术重点实验室, 吉林 长春 130033)

摘要: 视频快照压缩成像 (SCI) 是一种新型的成像技术, 通过在单个曝光时间内使用一个二维探测器捕获三维视频数据, 然后采用合适的算法重建原始的视频数据。尽管目前的许多算法在视频 SCI 的重建任务中有着非常出色的表现, 但它们重建质量的提升往往需要以牺牲重建速度为代价, 使算法的实时性大幅降低。为兼顾重建质量与重建速度, 本文提出了一种基于 Mamba-2 的端到端深度视频 SCI 重构方法——M2BA-SCI。M2BA-SCI 网络由预处理模块、token 生成块、Mamba 注意力块和视频重建块组成, 其中 Mamba 注意力块主要由 Mamba-2 线性注意力块和前馈神经网络构成。在模拟和真实视频数据集上的大量实验表明, M2BA-SCI 与先前算法相比取得了更为均衡的效果, 在提高重建质量的同时仍保持较快的重建速度。在基准灰度视频数据集中, 平均 PSNR 为 34.85, 平均 SSIM 为 0.966, 运行时间为 0.23 s。在基准彩色视频数据集上的平均 PSNR 为 36.21, 平均 SSIM 为 0.963, 运行时间为 1.03 s。M2BA-SCI 为视频 SCI 重建带来了新的思路, 为基于 Mamba 模型设计出更高重建质量的算法提供了参考。

关键词: 视频快照压缩成像; 压缩感知; Mamba-2; 深度学习

中图分类号: TP391 文献标识码: A doi: 10.37188/CJLCD.2024-0356 CSTR: 32172.14.CJLCD.2024-0356

Reconstruction method of video snapshot compressive imaging based on Mamba-2

SHI Dunpan^{1,2,3,4}, XU Wei^{1,3,4*}, PIAO Yongjie^{1,3,4}, FANG Yinghong^{1,3,4}, JI Haolin^{1,3,4}, LI Pengfei^{1,2,3,4}

(1. Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China;

3. Key Laboratory of Space-based Dynamic & Rapid Optical Imaging Technology, Chinese Academy of Sciences, Changchun 130033, China;

4. Jilin Provincial Key Laboratory of Aerospace Advanced Optical Imaging Technology, Changchun 130033, China)

Abstract: Video snapshot compressive imaging (SCI) is a novel imaging technique. It captures three-

收稿日期: 2024-12-27; 修订日期: 2025-01-27.

基金项目: 国家重点研发计划 (No. 2022YFB3705702)

Supported by National Key Research and Development Program of China (No. 2022YFB3705702)

*通信联系人, E-mail: xwciomp@126.com

dimensional video data using a two-dimensional detector within a single exposure time and then reconstructs the original video data with appropriate algorithms. Although many current algorithms have outstanding performance in the reconstruction tasks of video SCI, the improvement of their reconstruction quality often comes at the cost of sacrificing the reconstruction speed, which significantly reduces the real-time performance of the algorithms. To balance reconstruction quality and speed, this paper proposes an end-to-end deep video SCI reconstruction method based on Mamba-2, namely M2BA-SCI. The M2BA-SCI network consists of a preprocessing module, a token generation block, Mamba attention blocks, and a video reconstruction block. Among them, the Mamba attention blocks are mainly composed of Mamba-2 linear attention blocks and feed-forward neural networks. A large number of experiments on simulated and real video datasets show that M2BA-SCI achieves a more balanced effect compared with previous algorithms. It maintains a relatively fast reconstruction speed while improving the reconstruction quality. In the benchmark grayscale video dataset, the average PSNR is 34.85, the average SSIM is 0.966, and the running time is 0.23 s. In the benchmark color video dataset, the average PSNR is 36.21, the average SSIM is 0.963, and the running time is 1.03 s. M2BA-SCI brings new ideas to video SCI reconstruction and provides a reference for designing algorithms with higher reconstruction quality based on the Mamba model.

Key words: video snapshot compressive imaging; compressive sensing; Mamba-2; deep learning

1 引 言

在当今的数字时代,视频内容的采集和处理速率已成为研究和应用的热点问题。特别是对于记录高速动态事件,如化学反应、生物医学动态监测以及体育运动分析等,高帧率成像技术的需求愈发增长。然而,由于硬件限制,传统的高帧率摄像机在成本、尺寸以及功耗方面存在较大挑战。为了应对这些挑战,视频快照压缩成像(Snapshot Compressive Imaging, SCI)^[1]应运而生,它利用压缩感知(Compressive Sensing, CS)^[2]理论和先进的光学编码技术,具备从单个或少量二维图像中重建高速连续视频帧的能力,实现了对高维数据的有效获取。

目前已经出现了各种视频 SCI 系统^[3-7]。随着计算机科学的不断发展,国内外科研工作者也提出了各种各样的视频重建方法。重构算法可以分为基于模型的方法和基于深度学习的方法。由于单像素摄像机重构是一个病态问题,基于模型的方法通常使用视频的先验知识来解决这个逆问题,例如基于稀疏性的变换域^[8]、字典学习^[9-10]、总变分(Total Variation, TV)^[11]和高斯混合模型(Gaussian Mixture Models, GMM)^[12-13]。然而,这些模型将视频视为图像集,严重依赖先验

知识,无法充分利用视频数据的时空相关性。在此基础上,Liu 等人提出了一种利用视频帧的非局部自相似性对空间相似块进行聚类的解压缩 SCI 方法(DeSCI),加权核范数最小化(Weighted Nuclear Norm Minimization, WNNM)被用作图像块分组的低秩正则化方法^[14-15]。该方法获得了令人印象深刻的重建结果,但是这类方法在处理大规模数据或要求实时重建的应用场景中,往往受限于其计算复杂度和重建时间的需求。为了提高重建质量和运行速度,PnP-FFDNet^[16]和PnP-FastDVDnet^[17]将预先训练好的去噪网络集成到迭代优化算法中,从而形成即插即用(Plug-and-Play, PnP)框架。最新的在线 PnP^[18]可以在 PnP 迭代中自适应地更新网络的参数,使得去噪网络更适用于 SCI 重建任务中的所需数据。尽管如此,它们仍然需要在大规模数据集上进行长时间的重建,其中,PnP-FastDVDNet 需要数小时才能从高压缩比的单个压缩测量中重建高清视频。近年来,随着深度学习的蓬勃发展,研究人员构建了许多基于深度学习的模型。BIRNAT^[19]首次使用双向递归神经网络和生成对抗方法超越了基于模型的方法 DeSCI。MetaSCI^[20]对模型适应不同掩码进行了一些探索,显著减少了模型训练时间。DUN-3DUnet^[21]和 ELP-Unford^[22]将迭代优

化思想与深度学习模型相结合,进一步提高了重建质量。STFormer^[23]是第一个具有空间自注意分支和时间自注意分支的最先进(SOTA)的基于视觉Transformer的视频SCI重构算法。

尽管目前的许多算法在视频SCI的重建任务中有着非常出色的表现,但它们的计算复杂度限制了其效率和性能,特别是在处理远程视觉依赖关系(如高分辨率图像)时,在速度和内存使用方面提出了挑战^[24]。而重建速度快的方法,由于其模块过于单一,特征信息获取不全,最终使得重建质量普遍不高。最近,自然语言处理(NLP)领域的一种新型状态空间模型(State Space Model, SSM)^[25-27]——Mamba^[25],已经成为一种非常有前途的具有线性复杂性的长序列建模方法,它在图像和视频分类以及医学图像分割方面表现良好^[28]。受此启发,本文利用Mamba模型进行序列混合建模,捕捉空间全局上下文的信息,从而建立一个高效的视频SCI重建网络。本文所做贡献如下:

(1)对Mamba原理进行总结与分析,在此基础上,最终成功将Mamba-2模型首次应用于视频SCI重建任务中。

(2)构建了基于Mamba-2的端到端深度视频SCI重建网络,将其命名为M2BA-SCI(Mamba-2 Based Attention Snapshot Compressive Imaging),该网络可以很好地应用于多种视频SCI重建任务。

在模拟和真实视频数据集上的大量实验表明,与以前的算法相比,M2BA-SCI模型取得了更为均衡的效果,在提高重建质量的同时,仍然保持较快的重建速度。

2 相关工作

2.1 视频快照压缩成像

如图1所示,视频SCI是一个硬件编码器加软件解码器的集成系统。对于硬件编码器部分,原始视频的每一帧由不同的掩码编码,然后由二维相机对一系列编码帧进行逐像素求和以生成压缩测量。如此,视频SCI便能在光域成像过程中实现高效压缩,提高视频存储和传输的效率。掩码通常由数字微镜装置(DMD)或空间光调制器(SLM)生成,压缩测量值通常由CCD或CMOS

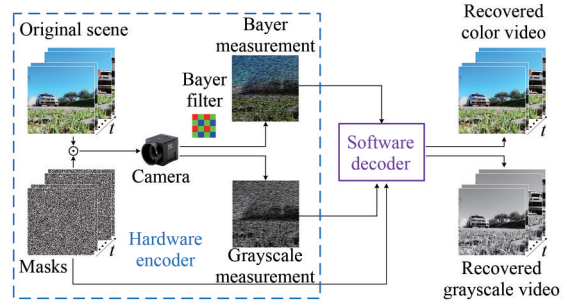


图1 视频SCI系统组成

Fig. 1 Composition of the video SCI system

相机捕获。此外,对于彩色视频SCI^[29],通常在传感器阵列前设置拜耳滤波器,捕捉不同像素点的红、绿、蓝等不同颜色分量,最终由相机输出调制后的拜耳测量值。对于软件解码器部分,通常将测量和掩码输入重构算法以恢复所需的高速视频。

在硬件编码器部分,首先对三维视频数据进行多个掩码调制。然后,通过跨时间维度的积分,由二维探测器捕获编码后的高速场景。假设一个 B 帧的视频数据 $X \in R^{n_x \times n_y \times B}$ 由 B 个不同的掩码矩阵 $M \in R^{n_x \times n_y \times B}$ 进行调制。二维探测器捕获的单帧快照压缩测量值 $Y \in R^{n_x \times n_y}$ 可以表示为:

$$Y = \sum_{b=1}^B X_b \odot M_b + Z, \quad (1)$$

其中: \odot 表示矩阵的Hadamard积,即逐元素乘积; $X_b = X(:, :, b) \in R^{n_x \times n_y}$ 表示第 b 帧视频数据; $M_b = M(:, :, b) \in R^{n_x \times n_y}$ 表示第 b 帧视频数据的调制掩码; $Z \in R^{n_x \times n_y}$ 表示测量噪声。

使用矢量算子,做如下定义:

$$\begin{cases} y = \text{vec}(Y) \in R^{n_x \times n_y} \\ z = \text{vec}(Z) \in R^{n_x \times n_y} \\ x_b = \text{vec}(X_b) \in R^{n_x \times n_y} \end{cases}, \quad (2)$$

由此,原始数据可以表示为:

$$x = [x_1^T, x_2^T, \dots, x_B^T] \in R^{n_x \times n_y \times B}. \quad (3)$$

然后,定义:

$$D_b = \text{diag}[\text{vec}(M_b)] \in R^{n_x \times n_y \times n_x \times n_y}, \quad (4)$$

其中, D_b 是一个对角矩阵且对角元素由 $\text{vec}(M_b)$ 填充。令:

$$H = [D_1, D_2, \dots, D_B] \in R^{n_x \times n_y \times B}. \quad (5)$$

基于以上定义,式(1)的矢量表达式为:

$$y = Hx + z. \quad (6)$$

视频 SCI 系统经由硬件编码器获得单帧测量值 y 以及测量矩阵 H 后,将它们输入设计的软件解码器,即可求解视频数据 x ,完成视频重建任务。

2.2 Mamba

状态空间模型(SSM)代表了序列建模的架构范例,擅长管理长依赖 token。尽管最初在训练中面临挑战,但由于其计算和内存强度,最近的进展显著改善了这些问题,使深度 SSM 成为 CNN 和 Transformer 的强大竞争对手^[30]。特别是,结构化状态空间序列(Structured State Space Sequence, S4)模型^[31]引入了一种高效的正态加低秩(Normal Plus Low-Rank, NPLR)表示,利用 Woodbury 恒等式来加速矩阵求逆,从而简化了卷积核计算。在此基础上,Mamba^[25]通过将特定于输入的参数化与可扩展的硬件优化计算方法结合起来,进一步完善了 SSM,在处理跨语言和基因组的大量序列方面实现了前所未有的效率和简单性。

S4 和 Mamba 基于 SSM,代表深度学习中的一类序列模型,其灵感来自连续系统。连续系统通过中间潜在状态 $h(t) \in R^N$ 将一维输入序列 $x(t) \in R$ 映射到输出 $y(t) \in R$:

$$\begin{cases} h'(t) = Ah(t) + Bx(t) \\ y(t) = C^T h(t) \end{cases}, \quad (7)$$

其中: $h(t)$ 的导数 $h'(t)$ 表示状态的变化量,系统使用矩阵 $A \in R^{N \times N}$ 作为演化参数, $B, C \in R^{N \times 1}$ 作为投影参数, $h(t)$ 可以表示为任意给定时间 t 的潜在状态表示, $h'(t)$ 其实就是下一个状态的表示。

S4 和 Mamba 是连续系统的离散版本,其中引入一个时间尺度参数 Δ ,将连续参数 A, B 转换为离散参数 \bar{A}, \bar{B} 。常用的转换方法是零阶保持,定义如下:

$$\begin{cases} \bar{A} = \exp(\Delta A) \\ \bar{B} = (\Delta A)^{-1}(\exp(\Delta A) - I) \cdot \Delta B \end{cases}. \quad (8)$$

由此得到离散后步长为 Δ 的模型:

$$\begin{cases} h_t = \bar{A} h_{t-1} + \bar{B} x_t \\ y_t = C^T h_t \end{cases}. \quad (9)$$

最后,模型通过全局卷积计算输出:

$$\begin{cases} \bar{K} = (C\bar{B}, C\bar{A}\bar{B}, \dots, C\bar{A}^{M-1}\bar{B}) \\ y = x * \bar{K} \end{cases}, \quad (10)$$

其中: M 是输入序列 x 的长度, $\bar{K} \in R^M$ 是结构化卷积核。

尽管 SSM 框架采用了硬件感知的选择性扫描算法来促进并行计算,但在有效利用硬件内的矩阵计算单元方面遇到了限制。这种低效源于其不依赖矩阵运算作为基本方法。此外,Mamba 的硬件感知算法计算核心完全驻留在 GPU 的 SRAM 中,而该 SRAM 的容量通常有限,这限制了模型计算中的特征维度,从而阻碍了模型在高维空间中的运行能力。

从计算效率的角度出发,Gu 等人提出 Mamba-2^[32],开创性地将矩阵 A 简化为标量形式,为 SSM 矩阵化计算开辟全新路径,揭示了 SSM 变体与掩码注意力机制之间的对偶关联,并顺势借助半可分离矩阵特性解锁线性注意力计算模式,提出了结构化状态空间对偶性(State Space Duality, SSD)理论框架。在此全新架构下,SSM 以式(11)所示的简洁形式呈现:

$$Y = \text{SSM}(A, B, C)(X) = MX. \quad (11)$$

通过精简矩阵运算实现了计算效率的飞跃式提升,同时揭开了 SSM 与线性注意力机制的内在等价关系:

$$Y = L \circ CB^T X = L \circ QK^T V. \quad (12)$$

2.3 Mamba 块

Mamba-1 和 Mamba-2 的模块结构如图 2 所示。Mamba-1 的设计以 SSM 为中心,其中选择性 SSM 层的任务是执行从输入序列 X 到 Y 的映射。在这种设计中,参数 (A, B, C) 被视为辅助参数,是 SSM 输入的函数,因此,定义 (A, B, C) 的线

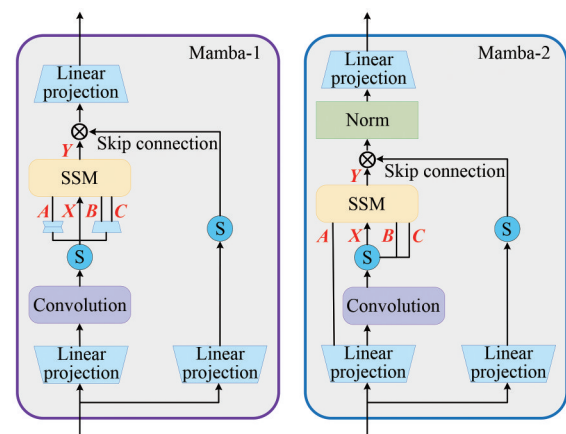


图 2 Mamba-1 和 Mamba-2 模块结构
Fig. 2 Structures of Mamba-1 and Mamba-2 module

性映射发生在创建 X 的初始线性映射之后。而在 Mamba-2 中,则是引入了 SSD 层来创建从 (X, A, B, C) 到 Y 的映射。其实现方式是在块的起始位置使用单个投射来并行生成 (X, A, B, C) , 与标准注意力架构相似, X, B, C 对应并行创建的 Q, K, V 投影。

也就是说,通过移除序列线性投影, Mamba-2 模块在 Mamba-1 模块的基础上进行了简化,这能使 SSD 结构的计算速度超过 Mamba-1 的并行选择式扫描。此外,为了提升训练稳定性, Mamba-2 还在跳跃连接之后添加了一个归一化层。其中,跳跃连接用以鼓励特征复用以及缓解常在模型训练过程中发生的性能下降问题。

3 方 法

3.1 模型架构

本文提出的视频重构模型 M2BA-SCI 网络架构如图 3 所示,主要由 3 部分组成。在图 3(a)所示的预处理阶段,本文参考 RevSCI^[33]、GAP-net^[34] 和 Two-stage^[35] 设计了输入初始化模块。考虑到测量帧 Y 是一个二维矩阵,首先对原始测量帧进行归一化,然后将掩码和归一化测量相结合,产

生调制帧的粗略估计,以达到减轻学习成像系统前向算子的负担的目的,如式(13)所示:

$$\bar{Y} = \frac{Y}{\sum_{b=1}^B M_b}, x_{init} = \bar{Y} \odot M \in R^{n_x \times n_y \times IC \times B}, \quad (13)$$

其中: \odot 表示矩阵逐元素乘积, $IC=1$ 或 3 表示输入的通道数, B 表示压缩比。

随后, x_{init} 被输入到图 3(b) 部分,这也是 M2BA-SCI 网络模型中最重要的部分,主要由 token 生成(TG)块^[33]、Mamba 注意力(MA)块和视频重建(VR)块^[33]组成。 x_{init} 首先经过 TG 块以获得一系列连续的 token,如图 4(a)所示, TG 块由 5 个三维卷积层组成,每个层后面都有一个 LeakyReLU 激活函数^[36]。通过卷积层中的步长和特征映射,最终生成的 token 数量为 $n_x/2 \times n_y/2 \times B$, 每个 token 的特征维度为 C 。与大多数 token 生成方法不同,本文使用的方法不将 x_{init} 划分成不重叠的部分,而是使用三维卷积进行特征映射,然后将特征图的每个点视为一个 token。这有利于减少局部细节丢失的现象。并且通过堆叠多个卷积层,能够逐渐提取出不同尺度的特征,从而捕捉到视频数据的细节和上下文信息^[23]。本文设计的 Mamba 注意力块(MA)可以很好地

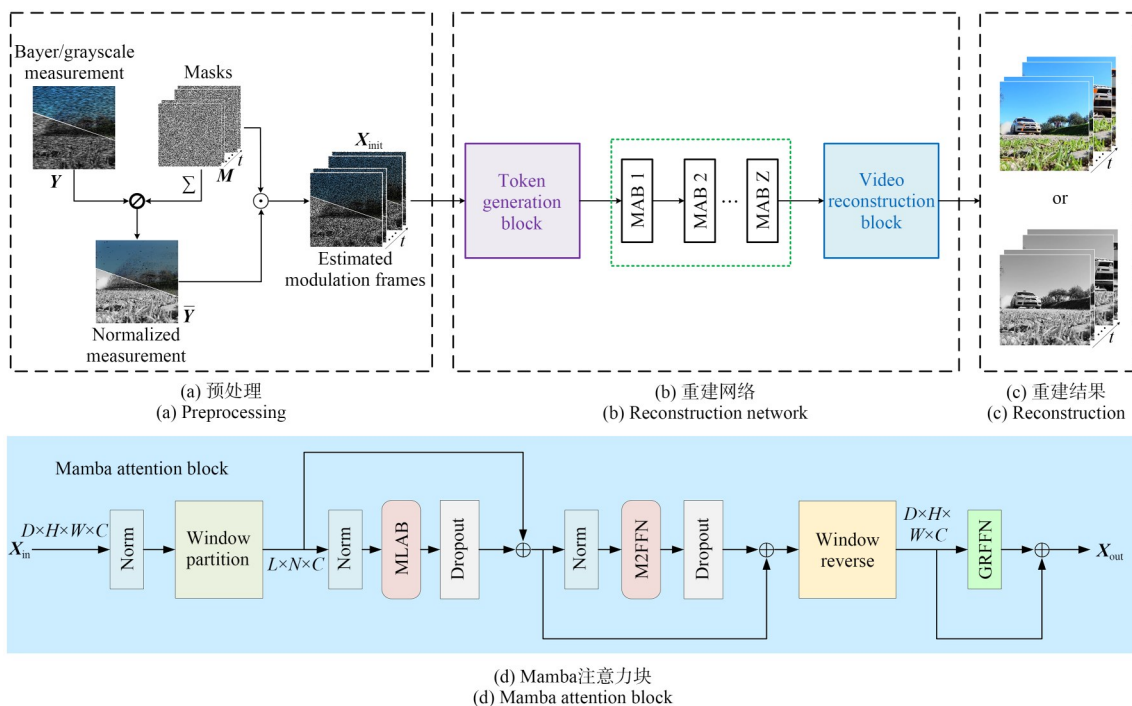


图 3 M2BA-SCI 架构图

Fig. 3 Architecture diagram of M2BA-SCI

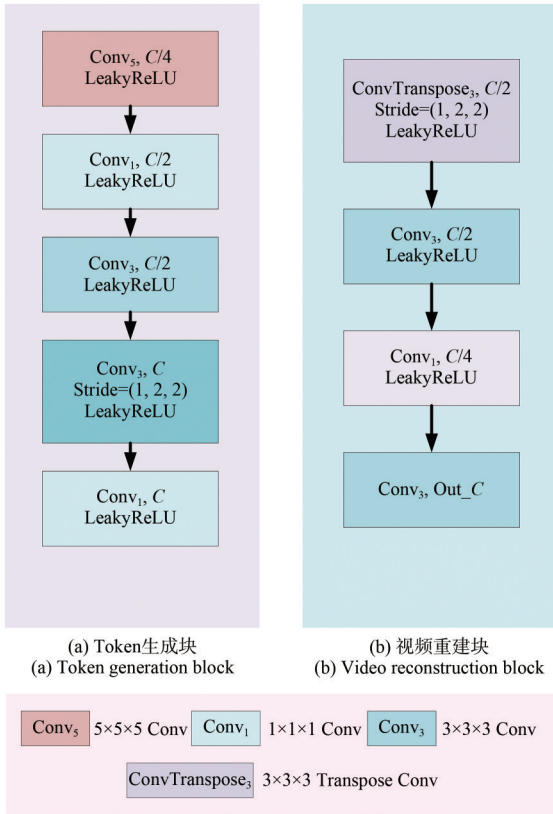


图4 token生成块与视频重建块

Fig. 4 Token generation block and video reconstruction block

探索每个 token 之间的相关性。值得注意的是,在每个 MA 块中没有使用任何下采样方法,并且每个 MA 块的输入和输出维度保持一致,这也有利于防止局部细节的丢失。

在原始的 token 被堆叠的 Z 个 MA 块映射之后,token 之间的相关性已经建立起来。最终,由 VR 块将经过 MA 块处理后的特征图转换为重建的视频帧。VR 块的结构如图 4(b) 所示。它由多个卷积层和激活函数的组成,最后一层卷积的输出通道数 Out_C 根据重建任务的不同而变化,对于灰度视频重建为 1,对于彩色视频重建为 3。

3.2 Mamba 注意力块

如图 3(d) 所示, Mamba 注意力块主要由 Mamba-2 线性注意力 (Mamba-2 Linear Attention, MLA) 块和两个不同的前馈神经网络 (Feedforward Neural Networks, FFN)^[37] 构成。其中, MLA 块可以很好地建立每个 token 块之间的相关性,两个 FFN 则有助于增强模型的表达能力以及进一步研究 token 相关性。

在 Mamba 注意力块中,首先将特征图 $X_{in} \in R^{D \times H \times W \times C}$ 划分为一系列互不重叠的局部窗口 $windows \in R^{L \times N \times C}$, 其中 $N = W_d W_h W_w$ 表示每个局部窗口中的 token 数量, $L = D \frac{HW}{N}$ 表示局部窗口的数量, W_d 、 W_h 、 W_w 分别表示局部窗口的时间维度、高度和宽度,本模型中 W_d 设置为 1, W_h 和 W_w 的默认值为 7。然后,将局部窗口序列输入 MLA 块,对每个不重叠的局部窗口进行 Mamba 注意力的计算。

3.2.1 Mamba-2 线性注意力块

在前期的实验验证中,发现引入 Mamba-1 模块能够得到较好的重建质量,但是训练速度远不及注意力机制。原因在于, Mamba-1 放弃了固定卷积核带来的并行性,而目前常用的 GPU、TPU 等加速器是为矩阵乘法进行过专门优化的。Mamba-2 提出的 SSD 在利用 GPU 张量核心来加速 SSM 的同时,还兼具 Transformer 的并行优化技术。因此本文基于 Mamba-2 模块设计了图 5 所示的 Mamba-2 线性注意力块。首先,本文将 Mamba-2 模型与线性注意力机制相结合设计出了 VMamba-2 模块,再将 VMamba-2 模块与卷积位置编码 (Convolutional Positional Encoding, CPE)、多层感知机 (Multilayer Perceptron, MLP) 以及残差连接相融合,最终形成了一个具有强大特征提取和表示能力的复合模块,即 Mamba-2 线

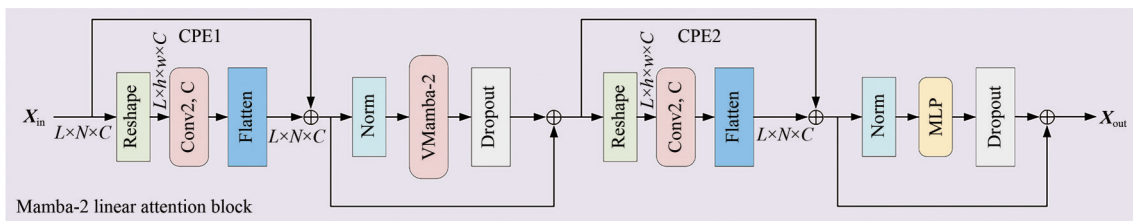


图5 Mamba-2线性注意力块网络

Fig. 5 Network of the Mamba-2 linear attention block

性注意力块。

在Mamba-2线性注意力块中,局部窗口特征序列作为输入 X_{in} 首先进入CPE1,实现对特征图的位置编码,为后续的Mamba注意力计算提供位置信息,增强模型对空间位置的感知能力。随后,将输入特征 X_{in} 与CPE1的输出相加,实现位置编码的融合,然后将结果输入到VMamba-2模块进行Mamba注意力计算。接着,将经过注意力机制处理的特征与其输入进行残差连接并将结果输入CPE2,进一步增强模型对空间位置的感知能力。最后,将上述得到的输出输入到MLP层进行非线性变换,得到最终的输出特征。在CPE1与CPE2中,均有一个Reshape的操作,其中 h 、 w 分别等于 W_h 和 W_w 。

3.2.2 两个不同的前馈神经网络

如图3(d)所示,局部窗口特征序列经过Mamba-2线性注意力块处理后进入第一个FFN,即M2FFN。M2FFN本质上也是一个MLP,不过它还包含一个权重初始化的操作,针对MLA块的线性层、归一化层和二维卷积层分

别进行了特定的权重初始化操作,有助于避免梯度消失或梯度爆炸等问题,提升模型的性能和训练效率。

随后,对经过M2FFN的局部窗口特征序列进行重组,得到与输入相同大小的特征图 $\tilde{X} \in R^{D \times H \times W \times C}$ 。将 \tilde{X} 输入到第二个FFN——分组残差前馈神经网络(Grouping Resnet FFN, GRFFN),其网络结构如图6所示。GRFFN由两个相同的残差网络构成,每个残差网络由两个三维卷积层和一个激活函数组成。首先,将输入特征图沿着通道维度分为两部分,第一部分被发送到第一个残差网络模块,其输出被添加到特征图的第二部分,并且求和的特征被用作另一个残差网络模块的输入。在此之后,两个残差网络模块的输出沿着通道维度连接以获得GRFFN的最终输出。总体而言,GRFFN通过分组并利用残差网络机制进行更多的特征映射,实现了多层信息融合,并使用卷积增强了相邻特征点之间的信息交互,有利于提升视频SCI重建的质量。

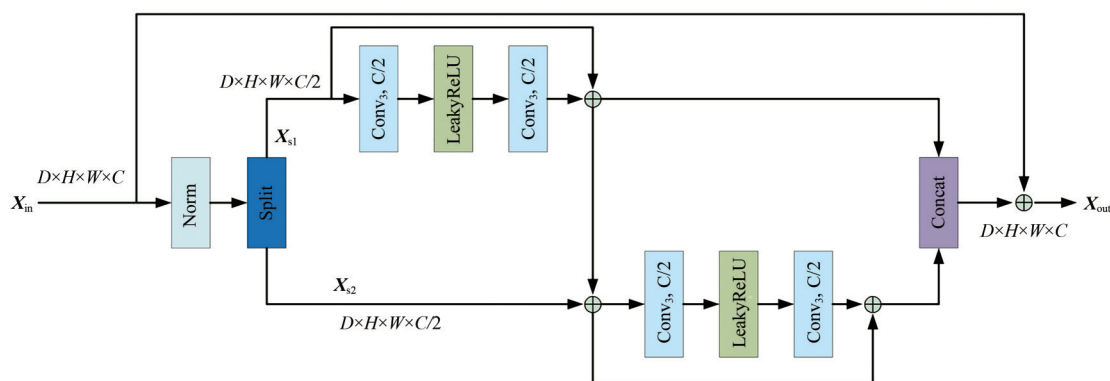


图6 分组残差前馈神经网络^[23]

Fig. 6 Grouping resnet feed forward network^[23]

4 实验结果与分析

本文在灰度模拟视频测试数据集、彩色模拟视频测试数据集和真实视频数据集上比较了提出的M2BA-SCI网络与几种SOTA视频SCI重建方法的性能,使用峰值信噪比(Peak Signal-to-Noise Ratio, PSNR)和结构相似性指数度量(Structural Similarity Index Metrics, SSIM)^[38]来评估不同视频SCI重建方法在模拟数据集上的性能。

4.1 数据集介绍

本文使用DAVIS2017^[39]作为模型的训练数据集。DAVIS2017包含90个不同的场景,共6208帧,分辨率为 480×894 和 1080×1920 。对于灰度模拟视频测试数据集,本文沿用DeSCI^[15]和Tensor ADMM-Net^[40]中的设置,使用Kobe、Runner、Drop、Traffic、Aerial和Vehicle六个基准数据集,尺寸为 256×256 ,压缩比为8。这些数据集用以测试预训练模型对于灰度压缩测量帧的

视频重建性能。对于彩色模拟视频测试数据集,本文按照 PnP-FastDVDnet^[17]中的设置,使用 6 个基准彩色模拟数据集,分别是 Beauty、Bosphorus、Jockey、Runner、ShakeNDry 和 Traffic,这些数据集的尺寸为 $512 \times 512 \times 3$,压缩比为 8,可以用来测试预训练模型对于彩色压缩测量帧的视频重建性能。真实视频数据集由真实的视频 SCI 摄像机^[6,29]捕获,共包含 4 个数据集,分别为 Duomino、Water Balloon、Hand 和 Hammer。这些真实视频数据的尺寸为 512×512 ,压缩比为 10,能够很好地反映实际视频 SCI 过程的重建需求。

4.2 实验设置与结果分析

本文选择的训练数据集的分辨率为 480×894 。对视频数据采用旋转、翻转、剪切等数据增强方式,同时将视频段裁剪为 $128 \times 128 \times 8$ 的数据立方体,得到 26 000 多个训练数据对作为训练集。在训练过程中,通过将测量帧和掩码作为输入来训练 M2BA-SCI 网络,并使用 Adam 优化器^[41]来优化模型。所有的实验都在 PyTorch 框架上运行。在模型训练阶段,为加快训练速度,本文选择租用 NVIDIA RTX 4090 GPU。在测试阶

段,统一在实验室配备的 NVIDIA RTX 4070 GPU 上进行。

4.2.1 灰度模拟视频数据集

目前,视频 SCI 的重建方法有很多种。将本文提出的方法与一些 SOTA 方法进行视频重建性能的比较,其中有基于模型的迭代优化方法 GAP-TV^[11],端到端深度学习的方法 U-net^[6]、BIRNAT^[19]、RevSCI^[33]和 Efficientsci^[42],即插即用方法 PnP-FFDNet^[16]和 PnP-FastDVDnet^[17],以及深度展开方法 GAP-CCoT^[43]。表 1 记录了 6 个基准灰度视频数据集上不同重建方法的 PSNR、SSIM 值以及重建一个测量帧的平均所需时间,其中最佳结果用加粗字体表示,次佳结果用下划线表示。图 7 展示了不同重建方法在 Aerial、Traffic、Kobe 和 Crash 4 个灰度模拟视频测试数据集上的可视化结果,为便于观察,对局部细节进行放大显示。通过分析表 1 和图 7 中数据,得出以下结论:

(1) 本文提出的 M2BA-SCI 模型实现了 34.85 的平均 PSNR 和 0.966 的平均 SSIM,相较于之前的 SOTA 方法中最优的 Efficientsci,分别提升了 0.63 和 0.005,且重建速度有接近 3 倍的提

表 1 不同重建算法在灰度模拟视频测试数据集上的 PSNR、SSIM 值和运行时间

Tab. 1 PSNR, SSIM values and running time of different reconstruction algorithms on the grayscale simulated video test dataset

Dataset	Evaluation	Grayscale simulated video test dataset							Running time/s
		Aerial	Crash	Drop	Kobe	Runner	Traffic	Average	
PnP-FastDVDnet	PSNR	27.87	26.33	41.92	32.33	36.14	26.17	31.79	6.20
	SSIM	0.895	0.915	0.989	0.943	0.962	0.917	0.937	
PnP-FFDNet	PSNR	24.29	24.66	39.69	30.32	32.44	23.87	29.21	0.70
	SSIM	0.820	0.837	0.987	0.923	0.934	0.829	0.888	
GAP-TV	PSNR	25.02	24.63	34.49	26.64	30.13	20.65	26.93	0.74
	SSIM	0.826	0.826	0.967	0.840	0.914	0.697	0.845	
GAP-CCoT	PSNR	29.40	28.52	42.54	32.58	39.12	29.03	33.53	0.04
	SSIM	0.923	0.941	0.992	0.949	0.980	0.938	0.954	
BIRNAT	PSNR	28.99	27.84	42.28	32.71	38.70	29.33	33.31	0.11
	SSIM	0.918	0.927	0.992	0.951	0.977	0.943	0.951	
RevSCI	PSNR	29.35	28.13	42.92	<u>33.72</u>	39.40	<u>30.02</u>	33.92	0.17
	SSIM	0.925	0.936	0.992	0.957	0.978	<u>0.950</u>	0.956	
U-net	PSNR	27.94	26.97	38.13	29.14	34.93	24.94	30.34	0.02
	SSIM	0.886	0.894	0.961	0.893	0.957	0.849	0.907	
Efficientsci	PSNR	<u>30.32</u>	<u>29.27</u>	43.56	33.45	<u>39.51</u>	29.20	<u>34.22</u>	0.95
	SSIM	<u>0.937</u>	<u>0.954</u>	0.993	<u>0.960</u>	<u>0.981</u>	0.942	<u>0.961</u>	
M2BA-SCI(Ours)	PSNR	30.74	30.28	<u>43.31</u>	33.90	40.50	30.37	34.85	0.23
	SSIM	0.943	0.962	<u>0.992</u>	0.961	0.984	0.953	0.966	

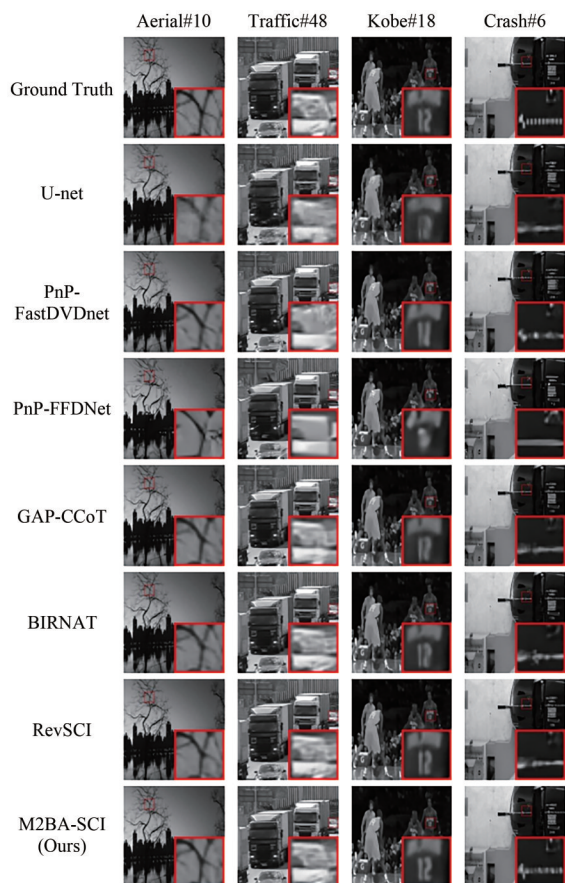


图 7 灰度模拟视频测试数据集的重建视频帧对比图
Fig. 7 Comparison chart of reconstructed video frames on the grayscale simulated video test dataset

升,虽仍慢于同样表现良好的 RevSCI,但综合来看,本文所提方法是一个较为均衡的方法,在提高重建质量的同时,仍然保持较快的重建速度。

(2)从重建视频的可视化来看,本文所提方法能够恢复出更多的细节和边缘信息。从图 7 中 Traffic 这一组对比图来看,以前的 SOTA 方法要么复原得过于模糊,看不清小车的轮廓信息;要么过于平滑,丢失了车灯等细节。M2BA-SCI 可以得到与 Ground Truth 最为贴近的重建效果。而在对高速场景 Crash 的重建中,虽然本文方法的复原效果也不太理想,但对于车身上的条纹信息,重建效果仍为最佳,依稀能够分辨出是一条一条的纹路,而不是像其他方法那样无法看清条纹边界。

4.2.2 彩色模拟视频数据集

为了验证本文方法对各类视频 SCI 重建任务的有效性,本文将 M2BA-SCI 网络扩展到彩色视频 SCI 重建任务中。在空间尺寸为 $512 \times$

512×3 的 6 个基准彩色视频数据集^[17]上进行了相关实验,其中 3 表示 RGB 通道。与基准灰度视频数据集类似,以 $B=8$ 的压缩比压缩视频。如图 1 所示,使用具有拜耳滤镜的相机捕获压缩的拜耳测量,将每个数据集的 32 个彩色视频帧压缩得到 4 个拜耳测量。由于之前的许多方法并没有考虑彩色视频重建的场景,本文仅与 GAP-TV、PnP-FFDNet、PnP-FastDVDnet 以及 STFormer 进行比较,其中即插即用模型根据降噪器的不同分为灰度版本和彩色版本^[23]。各类算法的重建结果如图 8 和表 2 所示,加粗数字为最优值,次佳结果用下划线表示。通过对结果进行分析,得出以下结论:

(1)M2BA-SCI 网络在基准彩色视频数据集上的平均 PSNR 达到了 36.21,比之前重建效果最好的 STFormer 方法提高了 0.12,特别是在高速运动场景 Bosphorus 中提高了 0.98,表明 M2BA-SCI 在彩色高速运动场景中也是有效的。并且, M2BA-SCI 的重建速度相较于其他几种算法也是最快的,展现出了更高的实时性能。

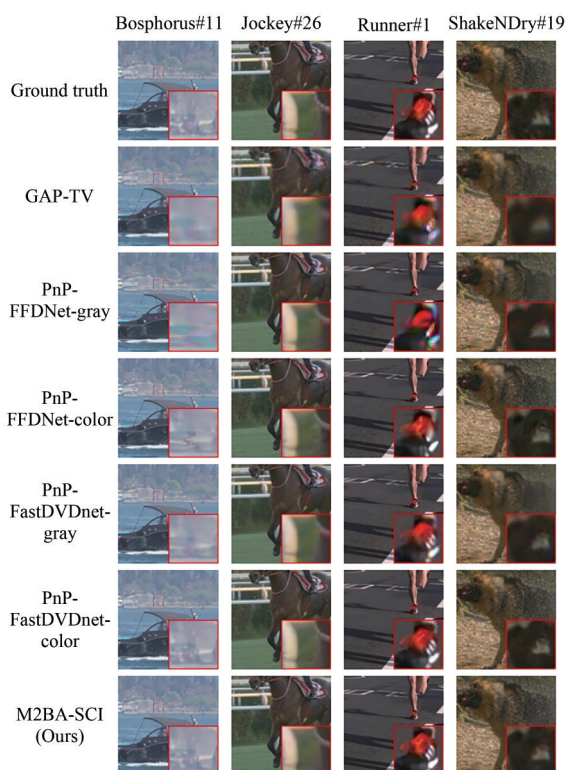


图 8 彩色模拟视频测试数据集的重建视频帧对比图
Fig. 8 Comparison chart of reconstructed video frames on the color simulated video test dataset

表 2 不同重建算法在彩色模拟视频测试数据集上的 PSNR、SSIM 值和运行时间

Tab. 2 PSNR, SSIM values and running time of different reconstruction algorithms on the color simulated video test dataset

Dataset	Evaluation	Color simulated video test dataset							Running time/s
		Beauty	Bosphorus	Jockey	Runner	ShakeNDry	Traffic	Average	
PnP-FastDVDnet-gray	PSNR	34.58	34.03	33.78	35.21	30.40	24.47	32.08	26.01
	SSIM	0.968	0.953	0.932	0.930	0.887	0.832	0.917	
PnP-FastDVDnet-color	PSNR	36.26	37.10	36.27	39.85	33.89	28.40	35.31	102.22
	SSIM	0.976	0.975	0.956	0.974	0.947	0.924	0.959	
PnP-FFDNet-gray	PSNR	33.15	28.30	32.31	30.63	27.72	20.73	28.81	3.98
	SSIM	0.962	0.897	0.920	0.883	0.844	0.695	0.867	
PnP-FFDNet-color	PSNR	34.60	33.34	35.21	35.49	32.65	24.78	32.68	54.89
	SSIM	0.970	0.956	0.948	0.941	0.939	0.841	0.932	
GAP-TV	PSNR	33.46	29.60	29.49	29.69	29.83	19.40	28.58	3.11
	SSIM	0.966	0.906	0.890	0.874	0.885	0.611	0.855	
STFormer	PSNR	<u>36.83</u>	<u>38.36</u>	<u>37.09</u>	<u>40.56</u>	<u>34.67</u>	<u>29.00</u>	<u>36.09</u>	2.96
	SSIM	<u>0.980</u>	<u>0.988</u>	<u>0.963</u>	<u>0.980</u>	<u>0.952</u>	<u>0.923</u>	<u>0.963</u>	
M2BA-SCI(ours)	PSNR	<u>37.06</u>	<u>39.34</u>	<u>36.80</u>	<u>40.95</u>	<u>34.44</u>	<u>28.69</u>	<u>36.21</u>	1.03
	SSIM	<u>0.979</u>	0.985	<u>0.961</u>	<u>0.980</u>	<u>0.949</u>	0.922	<u>0.963</u>	

(2)从可视化结果来看,本文方法可以恢复更清晰的边缘和更多的局部细节。以Bosphorus为例,对于远处背景中的物体,M2BA-SCI可以得到几乎等同于真实值(Ground Truth)的结果。对于Runner中的懈怠和ShakeNDry中的牙齿也复原得更为清晰。GAP-TV、PnP-FFDNet-gray和PnP-FastDVDnet-gray方法的重建结果存在一定的伪影,而PnP-FFDNet-color和PnP-FastDVDnet-color方法的重建结果边缘模糊,细节丢失严重。

4.2.3 真实视频数据集

本文提出的M2BA-SCI网络在模拟视频数据集上表现出优秀的重建性能。为验证本方法在真实场景下的视频重建质量,本文进一步对Qiao等人设计的真实视频SCI相机^[6]捕获的Duomino、Water Balloon和Hand视频数据进行重建分析。这些视频数据含有多个压缩比 $B=\{10,20,30,40,50\}$,所有压缩测量值的空间尺寸为 512×512 。

如图9和图10所示,在压缩比为10的场景中,本文将提出的M2BA-SCI网络与几种SOTA重建算法进行比较,即GAP-TV、PnP-FFDNet、PnP-FastDVDnet。通过放大重建结果局部区域,可以观察到本文所提方法可以恢复Duomino数

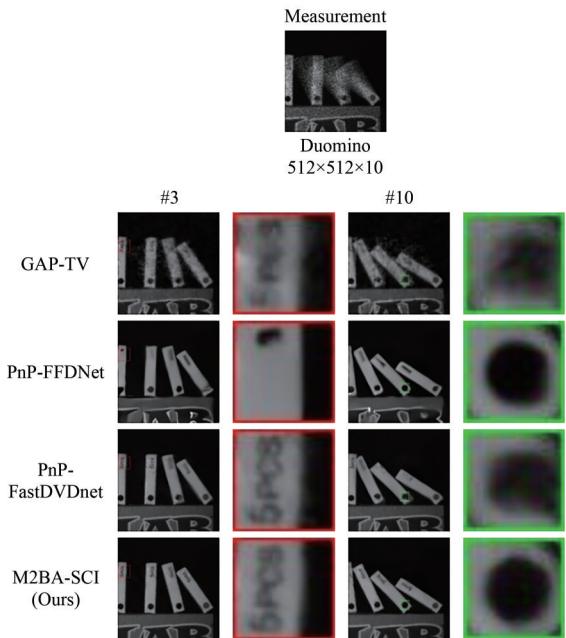


图9 真实视频数据 Duomino 的重建视频帧对比图

Fig. 9 Comparison chart of reconstructed video frames of real video data Duomino

据和 Water Balloon 数据中的清晰字母和尖锐边缘,而GAP-TV、PnP-FFDNet和PnP-FastDVDnet算法的重建结果过度平滑这些区域,带有一些伪影。特别是在 Duomino 数据集的复原中,PnP-FFDNet的结果连字迹都已无法观察到,

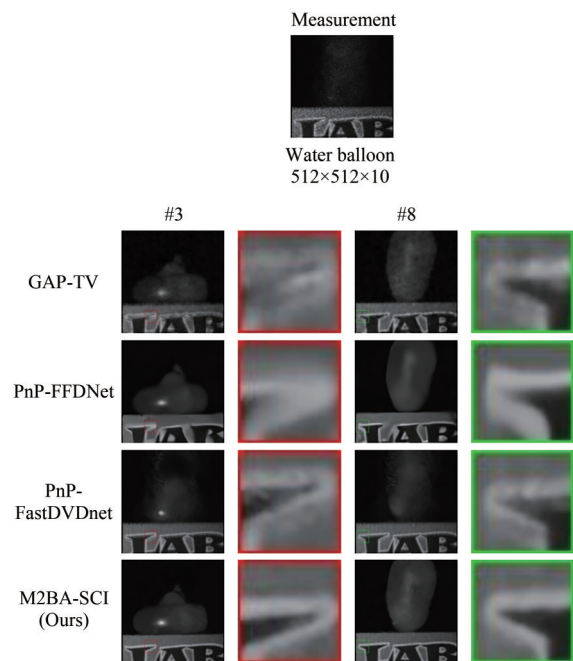


图10 真实视频数据 Water Ballon 的重建视频帧对比图
Fig. 10 Comparison chart of reconstructed video frames of real video data Water Ballon

PnP-FastDVDnet虽然字迹复原较为清晰,但对于圆圈这类特征复原效果较差。在 Water Balloon 数据集的复原结果中可以观察到,PnP-FastDVDnet 算法有一块较大的虚影没有去除,使得重建效果很差。除此之外,其他3种方法重建视频所需的时间也远高于本文所提方法。可见,M2BA-SCI 网络在真实视频数据的重建任务中表现出较高的性能。

4.3 消融实验

本文提出的 M2BA-SCI 模型重点在于引入 Mamba-2 线性注意力块 (MLAB)。在对 Mamba 进行改进的工作中, SiMBA^[44] 的作者提到在 Mamba 模块后引入一种名为 EinFFT 的通道建模方法来提升模型的性能。因此,为了验证这两个模块对于整体重建质量的影响,本文针对它们做了一些消融实验,实验结果如表3所示,其中√表示实验模型包含该模块,参数量 (Params) 与浮点数运算次数 (FLOPs) 可以反应模型的复杂度。根据实验结果,可以得到以下结论:

表3 M2BA-SCI 在6个基准灰度视频数据集上的消融实验

Tab. 3 Ablation experiments of M2BA-SCI on six benchmark grayscale video datasets

实验	MLAB	Mamba-2	M2FFN	EinFFT	Params/M	FLOPs/G	PSNR	SSIM	Running time/s
1		√			5.16	807.39	34.32	0.962	0.47
2	√				5.17	808.12	34.40	0.963	0.20
3	√			√	5.17	808.70	34.05	0.961	0.38
4	√		√		5.7	882.89	34.85	0.966	0.23

(1) MLA 模块相对于纯 Mamba-2 模块,得益于引入线性注意力机制,在拥有略高于 Mamba-2 的模型复杂度的基础上,重建质量和重建速度均优于纯 Mamba-2,特别是重建速度提升了 57%。

(2) 在 MLA 模块之后添加 EinFFT 方法,模型复杂度虽没有明显增加,但由于 EinFFT 引入了傅里叶变换,使得重建速度大幅降低。然而,速度的牺牲并没有换来质量的提升,PSNR 和 SSIM 还分别下降了 0.35 和 0.002。

(3) 在 MLA 模块之后引入 PVT2FFN,这也是 M2BA-SCI 模型的构成,相较于只引入 MLA 模块,模型复杂度有所升高,但 PSNR 和 SSIM 分别提高了 0.45 和 0.003,重建时间仅慢了 0.03 s。

为了验证 M2BA-SCI 网络的宽度和深度对重建质量的影响,本文设计了 3 组实验,其中通道

数用于调整网络的宽度,MLAB 的数量则用以调节网络的深度。实验结果如表4所示,增加模型的宽度和深度有利于重建质量的提高,但也导致运行时间有所加长。

综合来看,MLAB+PVT2FFN 是最优的组

表4 使用不同通道和块数的 M2BA-SCI 在6个灰度基准数据集上的重建质量和运行时间

Tab. 4 Reconstruction quality and running time on 6 grayscale benchmark datasets using M2BA-SCI with different number of channels and blocks

实验	Channel	Block	PSNR	SSIM	Time/s
1	64	2	32.55	0.943	0.12
2	128	2	34.85	0.966	0.23
3	128	4	34.97	0.968	0.64

合方案,它们都对 M2BA-SCI 模型性能的提升做出了贡献,通过消融实验验证了它们对于模型的重要性的必要性。此外,M2BA-SCI 中引入的 TG 块和 GRFFN 已在 STFormer 的消融实验中充分证明了它们对于整个视频 SCI 重建网络具备大于 1 dB 的有效增益。

5 结 论

本文成功地将 Mamba-2 首次应用于视频 SCI 重建任务中,提出了一种基于 Mamba-2 的端

到端深度视频 SCI 重构方法 M2BA-SCI。通过在模拟和真实视频数据集上的大量实验,证明了 M2BA-SCI 在提升重建质量的同时,仍然具备较高的重建速度。在灰度模拟视频数据集中,M2BA-SCI 模型重建结果的平均 PSNR 和 SSIM 分别达到了 34.85 和 0.966;在彩色模拟视频数据集中,分别达到了 36.21 和 0.963;在真实视频数据集中,也表现出对字母细节和尖锐边缘更强的重建能力。M2BA-SCI 的提出为视频 SCI 重建带来了新的思路,为基于 Mamba 模型设计出更高重建质量的算法提供了参考。

参 考 文 献:

- [1] YUAN X, BRADY D J, KATSAGGELOS A K. Snapshot compressive imaging: theory, algorithms, and applications [J]. *IEEE Signal Processing Magazine*, 2021, 38(2): 65-88.
- [2] DONOHO D L. Compressed sensing [J]. *IEEE Transactions on Information Theory*, 2006, 52(4): 1289-1306.
- [3] LLULL P, LIAO X J, YUAN X, *et al.* Coded aperture compressive temporal imaging [J]. *Optics Express*, 2013, 21(9): 10526-10545.
- [4] HITOMI Y, GU J W, GUPTA M, *et al.* Video from a single coded exposure photograph using a learned over-complete dictionary [C]//*Proceedings of 2021 International Conference on Computer Vision*. Barcelona: IEEE, 2011: 287-294.
- [5] REDDY D, VEERARAGHAVAN A, CHELLAPPA R. P2C2: programmable pixel compressive camera for high speed imaging [C]//*Proceedings of Computer Vision & Pattern Recognition 2021*. Colorado Springs: IEEE, 2011: 329-336.
- [6] QIAO M, MENG Z Y, MA J W, *et al.* Deep learning for video compressive sensing [J]. *APL Photonics*, 2020, 5(3): 030801.
- [7] SUN Y Y, YUAN X, PANG S. Compressive high-speed stereo imaging [J]. *Optics Express*, 2017, 25(15): 18182-18190.
- [8] YANG P H, KONG L H, LIU X Y, *et al.* Shearlet enhanced snapshot compressive imaging [J]. *IEEE Transactions on Image Processing*, 2020, 29: 6466-6481.
- [9] MAGGIONI M, BORACCHI G, FOI A, *et al.* Video denoising, deblocking, and enhancement through separable 4-d nonlocal spatiotemporal transforms [J]. *IEEE Transactions on Image Processing*, 2012, 21(9): 3952-3966.
- [10] ZHAO C, MA S W, ZHANG J, *et al.* Video compressive sensing reconstruction via reweighted residual sparsity [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2017, 27(6): 1182-1195.
- [11] YUAN X. Generalized alternating projection based total variation minimization for compressive sensing [C]//*Proceedings of 2016 IEEE International Conference on Image Processing*. Phoenix: IEEE, 2016: 2539-2543.
- [12] YANG J B, LIAO X J, YUAN X, *et al.* Compressive sensing by learning a Gaussian mixture model from measurements [J]. *IEEE Transactions on Image Processing*, 2015, 24(1): 106-109.
- [13] YANG J B, YUAN X, LIAO X J, *et al.* Video compressive sensing using Gaussian mixture models [J]. *IEEE Transactions on Image Processing*, 2014, 23(11): 4863-4878.
- [14] GU S H, XIE Q, MENG D Y, *et al.* Weighted nuclear norm minimization and its applications to low level vision [J]. *International Journal of Computer Vision*, 2017, 121(2): 183-208.
- [15] LIU Y, YUAN X, SUO J L, *et al.* Rank minimization for snapshot compressive imaging [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, 41(12): 2990-3006.
- [16] YUAN X, LIU Y, SUO J L, *et al.* Plug-and-play algorithms for large-scale snapshot compressive imaging [C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle:

- IEEE, 2020: 1444-1454.
- [17] YUAN X, LIU Y, SUO J L, *et al.* Plug-and-play algorithms for video snapshot compressive imaging [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 44(10): 7093-7111.
- [18] WU Z L, YANG C S, SU X F, *et al.* Adaptive deep PnP algorithm for video snapshot compressive imaging [J]. *International Journal of Computer Vision*, 2023, 131(7): 1662-1679.
- [19] CHENG Z H, LU R Y, WANG Z J, *et al.* BIRNAT: bidirectional recurrent neural networks with adversarial training for video snapshot compressive imaging [C]//*Proceedings of the 16th European Conference on Computer Vision*. Glasgow: Springer, 2020: 258-275.
- [20] WANG Z J, ZHANG H, CHENG Z H, *et al.* MetaSCI: scalable and adaptive reconstruction for video compressive sensing [C]//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville: IEEE, 2021: 2083-2092.
- [21] WU Z Y, ZHANG J, MOU C. Dense deep unfolding network with 3D-CNN prior for snapshot compressive imaging [C]//*Proceedings of 2021 IEEE/CVF International Conference on Computer Vision*. Montreal: IEEE, 2021: 4872-4881.
- [22] YANG C S, ZHANG S Y, YUAN X. Ensemble learning priors driven deep unfolding for scalable video snapshot compressive imaging [C]//*Proceedings of the 17th European Conference on Computer Vision*. Tel Aviv: Springer, 2022: 600-618.
- [23] WANG L S, CAO M, ZHONG Y, *et al.* Spatial-temporal transformer for video snapshot compressive imaging [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(7): 9072-9089.
- [24] ZHU L, LIAO B, ZHANG Q, *et al.* Vision mamba: efficient visual representation learning with bidirectional state space model [C]//*Proceedings of the 41st International Conference on Machine Learning*. Vienna: PMLR, 2024: 62429-62442.
- [25] GU A, DAO T. Mamba: linear-time sequence modeling with selective state spaces [J/OL]. *arXiv*, 2023: 2312.00752.
- [26] MEHTA H, GUPTA A, CUTKOSKY A, *et al.* Long range language modeling via gated state spaces [C]//*Proceedings of the 11th International Conference on Learning Representations*. Kigali: ICLR, 2023: 1-20.
- [27] WANG J, ZHU W T, WANG P C, *et al.* Selective structured state-spaces for long-form video understanding [C]//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver: IEEE, 2023: 6387-6397.
- [28] ZHENG Z R, WU C. U-shaped vision mamba for single image dehazing [J/OL]. *arXiv*, 2024: 2402.04139.
- [29] YUAN X, LLULL P, LIAO X J, *et al.* Low-cost compressive sensing for color video and depth [C]//*Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition*. Columbus: IEEE, 2014: 3318-3325.
- [30] HUANG T, PEI X H, YOU S, *et al.* LocalMamba: visual state space model with windowed selective scan [J/OL]. *arXiv*, 2024: 2403.09338.
- [31] GU A, GOEL K, RE C. Efficiently modeling long sequences with structured state spaces [C]//*Proceedings of the 10th International Conference on Learning Representations*. Online: ICLR, 2021: 1-27.
- [32] DAO T, GU A. Transformers are SSMS: generalized models and efficient algorithms through structured state space duality [C]//*Proceedings of the Forty-First International Conference on Machine Learning*. Vienna: ICML, 2024: 1-31.
- [33] CHENG Z H, CHEN B, LIU G L, *et al.* Memory-efficient network for large-scale video compressive sensing [C]//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville: IEEE, 2021: 16241-16250.
- [34] MENG Z Y, JALALI S, YUAN X. GAP-net for snapshot compressive imaging [J/OL]. *arXiv*, 2020: 2012.08364.
- [35] ZHENG S M, YANG X Y, YUAN X. Two-stage is enough: a concise deep unfolding reconstruction network for flexible video compressive sensing [J/OL]. *arXiv*, 2022: 2201.05810.
- [36] XU B, WANG N Y, CHEN T Q, *et al.* Empirical evaluation of rectified activations in convolutional network [J/OL]. *arXiv*, 2015: 1505.00853.

- [37] VASWANI A, SHAZEER N, PARMAR N, *et al.* Attention is all you need [C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: Curran Associates Inc., 2017: 6000-6010.
- [38] WANG Z, BOVIK A C, SHEIKH H R, *et al.* Image quality assessment: from error visibility to structural similarity [J]. *IEEE Transactions on Image Processing*, 2004, 13(4): 600-612.
- [39] PONT-TUSET J, PERAZZINI F, CAELLES S, *et al.* The 2017 DAVIS challenge on video object segmentation [J/OL]. *arXiv*, 2017: 1704.00675.
- [40] MA J W, LIU X Y, SHOU Z, *et al.* Deep tensor ADMM-net for snapshot compressive imaging [C]//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*. Seoul: IEEE, 2019: 10222-10231.
- [41] KINGMA D P. Adam: a method for stochastic optimization [C]//*Proceedings of the 3rd International Conference on Learning Representations*. San Diego: ICLR, 2015: 1-15.
- [42] WANG L S, CAO M, YUAN X. EfficientSCI: densely connected network with space-time factorization for large-scale video snapshot compressive imaging [C]//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver: IEEE, 2023: 18477-18486.
- [43] WANG L S, WU Z L, ZHONG Y, *et al.* Snapshot spectral compressive imaging reconstruction using convolution and contextual transformer [J]. *Photonics Research*, 2022, 10(8): 1848-1858.
- [44] PATRO B N, AGNEESWARAN V S. SiMBA: simplified mamba-based architecture for vision and multivariate time series [J/OL]. *arXiv*, 2024: 2403.15360.

作者简介:



石敦攀,男,硕士研究生,2022年于武汉大学获得学士学位,主要从事视频快照压缩成像重建算法的研究。E-mail: 2578452996@qq.com



徐 伟,男,博士,研究员,2008年于中国科学院长春光学精密机械与物理研究所获得博士学位,主要从事光学小卫星总体技术的研究。E-mail: xwciomp@126.com