文章编号　1004-924X(2025)18-2980-16

高效Mamba驱动的端到端光场图像压缩

封哲宇[1]，蒋志迪[2]，万立飞[1]，徐海勇[1]，蒋刚毅[1*]
(1. 宁波大学 信息科学与工程学院，浙江 宁波 315211；
2. 宁波大学 科技学院，浙江 宁波 315300)

摘要：光场图像因记录了光线的空间与角度信息，可提供比传统2D图像更丰富的视觉信息，但其高维特性导致现有压缩方法在全局特征利用、长距离相关性挖掘及计算复杂度上存在局限，限制了压缩性能和效率的提升。为此，本文提出了一种高效Mamba驱动的端到端光场图像压缩方法。首先，从4D光场图像中提取包含空间和极平面信息的2D切片，并利用Mamba充分捕捉其全局上下文信息。其次，为了在多个方向上扫描光场图像并避免计算复杂度的大幅增加，引入了一种通道高效的2D选择性扫描策略，以精确高效地提取光场特征。最后，在解码端设计了一个残差重建模块，该模块在降低参数量和减少编解码时间的基础上，显著提升了重建图像的质量。实验结果表明，与现有代表方法SADN相比，所提方法在7×7角度分辨率的光场图像上平均实现了7.4%的码率降低和0.37 dB的PSNR提升，同时在主观视觉质量上也表现更佳。在编解码时间方面，所提方法实现了10～20倍的显著提升。此外，与现有最新方法LFIC-DRASC相比，所提方法在13×13角度分辨率的光场图像上平均实现了19.5%的码率降低和0.58 dB的PSNR提升。

关　键　词：光场；图像压缩；端到端；Mamba
中图分类号：TP394.1　　**文献标识码**：A
doi：10.37188/OPE.20253318.2980　　**CSTR**：32169.14.OPE.20253318.2980

Efficient mamba-driven end-to-end light field image compression

FENG Zheyu[1], JIANG Zhidi[2], WAN Lifei[1], XU Haiyong[1], JIANG Gangyi[1*]

(1. *Faculty of Information Science and Engineering, Ningbo University, Ningbo* 315211, *China*;
2. *College of Science and Technology Ningbo University, Ningbo* 315300, *China*)
* *Corresponding author, E-mail: jianggangyi@nbu.edu.cn*

Abstract: Light field images capture both spatial and angular information of light rays, providing richer visual information than traditional 2D images. However, their high-dimensional nature poses challenges for existing compression methods in terms of global feature utilization, long-range correlation exploration, and computational complexity, limiting the improvements of compression performance and efficiency. To address these issues, this paper proposed an efficient Mamba-driven end-to-end light field image compression method. Firstly, 2D slices containing spatial and epipolar plane information were extracted from the 4D light field image, and Mamba was employed to fully capture their global contextual information. Secondly, to scan the light field image in multiple directions while avoiding a significant increase in computational complexity, a channel-efficient 2D selective scanning strategy was introduced to extract light field features accurately and efficiently. Finally, on the decoding end, a residual reconstruction module was designed to

收稿日期：2025-03-31；**修订日期**：2025-07-10.
基金项目：国家自然科学基金项目(No. 62271276, No. 62171243)

enhance the reconstructed image quality while reducing the number of parameters and decreasing the encoding and decoding time. The experimental results show that compared with the existing representative method SADN, the proposed method achieves an average bitrate reduction of 7.4% and a PSNR improvement of 0.37 dB on light field images with a 7×7 angular resolution, while also demonstrating superior subjective visual quality. In terms of encoding and decoding time, the proposed method has achieved a significant improvement of 10 to 20 times. Furthermore, compared to the state-of-the-art method LFIC-DRASC, the proposed method achieves an average bitrate reduction of 19.5% and a PSNR improvement of 0.58 dB on light field images with a 13×13 angular resolution.

Key words: light field; image compression; end-to-end; Mamba

1　引　言

光场图像(Light Field Images, LFI)能够同时记录光线的角度和空间信息，使其在数字重聚焦[1]、光谱光场成像[2]、光场渲染[3]等任务中展现出独特的优势。然而，光场图像因包含丰富的视角信息，其数据量远超传统2D图像，这给存储、传输和处理带来了巨大挑战。因此，开发高效的光场图像压缩(Light Field Image Compression, LFIC)方法对于光场图像的实际应用具有重要意义。

现有的光场图像压缩方法可分为非端到端和端到端两大类。非端到端方法主要沿两条技术路线展开：基于伪视频序列(Pseudo Video Sequence, PVS)转换的方法和基于空间相关性挖掘的方法。第一条技术路线的代表性研究有Liu等人[4]提出的子孔径图像(Sub-Aperture Images, SAIs)排序与PVS编码方案，通过对SAIs进行重排生成PVS，然后利用高效视频编码(High Efficiency Video Coding, HEVC)[5]或多功能视频编码(Versatile Video Coding, VVC)[6]等主流视频编码标准消除视图间冗余。比如，Shao等人[7]基于光场多视图图像之间的相似性，提出了一种新的二维预测结构，这种编码结构还可以实现渐进式传输功能，以满足不同场景的需求。后续研究还采用稀疏视图传输与解码端视图合成策略来进一步降低码率。如Huang等人[8]提出了一种低码率光场压缩方法，通过优化视差图几何一致性和SAIs预测算法，降低码率并保持光场结构的一致性。Zhang等人[9]提出了一种几何感知视图重建网络，通过融合光场的几何信息，提升了光场图像的压缩效率。Liu等人[10]提出多流密集视图重建网络，通过多尺度特征融合和视差感知补偿机制增强光场压缩性能。然而，PVS方法在遮挡区域易因一致性假设[11]失效引发重建伪影，降低光场图像的重建质量。第二条技术路线是直接建模宏像素的空间相关性。例如，Liu等人[12]采用基于高斯过程回归的光场图像压缩方法，通过内容分类和预测优化，提升压缩效率并降低计算复杂度。Schiopu等人[13]通过宏像素合成技术一次性合成整个光场图像并用于剩余视图的无损编码，进一步提升了编码效率。尽管挖掘空间相关性的方法通过消除空间冗余取得了一定进展，但预测模型的精度上限仍是制约其性能提升的关键障碍。此外，这两种技术路线均存在架构局限性，模块化设计导致各组件缺乏联合优化，这种系统性缺陷限制了非端到端方法的压缩潜能。

近年来，端到端图像压缩方法[14-18]在2D图像压缩任务上取得了显著进展。例如，Minnen等人[14]利用自回归上下文模型捕捉潜在特征间的相关性，他们的后续工作[15]对通道之间的上下文进行了切片建模，使得通道信息得到了有效利用。为了解决自回归上下文解码耗时长的问题，He等人[16]设计了棋盘上下文模型以实现并行计算。Jiang等人[17]结合了基于Transformer注意力机制的增强棋盘上下文模块和全局空间上下文模块，有效解决了局部和全局空间相关性问题，其改进版MLIC++[18]通过整合线性注意力计算进一步降低了计算复杂度。然而，光场图像不仅是更高维的数据，还受到角度一致性约束[19]。端到端2D图像压缩方法仅考虑图像的空间特性，难以利用光场图像视角间的相关性，不适合直接应用于4D结构的光场图像压缩。因此，现有的

端到端光场图像压缩方法主要致力于解耦光场图像的空间和角度特征。例如，Singh 等人[20]提出了一种结合视差辅助模型的光场图像压缩方法，确保重建光场的结构完整性。Zhong 等人[21]采用 3D 卷积自编码器，在角度或空间分辨率上进行上采样和下采样，以减少伪影。Tong 等人[22]设计了一种空间-角度解耦网络，通过步长卷积和膨胀卷积分别提取角度和空间信息，并进行联合压缩。Ye 等人[23]基于极平面图像（Epipolar Plane Image，EPI）引入角度补全模块和空间-角度联合变换，以减少冗余信息。Feng 等人[24]提出了一种基于解耦表示和不对称条带卷积的光场图像压缩方法，进一步解耦光场特征，增强了模型在表示复杂空间-角度关系方面的能力。然而，受限于卷积神经网络（Convolutional Neural Network，CNN）有限的感受野，这些方法仍然存在全局建模能力不足的问题，难以捕捉长距离空间-角度关系。

在近期的研究中，注意力机制[25]凭借其在捕捉长距离依赖关系和动态调整输入权重方面的强大能力，在计算机视觉任务中得到广泛应用，并被引入 LFIC 任务。例如，Tong 等人[26]通过全局注意力模块捕获光场图像的长距离依赖关系，提取具有大感受野的紧凑特征图，从而降低了像素级冗余。Liu 等人[27]进一步引入了局部-全局相关性学习策略，结合局部卷积和全局注意力机制，可有效处理聚焦型光场图像中的局部结构和长距离相关性。然而，现有基于注意力机制的光场图像压缩方法在处理下采样后的光场图像特征时，既面临着可提取信息减少影响重建质量的问题，又承受着较高内存成本和计算复杂度的负担。因此，如何在保证光场图像高效压缩的同时兼顾计算效率，仍然是一个亟待解决的挑战。

状态空间模型（State Space Model，SSM）[28-29]作为一种新兴的序列建模方法，近年来在深度学习领域展现出强大的长程依赖建模能力。与注意力机制不同，SSM 通过线性递归计算实现长距离信息建模，具有线性复杂度的优势。Mamba 模型[30]作为 SSM 的一种高效变体，通过引入选择性机制和硬件优化，在自然语言处理任务中超越了 Transformer，并成功应用于图像分类[31-32]和图像压缩[33]等视觉任务。例如，Qin 等人[33]提出了一种基于 Mamba 的 2D 选择性扫描方法，实现了对 2D 图像的高效压缩。然而，在 LFIC 领域，Mamba 的应用尚处于起步阶段，仍需进一步研究和探索其在 4D 光场数据上的适用性。

针对上述挑战，本文提出了一种高效 Mamba 驱动的端到端光场图像压缩方法（Efficient Mamba-driven Light Field Image Compression，EMLFIC）。EMLFIC 将 Mamba 高效地应用于 4D 光场数据，克服了 CNN 和 Transformer 的局限性，实现了更优的压缩性能和更低的计算复杂度，显著提升了光场图像压缩性能，为光场图像压缩研究提供了新的思路和技术路径。本文的主要贡献包括：

（1）将 Mamba 模型应用于 LFIC 任务中，提出了一种端到端的光场图像压缩方法 EMLFIC。利用高效 Mamba 提取光场图像的长距离空间-角度依赖关系，在保持高压缩率的同时提升重建质量。

（2）引入了通道高效的 2D 选择性扫描机制，通过多方向跳跃扫描光场图像，实现更高效的特征学习与提取。该方法不仅增强了模型的空间-角度信息整合能力，还降低了计算复杂度。

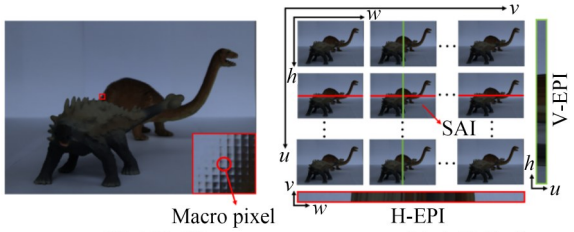
（3）设计了解码端的光场图像残差重建模块，通过残差学习机制，在减少模型参数量和编解码时间的同时，保证了重建光场图像的高质量生成。

2　提出方法

2.1　光场图像的表示形式

光场图像的基础是四维光场函数，它描述了通过空间中每个点的各个方向的光线量。这种四维函数的表示形式为 $L(u,v,h,w)$，其中 u 和 v 定义了角度维度，h 和 w 定义了空间维度。光场图像的表示形式主要有以下三种：第一种是微透镜图像（Micro Lens Image，MLI）形式，如图 1(a)所示，它由一系列宏像素组成，这些宏像素在空间上排列，捕捉了场景中不同位置的角度信息；第二种是子孔径图像（Sub-Aperture Images，SAIs）形式，如图 1(b)所示，每个 SAI 记录了从特定视角观察到的空间信息。第三种是极平面图

像（Epipolar Plane Image，EPI）形式，如图 1（b）中水平极平面图像（Horizontal EPI，H-EPI）和垂直极平面图像（Vertical EPI，V-EPI）所示，EPI 中的一条直线代表 3D 场景中同一个点在不同视角下的投影轨迹，其斜率反映了该点的深度和视差。



(a) 微透镜图像　(b) 子孔径图像阵列
(a) Micro lens image　(b) Sub-aperture image array

图 1　光场图像的不同表示形式

Fig. 1　Different representation of light field images

2.2　EMLFIC 网络总体框架

针对现有端到端光场图像压缩方法在全局特征提取和长距离相关性建模方面的问题，本文提出了高效 Mamba 驱动的端到端光场图像压缩方法 EMLFIC。该方法通过精心设计的编码器-熵模型-解码器结构，结合高效的状态空间建模能力，优化 4D 光场结构中的空间-角度信息表达，提升了光场图像的压缩质量和效率。整体框架如图 2 所示，主要包括光场编码器、熵模型以及光场解码器三个核心模块。光场编码器由高效 Mamba（Efficient Mamba，EMamba）和主编码器 g_a 构成，旨在将原始高维光场图像 L 转换为潜在特征表示 $\boldsymbol{y}$，并通过量化函数 Q 得到 $\hat{\boldsymbol{y}}$。同时，利用超先验编码器 h_a 从 $\boldsymbol{y}$ 中提取辅助信息 $\boldsymbol{z}$，以进一步提升熵模型的建模能力。在熵建模阶段，采用融合了通道自回归与空间上下文的熵模型，联合建模潜在变量的分布，提升码率估计精度。最后，使用由主解码器 g_s 和残差重建模块（Residual Reconstruction Module，RRM）组成的光场解码器，通过对 $\hat{\boldsymbol{y}}$ 的解码，实现对重建光场的高质量复原。熵编解码器沿用了与经典端到端光场图像压缩方法 SADN[22] 一致的算术编码器（Arithmetic Encoder，AE）和算术解码器（Arithmetic Decoder，AD）。光场图像的编码和解码过程可以表示为：

$$\boldsymbol{y}=g_a(H_E(L;\theta_{g_a}),\tag{1}$$

$$\hat{\boldsymbol{y}}=Q(\boldsymbol{y}),\tag{2}$$

$$\hat{L}=H_R(g_s(\hat{\boldsymbol{y}};\theta_{g_s})),\tag{3}$$

其中：L 和 $\hat{L}$ 分别表示原始光场图像和重建的光场图像。H_E 表示 EMamba，H_R 表示 RRM。θ_{g_a} 和 θ_{g_s} 分别是 g_a 和 g_s 的可学习参数，Q 表示量化函数。

L 经过 EMamba 的全局特征学习后，输入到主编码器 g_a 中进一步编码得到潜在变量 $\boldsymbol{y}$。$\boldsymbol{y}$ 经过量化函数 Q 量化为 $\hat{\boldsymbol{y}}$，再经 AE 和 AD 传输至主解码器 g_s。最后，通过 RRM 重建得到 $\hat{L}$。超先验

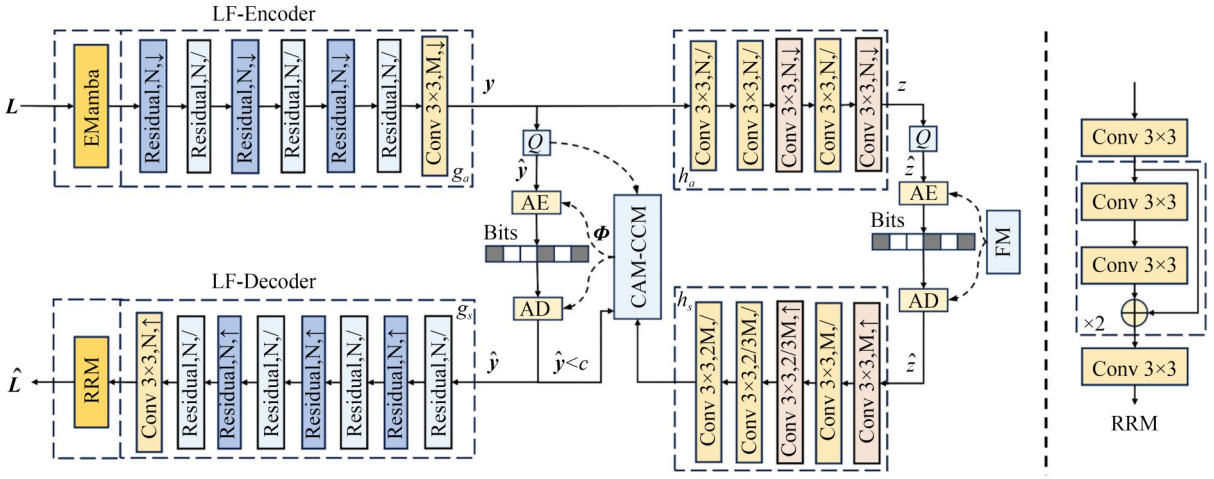


图 2　EMLFIC 的总体框图。↓ 表示下采样，/ 表示卷积步长为 1，↑ 表示上采样。FM 表示因式分解熵模型。

Fig. 2　Overall framework of EMLFIC. ↓ represents down-sampling, / represents a convolution stride of 1, and ↑ represents up-sampling. FM represents factorized entropy model

编解码和熵参数计算过程表示为：

$$z = h_a(\boldsymbol{y};\ \theta_{h_a}), \tag{4}$$

$$\hat{z} = Q(z), \tag{5}$$

$$\boldsymbol{\Phi} = H_{CC}(h_s(\hat{z};\ \theta_{h_s})), \tag{6}$$

其中：z是由超先验编码器h_a从$\boldsymbol{y}$中提取的超先验信息，$\hat{z}$表示z的量化版本。θ_{h_a}和θ_{h_s}分别是h_a和h_s的可学习参数。H_{CC}表示CAM－CCM熵模型，它是文献[15]和文献[16]的结合版本。具体来说，在通道上采用文献[15]提出的将潜在变量y按通道均匀划分为10个切片的通道自回归熵模型(Channel-wise Autoregressive Entropy Model，CAM)，在空间上则采用文献[16]提出的支持并行编解码的棋盘格上下文模型(Checkerboard Context Model，CCM)。$\boldsymbol{\Phi}$是将$\hat{z}$先输入到超先验解码器h_s，再通过CAM－CCM熵模型推导出来的均值和方差，用于指导AE和AD。

值得注意的是，上述公式中涉及的g_a，g_s，h_a和h_s都是文献[34]移除了注意力模块的简化版本。所提方法并未直接在光场解码器中引入对称结构的EMamba，而是设计了轻量的残差重建模块RRM，其具体结构如图2右侧所示。这不仅避免了冗余特征建模和参数堆叠，还进一步提升了解码效率与可扩展性。整套框架在保持模型紧凑性的同时，显著提升了在空间-角度建模、特征冗余抑制以及主客观质量上的综合表现。

2.3　高效Mamba(EMamba)

在光场图像压缩中，如何减少4D光场结构中的空间-角度冗余是一个关键问题。本文提出的EMamba旨在高效解耦光场图像中的空间和角度特征，以实现最大程度的冗余消除。如图3所示，EMamba由初始特征提取(Initial Feature Extraction，IFE)块、极平面特征提取(Epipolar Plane Feature Extraction，EFE)块、空间-角度特征提取(Spatial-Angular Feature Extraction，SAFE)块和特征融合(Feature Fusion，FF)块组成。其中，EMamba的基本组件残差状态空间块将在2.4节中进行详细介绍。

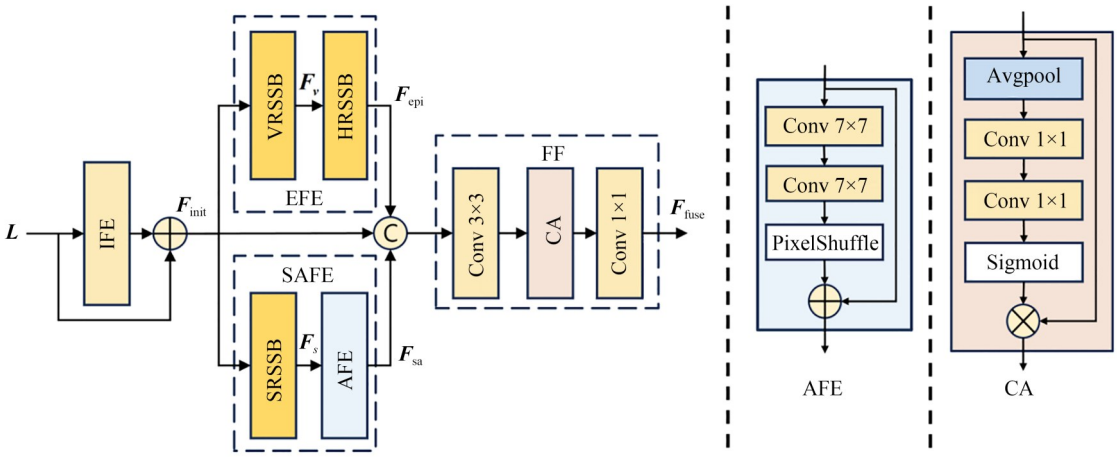


图3　EMamba的详细结构

Fig. 3　The detailed structure of EMamba

首先，通过与文献[35]～文献[37]类似的IFE块提取输入的光场图像$L \in \boldsymbol{R}^{UH \times VW \times C}$的初始特征，并重塑为$F_{\text{init}} \in \boldsymbol{R}^{U \times V \times H \times W \times C}$：

$$F_{\text{init}} = H_I(L), \tag{7}$$

其中：H_I表示IFE块，C表示通道维度，与g_a的通道数相同，在本文中设置为48。

随后，初始特征F_{init}被输入到并联的SAFE块和EFE块中。SAFE块由空间残差状态空间块(Spatial Residual State Space Block，SRSSB)和角度特征提取(Angular Feature Extraction，AFE)块组成。其中，SRSSB利用状态空间建模方式，从每个视角图像内提取空间长程依赖。具体来说，将初始特征F_{init}重塑为2D SAI形式$F_{\text{SAI}} \in \boldsymbol{R}^{UV \times H \times W \times C}$，通过SRSSB提取空间特征，

再将捕捉到的空间特征重塑为 $F_s \in \boldsymbol{R}^{U\times V\times H\times W\times C}$：

$$F_s = H_{\mathrm{SR}}(F_{\mathrm{init}}), \tag{8}$$

其中，H_{SR} 表示空间残差状态空间块SRSSB。

由于角度分辨率远小于空间分辨率，本文未采用角度残差状态空间块，而是使用由膨胀系数为7的膨胀卷积和像素Shuffle组成的AFE块来捕捉 F_s 的角度信息，使得特征的表达能力在角度维度上得到增强。具体来说，将空间特征 F_s 重塑为2D MLI形式 $F_{\mathrm{MLI}} \in \boldsymbol{R}^{HW\times U\times V\times C}$，以便更有效地提取角度特征，最终重塑为 $F_{sa} \in \boldsymbol{R}^{U\times V\times H\times W\times C}$：

$$F_{\mathrm{sa}} = H_A(F_s), \tag{9}$$

其中，H_A 表示AFE块。

EFE块由垂直残差状态空间块(Vertical Residual State Space Block, VRSSB)和水平残差状态空间块(Horizontal Residual State Space Block, HRSSB)串联组成，分别用于提取垂直极平面图像(Vertical EPI, V-EPI)和水平极平面图像(Horizontal EPI, H-EPI)的空间-角度混合信息，这能充分反映光场图像在角度维度上的几何结构。具体来说，将 F_{init} 重塑为 $\boldsymbol{F}_{\mathrm{V\text{-}EPI}} \in \mathrm{R}^{UH\times V\times W\times C}$ 输入到VRSSB得到 F_v，再把 F_v 重塑为 $F_{\mathrm{H\text{-}EPI}} \in \boldsymbol{R}^{VW\times U\times H\times C}$ 输入到HRSSB，并最终将提取到的EPI特征重塑为 $F_{\mathrm{epi}} \in \boldsymbol{R}^{U\times V\times H\times W\times C}$：

$$F_{\mathrm{epi}} = H_{\mathrm{HR}}(H_{\mathrm{VR}}(F_{\mathrm{init}})), \tag{10}$$

其中：H_{HR} 表示HRSSB，H_{VR} 表示VRSSB。

进一步，通过拼接操作将上述三个特征在通道维度组合，并输入到由卷积和通道注意力[38]组成的FF块中实现特征融合，得到 $F_{\mathrm{fuse}} \in \boldsymbol{R}^{U\times V\times H\times W\times C}$。该特征不仅包含了空间和角度维度的上下文信息，还保留了极平面方向的重要结构特征，为后续的熵建模和解码提供了坚实基础：

$$F_{\mathrm{fuse}} = H_F(\mathrm{Concat}(F_{\mathrm{init}}, F_{\mathrm{sa}}, F_{\mathrm{epi}})), \tag{11}$$

其中，H_F 表示FF块。

最后，将 F_{fuse} 输入到 g_a 中，进一步压缩得到潜在变量 $\boldsymbol{y}$。

2.4　残差状态空间块

如图4所示，本文采用残差状态空间块(Residual State Space Block, RSSB)[39]作为EMamba的基本组件，并结合所提出的高效视觉状态空间块(Efficient Visual State Space Block, EVSSB)，以更高效地完成图像压缩任务。RSSB通过整合SSM和卷积操作，能够高效捕捉全局依赖和局部特征。具体来说，给定一个省略了批次维度的2D特征 $F \in \boldsymbol{R}^{H\times W\times C}$，RSSB首先利用层归一化(Layer Normalization, LN)和EVSSB捕捉输入特征的长程依赖关系，并引入可学习的缩放因子 $\boldsymbol{s}_1 \in \boldsymbol{R}^C$ 来控制跳跃连接的信息传递：

$$\bar{F} = H_{\mathrm{EV}}(\mathrm{LN}(F)) + \boldsymbol{s}_1 \cdot F, \tag{12}$$

其中：H_{EV} 表示EVSSB，LN表示层归一化。

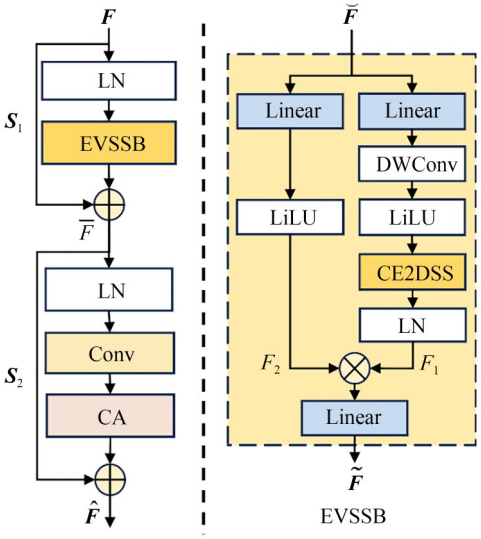


图4　RSSB的结构图

Fig. 4　Structural diagram of RSSB

接着，在对输出结果 $\bar{F}$ 进行归一化后，引入局部卷积以弥补SSM扁平化操作导致的局部像素关系丢失。此外，RSSB通过通道注意力(Channel Attention, CA)机制优化通道间的表达，消除冗余信息，进一步增强关键通道的表示能力。最终，通过可调节的缩放因子 $\boldsymbol{s}_2 \in \boldsymbol{R}^C$ 完成残差连接：

$$\hat{F} = \mathrm{CA}(\mathrm{Conv}(\mathrm{LN}(\bar{F})) + \boldsymbol{s}_2 \cdot \bar{F}. \tag{13}$$

为了减少计算复杂度和模型参数，EVSSB采用的通道高效的扫描机制(Channel-Efficient 2D Selective Scan, CE2DSS)，能在保持性能的同时，提高模型的效率。EVSSB的架构采用双流并行结构[31]，在第一个分支中，首先对输入特征 $\check{F} \in \boldsymbol{R}^{H\times W\times C}$ 进行线性变换和深度卷积，然后通过SiLU激活函数[40]、CE2DSS和LN得到 F_1。在第二个分支中，输入特征通过线性层扩展后，

直接进行SiLU激活得到F_2。两个分支的特征通过Hadamard积聚合后，再通过线性层将通道数投影回原始维度，生成与输入形状相同的输出$\tilde{F}$：

$$F_1 = \mathrm{LN}(H_{CE}(\mathrm{SiLU}(\mathrm{DWConv}(\mathrm{Linear}(\bar{F}))))), \tag{14}$$

$$F_2 = \mathrm{SiLU}(\mathrm{Linear}(\bar{F})), \tag{15}$$

$$\tilde{F} = \mathrm{Linear}(F_1 \odot F_2), \tag{16}$$

其中：DWConv表示深度卷积，H_{CE}表示CE2DSS，$\odot$表示Hadamard积，Linear表示线性投影。

2.5 通道高效的2D选择性扫描

4D光场图像可通过多种形式（如SAI、MLI和EPI）予以表示，选择合适的扫描方式对效率至关重要。鉴于原始Mamba[30]的因果性限制，需采用多向扫描以确保各视图中的Token充分交互。本文将4D光场图像视为多个2D切片的组合，将其投影至高维空间并展平为1D序列，作为SSM的输入。通过在空间和EPI维度分别扫描，有效缩短扫描长度，解决了长程记忆问题，同时保留了完整的4D全局信息。

如图5(a)所示，为提取全局空间信息，将4D光场特征转换为SAI阵列形式，并在各SAI中提取空间上下文信息。此外，如图5(c)所示，EPI蕴含丰富的空间与角度维度信息，通过在H-EPI（或V-EPI）子空间进行扫描，可融合空间与角度信息的水平（或垂直）方向特征。

如图5(b)～5(d)所示，为增强非因果2D图像的建模能力，CE2DSS将4D光场图像的2D切片数据沿通道维度划分为四组，通过引入跳跃采样机制[41]，显著减少了所需扫描的Token数量，在SSM序列建模中节省了多倍计算成本，同时

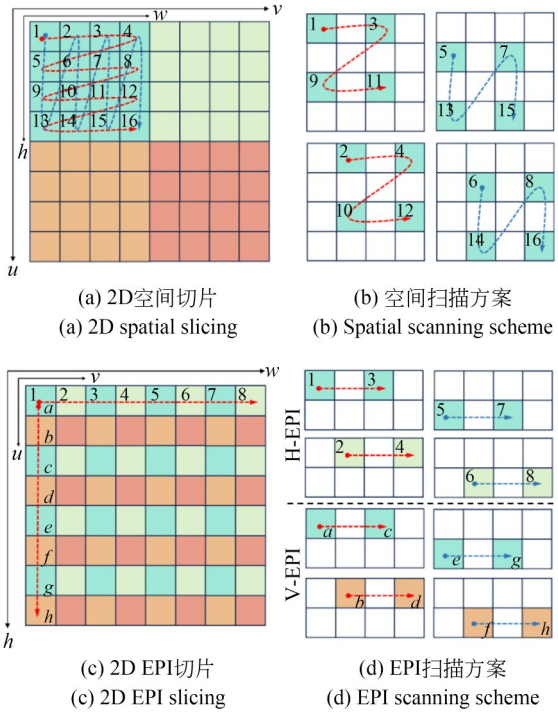


图5　空间和EPI的2D切片及其扫描策略

Fig. 5　Spatial and EPI 2D slices and the scanning strategies

保持了全局感受野。该策略高效提取全局依赖关系，并将计算复杂度从$O(N)$降至$O(N/4)$。随后，将不同位置跳跃采样的数据按不同顺序展平为1D序列以学习长程依赖，最终通过重塑操作将序列还原为原始2D切片数据并沿通道维度拼接，如图6所示。这种对通道进行分组和跳跃采样的扫描方案仅牺牲很少的学习能力即可大幅减少参数量。

值得注意的是，由于图5的例子中$W=H=4$，$U=V=2$，H-EPI（或V-EPI）的长度为$W\times U=8$，所以按CE2DSS的扫描方法会得到如图5(d)所示的内容。如果$U=V=4$，则CE2DSS

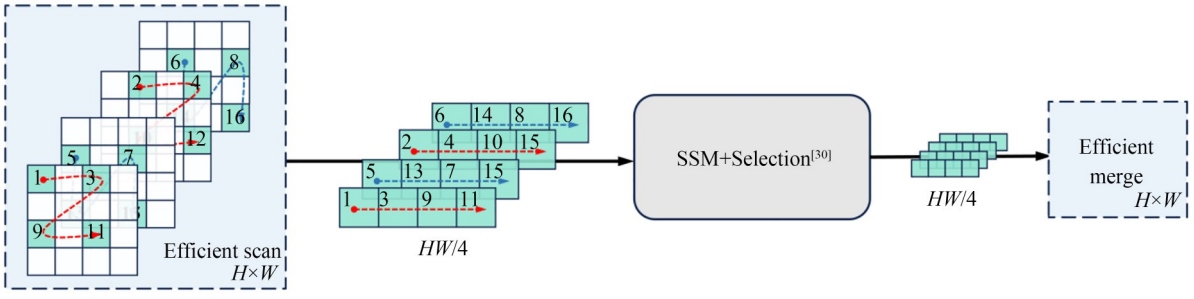


图6　以2D空间切片为例的CE2DSS

Fig. 6　Illustration of CE2DSS using 2D spatial slices

的 EPI 扫描路径与图 5(b)相同。

本文采用与 SADN[22]相同的率失真损失函数 J 来训练 EMLFIC：

$$J=\lambda D+R, \tag{17}$$

其中：R 表示码率，用每像素比特数来表示。D 是对输入 L 和输出 $\hat{L}$ 计算均方误差得到的失真。超参数 λ 设置为 {0.000 1, 0.000 15, 0.000 2, 0.000 5, 0.001, 0.002, 0.003}，用于控制码率和失真之间的权衡。

3　实验结果与分析

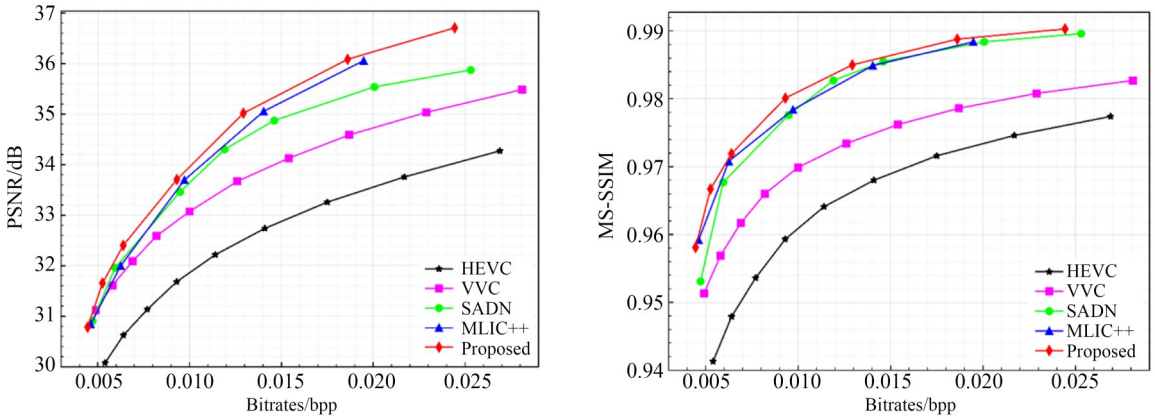
3.1　实验设置

本文在 CompressAI 平台[42]上实现了提出的 EMLFIC。采用 SADN[22]提出的大规模真实场景数据集 PINET 进行训练，并将选择的光场图像非重叠地裁剪成大小为 448×448 的图像块，即角度分辨率为 7×7，空间分辨率为 64×64。测试数据为 EPFL 数据集[43]提供的 12 个不同场景。整个架构在 RTX 4070 Ti Super GPU 上进行训练，CPU 为 Intel Core i7-12700。采用 Adam 优化器，Batch size 大小为 4，初始学习率为 10^{-4}，当优化停止改进时，学习率降低 0.1 倍。

对比方法包括基于光场 PVS 编码的 HEVC[5]和 VVC[6]、基于端到端光场图像编码的 SADN[22]以及基于端到端 2D 图像编码的 MLIC++[18]。其中，基于 PVS 的传统方法采用 YUV444 格式和蛇形排序，HEVC 和 VVC 分别采用了参考软件 HM16.26[44]和 VTM23.1[45]的低延时 P 模式。SADN，MLIC++和 EMLFIC 都是在 PINET 数据集的 40014 个非重叠块上重新训练的。此外，EMLFIC 和 MLIC++的潜在特征通道数分别为 48 和 192，超先验通道数分别为 80 和 320。

3.2　编码性能比较

图 7 给出了不同方法在 EPFL 测试数据集中 12 个不同场景下的平均率失真性能对比结果，如图所示，所提方法 EMLFIC 在所有测试码率下均表现最优，其平均 PSNR 和 MS-SSIM 指标明显优于其他对比方案。与基于 PVS 的方法相比，EMLFIC 的编码效率远高于 HEVC 和 VVC 的编码方案。与 SADN 相比，EMLFIC 对光场图像进一步的解纠缠，更有效地去除了光场图像内部的冗余信息，使得整体性能优于 SADN。另一方面，MLIC++在所有测试码率上的表现都低于 EMLFIC，这表明具有 4D 结构的光场图像与传统 2D 图像的差异较大，EMLFIC 更有利于光场图像的压缩。MLIC++的 PSNR 在较高的测试码率上高于 SADN，这是因为 MLIC++具有更大的特征通道数，且使用了更为复杂的上下文模型。此外，MLIC++与模型复杂度更低的 SADN 在 MS-SSIM 指标上表现出几乎相同的性能。这一现象表明，空间-角度特征提取能够有效学习到光场图像的结构一致性，保证了光场图



(a) PSNR率失真曲线
(a) PSNR rate-distortion curve

(b) MS-SSIM率失真曲线
(b) MS-SSIM rate-distortion curve

图 7　EPFL 测试数据集上不同方法的率失真曲线性能比较

Fig. 7　Performance comparison of rate-distortion curve of different methods on the EPFL test dataset

像的MS-SSIM性能。

为了进一步客观评价压缩性能，本文将Bjøntegaard Delta PSNR（BD-PSNR）和Bjøntegaard Delta BitRate（BD-BR）作为客观指标进行比较。BD-PSNR正值表示与对比方法相比，在相同码率下所提方法实现了更佳的客观质量，而负值则反之。BD-BR负值表示与对比方法相比，在相同质量下所提方法节省了码率，而正值则表示所提方法需要消耗更多的码率。如表1所示，与HEVC，VVC，SADN和MLIC++相比，所提方法EMLFIC分别平均节省了39.320%，11.229%，7.412%和7.882%的码率，而BD-PSNR分别提高了2.055 dB，0.752 dB，0.372 dB和0.315 dB。与传统方法HEVC和VVC相比，在平坦区域主导的场景（如Ankylosaurus&Diplodocus、Color Chart和Magnets）中，EMLFIC的率失真性能相对受限，而在纹理复杂的场景（如Bikes，Danger De Mort，Flowers及Fountain&Vincent）中，其性能提升尤为显著。相较于端到端对比方法，虽然在Danger De Mort和Flowers场景的特定码率下PSNR指标略低于SADN，在Bikes场景稍逊于MLIC++，但是EMLFIC在绝大多数测试场景上都展现出更优的压缩性能。其主要原因在于这些场景所具有的特殊物理特性与模型结构的适配性存在差异。Danger De Mort和Flowers场景中存在大量细粒度纹理、复杂边缘结构以及明显的局部视差变化，使得光场图像在角度维度上冗余性较低且局部非一致性较强，而EMLFIC所采用的全局空间-角度联合建模策略在应对这类纹理突变与非刚性运动时存在一定局限，反观SADN通过空间-角度解纠缠机制对这类局部细节建模更为敏感，因此在这两个场景中表现更优。另一方面，Bikes场景中包含大量重复的几何结构和高频纹理信息，如车轮辐条等规则图案，对熵建模的准确性要求更高，而MLIC++采用多参考点熵建模和更复杂的上下文结构，具备更强的特征捕捉能力，因而在该场景中取得更高的PSNR。综上所述，EMLFIC在面对局部变化剧烈或重复结构密集的场景时，其模型结构在细节建模与熵估计方面仍有提升空间。但是，综合所有12个测试场景的平均率失真性能，EMLFIC实现了最优表现，仍能说明所提方法的有效性。

表1　所提方法与对比方法在12个测试场景的BD-BR和BD-PSNR

Tab. 1　BD-BR and BD-PSNR of the proposed method compared with the comparative methods on 12 test scenarios

Images	Proposed vs.							
	HEVC		VVC		SADN		MLIC++	
	BD-BR /%	BD-PSNR/ dB	BD-BR /%	BD-PSNR /dB	BD-BR /%	BD-PSNR /dB	BD-BR /%	BD-PSNR/ dB
Ankylosaurus&Diplodocus	+3.134	+0.904	+29.826	+0.238	−25.726	+0.506	−22.207	+0.572
Bikes	−59.690	+2.942	−43.284	+1.865	+3.841	+0.017	+2.578	−0.078
Color Chart	−24.980	+1.198	+50.845	−1.586	−1.168	+1.063	−21.799	+1.042
Danger De Mort	−62.348	+3.278	−46.632	+2.080	+10.294	−0.232	−0.143	+0.016
Desktop	−24.144	+1.226	+11.859	−0.251	−7.597	+0.181	−8.692	+0.390
Flowers	−65.101	+2.955	−48.255	+1.947	+2.873	−0.055	−0.433	+0.019
Fountain&Vincent	−62.476	+2.912	−45.514	+1.812	−3.653	+0.181	−2.557	+0.107
Friends	−37.794	+1.824	−10.527	+0.631	−11.820	+0.543	−8.374	+0.330
ISO Chart	−39.443	+2.149	−2.514	+0.246	−19.127	+0.904	−3.580	+0.133
Magnets	+2.267	+0.217	+31.548	−0.842	−22.087	+0.729	−21.283	+0.939
Stone Pillars Outside	−47.149	+2.077	−25.520	+1.070	−8.021	+0.331	−3.273	+0.115
Vespa	−54.113	+2.975	−36.584	+1.810	−6.747	+0.299	−4.826	+0.194
Average	**−39.320**	**+2.055**	**−11.229**	**+0.752**	**−7.412**	**+0.372**	**−7.882**	**+0.315**

3.3　视觉质量比较

角度一致性。EPI能够有效捕捉自然景物的深度变化信息以及被遮挡区域的信息，从而反映光场图像的角度信息变化情况。因此，EPI可用于评估光场图像的角度一致性。本文选取了 Bikes、Danger De Mort 和 Fountain&Vincent 三个场景进行比较，结果如图 8 所示。在对比中，基于 PVS 的 HEVC 和 VVC 方法通过将 SAIs 转换为伪视频序列进行压缩，但由于 SAIs 的不连续性和误差累积，导致 EPI 出现了失真和扭曲。MLIC＋＋由于未充分考虑光场图像的角度信息，需要更大的特征通道数才能达到与所提方法相近质量的重建 EPI。尽管 SADN 考虑了空间和角度信息的耦合，但其重建的 LFI 在角度一致性方面仍有改进空间。相比之下，所提方法生成的 EPI 与原始 EPI 更为接近，表明 EMLFIC 能够更好地保留 LFI 的 4D 结构。

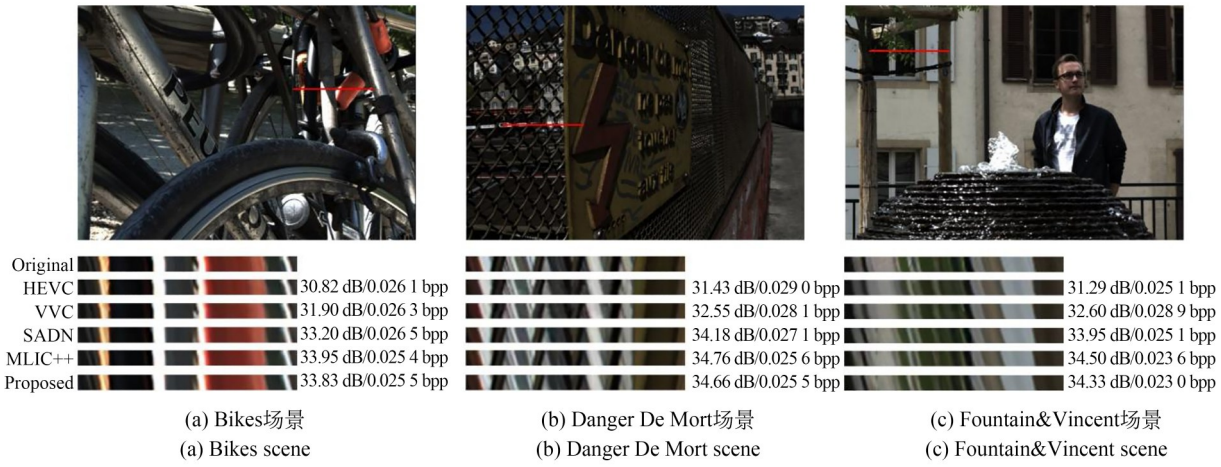


图 8　Bikes、Danger De Mort 和 Fountain&Vincent 场景的 EPI 质量比较

Fig. 8　EPI quality comparison for the Bikes, Danger De Mort, and Fountain&Vincent scenes

主观质量。鉴于 MLI 难以通过人眼进行直观观察，而中心 SAI 能够更清晰地呈现重建图像的主观质量，故本文选取中心 SAI 进行视觉对比分析。同时，还通过残差图直观展示了原始光场图像与重建光场图像的差异，如图 9 所示。从图中可以观察到，在三种测试场景下，HEVC 和 VVC 均出现了明显的边缘模糊和结构失真现象，其残差分布呈现显著的高频伪影特征。而

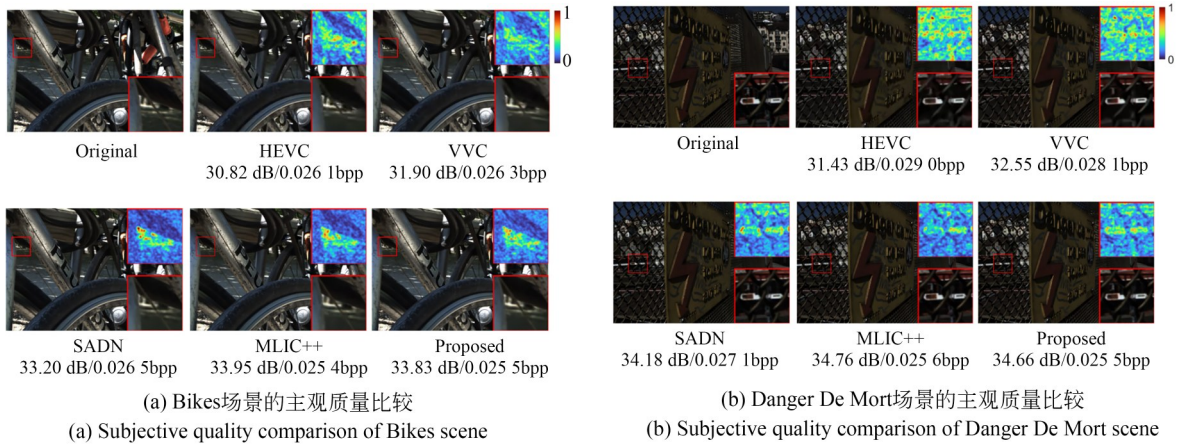


图 9　Bikes 和 Danger De Mort 场景的中心 SAI 主观质量比较

Fig. 9　Subjective quality comparison of the central SAI for the Bikes and Danger De Mort scenes

SADN在复杂纹理区域存在细节丢失问题。尽管MLIC++方法在部分场景重建精度和视觉质量方面与所提方法EMLFIC最为接近，但其存在模型参数量大、计算复杂度高的问题。相比之下，EMLFIC通过高效建模MLI中宏像素间的纹理关系，有效捕捉了光场图像的长程依赖特性。实验结果表明，EMLFIC在重建图像清晰度和视觉伪影抑制方面均表现出显著优势，特别是在复杂纹理区域能够更好地保留细节信息，克服了SADN在该区域存在的细节丢失问题。这一改进表明了EMLFIC在LFI细节信息保持方面的有效性和优越性。

3.4 消融研究

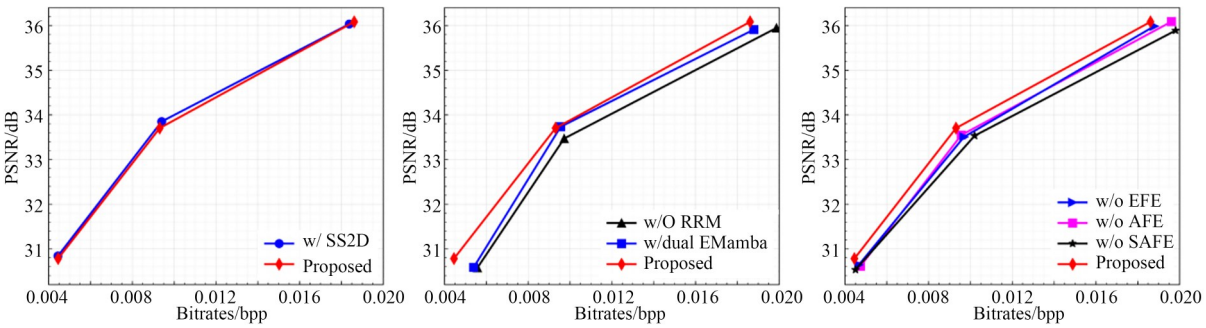
为深入探究EMLFIC的有效性，本文将其与若干经过调整的变体进行对比分析，旨在分别验证通道高效的扫描机制CE2DSS、光场解码器中的残差重建模块RRM和光场编码器中各模块的有效性。设定$\lambda\in\{0.000\,1, 0.000\,5, 0.002\}$作为率失真性能评估的参数范围，以全面比较各方法的性能表现。

CE2DSS和SS2D：为验证所提出的CE2DSS的有效性，采用VMamba[31]中的传统2D选择性扫描方案SS2D作为对比基准，记为“w/SS2D”。SS2D方案未对光场图像特征进行跳跃采样，而是将所有Token作为输入。此外，SS2D方案未对光场图像特征的通道进行分组，而是将所有通道复制四份，并按四个不同方向展平为一维序列输入到状态空间模型SSM中，最终将输出结果相加。如图10(a)所示，在多个码率点上对CE2DSS与SS2D进行的PSNR性能对比表明，二者在重建质量上表现相当。这一观察在表2的BD-BR和BD-PSNR度量结果中得到进一步验证，SS2D相较于CE2DSS在BD-PSNR和BD-BR指标上分别仅有0.084 dB和2.425%的优势，这表明所提出的CE2DSS方法仅以极小的性能损失为代价。为更精确地量化扫描模块本身的效率提升，表3进一步对比了CE2DSS与SS2D在Mamba扫描模块上的计算复杂度。结果表明，相较于SS2D，CE2DSS显著降低了扫描模块的计算负担。CE2DSS的FLOPs从SS2D的5.973 G大幅降至3.253 G，降幅约45.54%，参数量也从SS2D的36.480 K显著减少至20.352 K，降幅约44.21%。这种现象源于所提出的CE2DSS通过对通道进行分组并采用跳跃采样的扫描策略，大幅减少了特征提取时的参数量，同时保留了输入数据的完整性，从而能够有

表2 所提方法与其他变体在EPFL数据集上的BD-BR和BD-PSNR

Tab. 2 BD-BR and BD-PSNR of the proposed method compared with other variants on the EPFL test dataset

	Methods	BD-BR /%	BD-PSNR/dB
Proposed *v. s.*	w/SS2D	2.425	−0.084
	w/o RRM	−12.518	0.476
	w/dual EMamba	−6.260	0.217
	w/o EFE	−8.003	0.315
	w/o AFE	−6.958	0.268
	w/o SAFE	−11.659	0.463



(a) CE2DSS和SS2D (a) CE2DSS and SS2D　(b) RRM和dual EMamba (b) RRM and dual EMamba　(c) EFE,AF和dual SAFE (c) EFE, AFE and SAFE

图10 不同变体在EPFL测试集上消融实验的率失真曲线图

Fig. 10 Rate-distortion curves of ablation experiments for different variants on the EPFL test dataset

表 3　CE2DSS 和 SS2D 的计算复杂度对比

Tab. 3　Comparison of computational complexity between CE2DSS and SS2D

Methods	FLOPs/G	Params/K
SS2D	5.973	36.480
CE2DSS	3.253	20.352

效学习特征的长程依赖关系。因此，CE2DSS 仅以极小的学习能力损失为代价，实现了显著的效率提升。

RRM 和 dual EMamba：为验证所提出的 RRM 的有效性，将 RRM 从 EMLFIC 中移除，记为“w/o RRM”。如图 10(b)所示，与完整的 EMLFIC 相比，移除 RRM 后，模型的 BD-PSNR 减少了 0.476 dB，BD-BR 增加了 12.518%。这是因为所提出的 RRM 能有效学习初步解码后的特征相关性，进一步减少了冗余信息。为了确保特征编解码过程的对称性，通常在解码器中添加一个与主编码器相同的反向特征提取模块。基于此，进一步探讨了在主解码器 g_s 后串联 EMamba 的效果，记为“w/ dual EMamba”。实验结果表明，在解码端直接使用 EMamba 对性能的提升效果不如 RRM 显著。此外，如表 4 所示，“Propsoed/dual EMamba”还增加了编码和解码时间以及计算复杂度，这一结果进一步验证了 RRM 在设计上的合理性。

表 4　所提方法与其他方法的编解码时间和计算复杂度对比

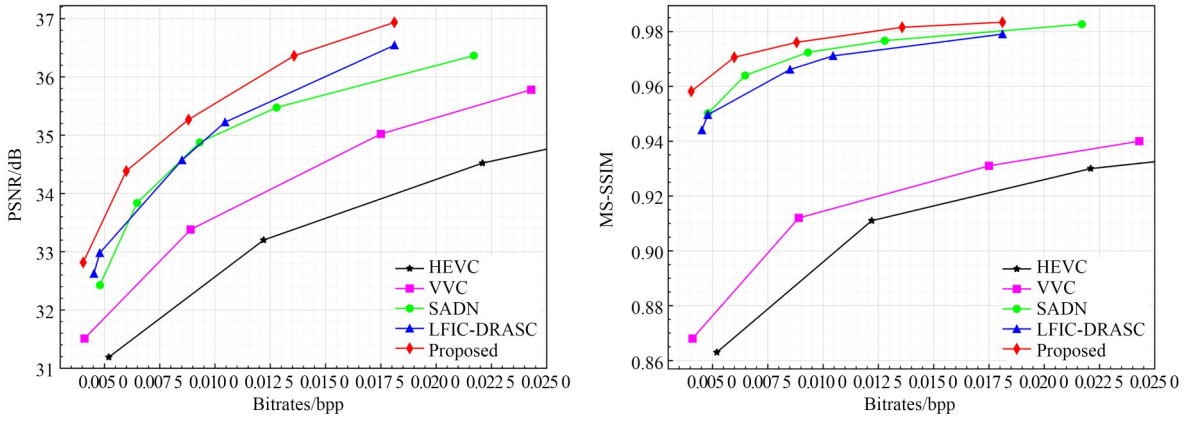
Tab. 4　Comparison of encoding/decoding time and computational complexity between the proposed method and other methods

Methods	Encoding time/s	Decoding time/s	FLOPs/G	Params/M
SADN	68.84	136.48	60.43	3.46
MLIC++	10.81	13.24	231.42	83.50
Proposed/dual EMamba	6.86	7.78	114.87	20.31
Proposed/SS2D	7.93	6.29	94.14	19.96
Proposed	6.75	6.18	91.77	19.95

EFE，AFE 和 SAFE：为验证 EMamba 中各模块的有效性，引入了三种变体进行对比分析。具体而言，分别将 EFE，AFE 以及 SAFE 从 EMamba 中移除，并分别记为“w/o EFE”，“w/o AFE”和“w/o SAFE”。如图 10(c)和表 2 所示，与完整的 EMLFIC 相比，这三种变体的 BD-PSNR 分别减少了 0.315 dB，0.268 dB 和 0.463 dB，BD-BR 分别增加了 8.003%，6.958% 和 11.659%。这一结果表明，仅提取光场图像的部分子空间信息仍会存在一定的冗余，无法完全解开光场图像中空间特征与角度特征之间的耦合关系。联合利用 EPI 域视差信息和光场图像中的空间与角度信息，能更好地降低 4D 光场的空间-角度冗余，实现更优的压缩性能。这进一步验证了 EMamba 在特征提取和冗余消除方面的有效性。

3.5　计算效率分析

本文将所提方法的编解码时间及计算复杂度与上述对比方法进行对比评估，结果如表 4 所示。SADN 的编码和解码时间分别为 68.84 s 和 136.48 s，MLIC++ 的编码和解码时间分别为 10.81 s 和 13.24 s，而所提方法 EMLFIC 的编码和解码时间仅为 6.75 s 和 6.18 s。相比于 SADN 采用的自回归上下文模型，所提方法和 MLIC++ 均采用了基于通道均匀切片与棋盘格上下文结合的熵模型网络，这显著缩短了编码和解码时间。由于所提方法和 SADN 在主编解码器中均采用了相对较低的 48 个通道数，而非 MLIC++ 中端到端图像编码常用的 192 个通道数，因此总体参数量更少。此外，相较于 MLIC++ 基于 Transformer 的全局和局部熵模型，EMLFIC 具有更少的浮点运算次数，从而实现了更低的计算复杂度和更短的编解码时间。尽管与 SADN 相比，EMLFIC 的浮点运算次数和参数量有所增加，但考虑到其在压缩效率和编解码时间上的显著提升，这一代价是可以接受的。



(a) PSNR率失真曲线
(a) PSNR rate-distortion curve

(b) MS-SSIM率失真曲线
(b) MS-SSIM rate-distortion curve

图 11　EPFL 测试数据集上不同方法的率失真曲线性能比较(角度分辨率为 13×13)

Fig. 11　Performance comparison of RD curve of different methods on the EPFL test dataset(Angle resolution is 13×13)

总体而言,所提方法在编解码时间上表现出色,同时实现了更优的压缩性能。

3.6　角度分辨率适用性研究

为了进一步验证所提方法 EMLFIC 对不同角度分辨率的光场图像压缩的适用性,本节针对角度分辨率为 13×13 的光场图像进行了系统性实验分析。实验设置严格遵循 SADN[22]的基准配置:训练集包含从 PINET[22]数据集中提取的 43 099 个光场图像块,每个图像块的角度分辨率为 13×13,尺寸为 832 pixels×832 pixels。测试数据为 EPFL 数据集[43]提供的 12 个不同场景。为确保结果的可比性,所有对比方法(包括 HEVC[5],VVC[6],SADN[22]和 LFIC-DRASC[24])的实验数据均由原作者团队直接提供。由于 MLIC++[18]的模型参数过大,无法在角度分辨率为 13×13 的光场图像上进行训练,因此本节没有对比 MLIC++[18]方法。如图 11 所示,所提方法 EMLFIC 在 13×13 角度分辨率条件下较 7×7 角度分辨率展现出更显著的性能提升。与 HEVC[5],VVC[6],SADN[22]和 LFIC-DRASC[24]相比,所提方法分别平均节省了 68.935%,52.997%,24.163% 和 19.575% 的码率,而 BD-PSNR 分别提高了 2.853 dB,1.876 dB,0.742 dB 和 0.587 dB。这一现象可归因于更高角度分辨率蕴含了更强的角度冗余特性,13×13 角度分辨相比于前文的 7×7 角度分辨率,其视角数量从 49 个增加到 169 个,角度冗余度提升了约 3.5 倍。这也从另一方面有效验证了所提方法在建模长程空间-角度依赖性方面的优势。实验结果表明,所提方法对不同角度分辨率的光场图像压缩均具有良好适应性,且性能增益与角度分辨率呈正相关性。

4　结　论

本文提出了一种高效 Mamba 驱动的端到端光场图像压缩方法 EMLFIC,针对光场图像高维特性带来的压缩挑战,通过高效建模全局特征和长程依赖关系,有效提升了压缩性能。具体而言,从 4D 光场图像中提取包含空间信息和极平面信息的 2D 切片,并利用提出的 EMamba 对光场图像进行全局建模,显著增强了其对长距离相关性的捕捉能力。在此基础上,设计了一种通道高效的 2D 选择性扫描策略,通过多方向跳跃扫描实现高效特征提取,同时有效降低了计算复杂度。此外,解码端的残差重建模块进一步提升了重建质量并优化了解码端的计算效率。实验结果表明,所提出的 EMLFIC 在率失真性能上表现出色。与代表性方法 SADN 相比,所提方法在 7×7 角度分辨率的光场图像上平均实现了 7.4% 的码率节省和 0.37 dB 的 PSNR 提升,同时编解码速度提升了 10~20 倍。此外,与最新方法 LF-IC-DRASC 相比,在 13×13 角度分辨率下,所提方法进一步实现了 19.575% 的码率节省和

0.587 dB的PSNR提升，展现出了更高的压缩性能。总的来说，从模型压缩效率、计算复杂度和编码解码时间的综合性能来看，EMLFIC在整体上展现出了更具竞争力的效果。未来研究可通过引入区域自适应建模、多尺度特征分析或增强角度一致性建模等方式进一步优化性能。

作者贡献声明：

封哲宇：测量方法的提出，论文构思和撰写；
蒋志迪：论文审核与编辑写作；
万立飞：论文审核与编辑写作；
徐海勇：论文审核与编辑写作；
蒋刚毅：论文指导，审核与编辑写作。

参考文献：

[1] SAMARAKOON T, ABEYWARDENA K, EDUSSOORIYA C U S. Arbitrary Volumetric Refocusing of Dense and Sparse Light Fields [EB/OL]. (2025-02-26). https://arxiv.org/abs/2502.19238.

[2] 吕晓波，刘宇丰，李毅威，等. 快照式光谱光场成像技术[J]. 光学精密工程，2021，29(2)：220-230.
LÜ X B, LIU Y F, LI Y W, *et al*. Snapshot spectral light-field imaging technology[J]. *Opt. Precision Eng.*, 2021, 29(2): 220-230. (in Chinese)

[3] LEVOY M, HANRAHAN P. *Light Field Rendering*[M]. New York: ACM Press, 2023: 441-452.

[4] DONG L, WANG L Z, LI L, *et al*. Pseudo-sequence-based light field image compression [C]. 2016 *IEEE International Conference on Multimedia & Expo Workshops* (*ICMEW*). *July* 11-15, 2016, *Seattle*, *WA*. IEEE, 2016: 1-4.

[5] SULLIVAN G J, OHM J R, HAN W J, *et al*. Overview of the high efficiency video coding (HEVC) standard[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2012, 22(12): 1649-1668.

[6] BROSS B, CHEN J L, OHM J R, *et al*. Developments in international video coding standardization after AVC, with an overview of versatile video coding (VVC)[J]. *Proceedings of the IEEE*, 2021, 109(9): 1463-1493.

[7] SHAO J R, BAI E J, JIANG X Q, *et al*. Light-field image compression based on a two-dimensional prediction coding structure[J]. *Information*, 2024, 15(6): 339.

[8] HUANG X P, AN P, CHEN Y L, *et al*. Low bitrate light field compression with geometry and content consistency[J]. *IEEE Transactions on Multimedia*, 2020, 24: 152-165.

[9] ZHANG Y Z, WAN L F, MAO Y F, *et al*. Geometry-aware view reconstruction network for light field image compression[J]. *Scientific Reports*, 2022, 12: 22254.

[10] LIU D Y, HUANG Y, FANG Y M, *et al*. Multi-stream dense view reconstruction network for light field image compression[J]. *IEEE Transactions on Multimedia*, 2022, 25: 4400-4414.

[11] SHENG H, ZHAO P, ZHANG S, *et al*. Occlusion-aware depth estimation for light field using multi-orientation EPIs[J]. *Pattern Recognition*, 2018, 74: 587-599.

[12] LIU D Y, AN P, MA R, *et al*. Content-based light field image compression method with Gaussian process regression[J]. *IEEE Transactions on Multimedia*, 2020, 22(4): 846-859.

[13] SCHIOPU I, MUNTEANU A. Deep-learning-based macro-pixel synthesis and lossless coding of light field images[J]. *APSIPA Transactions on Signal and Information Processing*, 2019, 8(1): 20.

[14] MINNEN D, BALLÉ J, TODERICI G D. Joint autoregressive and hierarchical priors for learned image compression[C]. *Advances in Neural Information Processing Systems*, 2018: 10794-10803.

[15] MINNEN D, SINGH S. Channel-wise autoregressive entropy models for learned image compression [C]. 2020 *IEEE International Conference on Image Processing* (*ICIP*). *October* 25-28, 2020. *Abu Dhabi*, *United Arab Emirates*. IEEE, 2020: 3339-3343.

[16] HE D L, ZHENG Y Y, SUN B C, *et al*. Checkerboard context model for efficient learned image compression[C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 20-25, 2021. *Nashville*, *TN*, *USA*. IEEE, 2021: 14766-14775.

[17] JIANG W, YANG J Y, ZHAI Y Q, *et al*. MLIC: multi-reference entropy model for learned image compression[C]. *Proceedings of the* 31*st*

ACM International Conference on Multimedia. Ottawa ON Canada. ACM, 2023: 7618-7627.

[18] JIANG W, WANG R. MLIC++: Linear complexity multi-reference entropy modeling for learned image compression[C]. *ICML 2023 Workshop Neural Compression: From Information Theory to Applications. Honolulu, HI*, USA, 2023.

[19] 程俊, 郁梅, 蒋刚毅. 结合视差补偿与3D数据处理的盲光场图像质量评价[J]. 光学 精密工程, 2023, 31(8): 1202-1216.
CHENG J, YU M, JIANG G Y. Blind light field image quality assessment combining disparity compensation with 3D data processing[J]. *Opt. Precision Eng.*, 2023, 31(8): 1202-1216. (in Chinese)

[20] SINGH M, RAMESHAN R M. Learning-based practical light field image compression using a disparity-aware model[C]. 2021 *Picture Coding Symposium (PCS). June* 29-*July* 2, 2021. *Bristol, United Kingdom.* IEEE, 2021: 1-5.

[21] ZHONG T T, JIN X, TONG K D. 3D-CNN autoencoder for plenoptic image compression [C]. 2020 *IEEE International Conference on Visual Communications and Image Processing (VCIP). December* 1-4, 2020, *Macau, China.* IEEE, 2020: 209-212.

[22] TONG K D, JIN X, WANG C, *et al*. SADN: learned light field image compression with spatial-angular decorrelation [C]. *ICASSP* 2022 - 2022 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). May* 23-27, 2022, *Singapore, Singapore.* IEEE, 2022: 1870-1874.

[23] YE K S, LI Y, LI G, *et al*. End-to-end light field image compression with multi-domain feature learning[J]. *Applied Sciences*, 2024, 14(6): 2271.

[24] FENG S Y, ZHANG Y, ZHU L W, *et al*. LFIC-DRASC: deep light field image compression using disentangled representation and asymmetrical strip convolution[J]. *IEEE Transactions on Broadcasting*, 2025, 71(3): 889-902.

[25] VASWANI A, SHAZEER N, PARMAR N, *et al*. Attention is all you need[C]. *Advances in Neural Information Processing Systems*, 2017.

[26] TONG K D, JIN X, YANG Y Q, *et al*. Learned focused plenoptic image compression with microimage preprocessing and global attention [J]. *IEEE Transactions on Multimedia*, 2023, 26: 890-903.

[27] LIU G S, YUE H J, WEN B H, *et al*. Learned focused plenoptic image compression with local-global correlation learning[J]. *IEEE Transactions on Multimedia*, 2025, 27: 1216-1227.

[28] GU A, GOEL K, RÉ C, *et al*. Efficiently Modeling Long Sequences with Structured State Spaces [EB/OL]. (2021-10-31) [2022-08-05]. https://arxiv.org/abs/2111.00396.

[29] GU A, JOHNSON I, GOEL K, *et al*. Combining recurrent, convolutional, and continuous-time models with linear state-space layers [C]. *Neural Information Processing Systems*, 2008.

[30] GU A, DAO T. Mamba: Linear-time Sequence Modeling with Selective State Spaces [EB/OL]. (2023-12-01) [2024-05-31]. https://arxiv.org/abs/2312.00752.

[31] LIU Y, TIAN Y, ZHAO Y, *et al*. VMamba: Visual state space model[J]. *Advances in Neural Information Processing Systems*, 2024, 37: 103031-103063.

[32] ZHU L, LIAO B, ZHANG Q, *et al*. Vision Mamba: Efficient Visual Representation Learning with Bidirectional State Space Model [EB/OL]. 2024: 2401.09417. https://arxiv.org/abs/2401.09417.

[33] QIN S, WANG J, ZHOU Y, *et al*. MambaVC: Learned Visual Compression with Selective State Spaces [EB/OL]. (2024-05-24) [2024-05-28]. https://arxiv.org/abs/2405.15413.

[34] CHENG Z X, SUN H M, TAKEUCHI M, *et al*. Learned image compression with discretized gaussian mixture likelihoods and attention modules[C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June* 13-19, 2020. *Seattle, WA, USA.* IEEE, 2020: 7939-7948.

[35] LIANG Z Y, WANG Y Q, WANG L G, *et al*. Learning non-local spatial-angular correlation for light field image super-resolution[C]. 2023 *IEEE/CVF International Conference on Computer Vision (ICCV). October* 1-6, 2023, *Paris, France.* IEEE, 2023: 12342-12352.

[36] LU Y, WANG S, WANG Z, *et al*. LFMamba: light field image super-resolution with state space model [EB/OL]. (2024-06-18). https://arxiv.org/abs/2406.12463.

[37] GAO R S, XIAO Z Y, XIONG Z W. Mamba-based light field super-resolution with efficient subspace scanning [C]. *Computer Vision-ACCV 2024. Singapore: Springer Nature Singapore*, 2025: 421-437.

[38] ZHANG Y L, LI K P, LI K, *et al*. Image super-resolution using very deep residual channel attention networks[C]. *Computer Vision-ECCV 2018. Cham: Springer International Publishing*, 2018: 294-310.

[39] GUO H, LI J M, DAI T, *et al*. *MambaIR: A Simple Baseline for Image Restoration with State-Space Model* [M]. Computer Vision - ECCV 2024. Cham: Springer Nature Switzerland, 2024: 222-241.

[40] ELFWING S, UCHIBE E, DOYA K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning [J]. *Neural Networks*, 2018, 107: 3-11.

[41] PEI X, HUANG T, XU C. EfficientVmamba: Atrous Selective Scan for Light Weight Visual Mamba [EB/OL]. (2024-03-15). https://arxiv.org/abs/2403.09977.

[42] BÉGAINT J, RACAPÉ F, FELTMAN S, *et al*. CompressAI: a PyTorch library and evaluation platform for end-to-end compression research [EB/OL]. (2020-11-05). https://arxiv.org/abs/2011.03029.

[43] RERABEK M, EBRAHIMI T. New light field image dataset[C]. *8th International Conference on Quality of Multimedia Experience* (*QoMEX*). *Lisbon*, *Portugal*, 2016.

[44] HEVC Official Test Model[EB/OL]. https://vcgit.hhi.fraunhofer.de/jvet/HM/-/tags.

[45] VVC Official Test Model[EB/OL]. https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM.

作者简介：

封哲宇（2000－），男，浙江海宁人，硕士研究生，2023年于宁波大学获得学士学位，现为宁波大学信息科学与工程学院研究生，主要从事光场图像压缩等方面的研究。E-mail：15068368500@163.com

通讯作者：

蒋刚毅（1964－），男，浙江绍兴人，博士，教授，博士生导师，2000年于韩国Ajou大学获得博士学位，主要从事多媒体信号处理与通信、计算成像、视觉感知与编码、图像与视频质量评价等方面的研究工作。E-mail：jianggangyi@nbu.edu.cn