计 算 机 研 究 与 发 展
Journal of Computer Research and Development
https://crad.ict.ac.cn
63(2): 448−459, 2026

基于重构难度感知的视频快照压缩成像方法

陈 灿[1,2] 石鹏飞[1] 周 超[1] 胡 俊[2] 张登银[1]
[1](南京邮电大学物联网学院 南京 210003)
[2](江苏有线技术研究院 南京 210018)
(chencan@njupt.edu.cn)

Video Snapshot Compressive Imaging Method Based on Reconstruction Difficulty Perception

Chen Can[1,2], Shi Pengfei[1], Zhou Chao[1], Hu Jun[2], and Zhang Dengyin[1]
[1](*College of Internet of Things*, *Nanjing University of Posts and Telecommunications*, *Nanjing* 210003)
[2](*Jiangsu Cable Television Research Institute*, *Nanjing* 210018)

Abstract Artificial intelligence applications require highly advanced high-speed video imaging technologies to perceive the surrounding environment better. Deep learning-based snapshot compressive imaging (SCI) offers a promising solution. How to reconstruct high-speed videos from observed data using deep learning techniques is a frontier hotspot in the field. However, existing reconstruction methods focus on mining prior information, neglecting the direct influence of masks and image textures on reconstruction difficulty. There is still room for further improvement in reconstruction quality. To address this issue, we propose a reconstruction difficulty perception-based SCI (RdpSCI) method. Based on the observation that masks and image textures jointly determine the information contained in the observed data, exploration of the correlation between masks, image textures, and reconstruction difficulty are first proposed by RdpSCI to guide deep networks for reconstruction. Specifically, it introduces an improved residual dense network (I-ResDNet) module, innovatively incorporating channel shuffling operations into ResDNet to reduce the dependency of feature fusion effects on channel partitioning methods. The proposed I-ResDNet also introduces a reconstruction difficulty weight vector to guide feature fusion, enhancing feature fusion capability without significantly increasing model parameters. Experiments show that compared with the state-of-the-art methods STFormer and EfficientSCI, RdpSCI achieves improvements of 0.68 dB and 0.54 dB in reconstruction quality on benchmark grayscale and colour datasets, respectively.

Key words computational imaging; snapshot compressive imaging (SCI); video compressive sensing; video reconstruction; reconstruction difficulty perception

摘 要 人工智能应用需要有与之适配的高速视频成像新技术以便更好地感知周围环境,基于深度学习的快照压缩成像(snapshot compressive imaging, SCI)提供了一种具有前景的解决方案。如何利用深度学习技术从观测值中重构高速视频是领域的前沿热点。然而,现有重构方法注重挖掘先验信息以提升重构质量,忽略了掩码和图像纹理对重构难度的影响,使得重构质量仍有进一步提高的空间。针对这个问题,提出一种基于重构难度感知的快照压缩成像方法RdpSCI(reconstruction difficulty perception-based SCI)。

收稿日期: 2024-08-13; **修回日期**: 2025-04-29
基金项目: 国家自然科学基金项目(62471241); 南京邮电大学引进人才科研启动基金项目(NY222084)
This work was supported by the National Natural Science Foundation of China (62471241) and the Startup Foundation for Introducing Talent of Nanjing University of Posts and Telecommunications (NY222084).
通信作者: 周超(zhouchao@njupt.edu.cn)

基于观测值中包含的信息由掩码和图像纹理共同决定观察，所提方法提出挖掘掩码和图像纹理与重构难度之间的关联，引导深度网络进行重构。特别地，提出一种改进残差密集连接网络模块 I-ResDNet（improved ResDNet），通过引入重构难度权重向量引导特征融合，并创新地在 ResDNet 中引入通道打乱操作，降低特征融合效果对于通道划分方式的依赖，能够在不显著增加模型参数量的情况下增强特征融合能力。实验表明，RdpSCI 相比于领域现有最优方法 STFormer 和 EfficientSCI，在基准灰度数据集和基准彩色数据集上，重构结果的峰值信噪比分别有 0.68 dB 和 0.54 dB 的提升。

关键词 计算成像；快照压缩成像；视频压缩感知；视频重构；重构难度感知

中图法分类号 TP399；TN919.8

DOI: 10.7544/issn1000-1239.202440681 **CSTR:** 32373.14.issn1000-1239.202440681

随着人工智能应用的演进，由于存储和传输成本问题，高维视频数据的高效处理正成为大规模深度学习的瓶颈[1]。在自动驾驶、机器人等人工智能应用中，目标跟踪[2]、语义分割[3]等机器视觉技术的成功运用依赖高维视频成像，以便更好地感知周围环境，提高智能决策的精确性。通过将光学传感器和优化方法相结合，快照压缩成像（snapshot compressive imaging, SCI）提供了一种高效处理高维视频数据的方案。SCI 首先在采样端用 2D 相机拍摄快照以捕捉视频数据的观测值，再在重构端利用优化方法恢复原始视频数据[4-5]。与传统成像方法相比，SCI 具有低带宽、低能耗以及快采集等优势，具有广泛的应用，例如，高光谱成像[6]和高速视频成像[7]。本文主要研究高速视频成像，即如何从观测值中恢复原始视频数据。当前主流 SCI 重构方法可以分为基于分析模型的方法和基于深度学习的方法 2 类。

基于分析模型的 SCI 重构方法首先利用先验信息构建分析模型，再通过迭代求解该模型重构视频。文献 [8] 对传统彩色摄像机进行改进，设计了一种彩色压缩感知相机，该相机利用广义交替投影（generalized alternating projection, GAP）方法从观测值中重构图像，能够恢复聚焦在不同深度的多个彩色帧；文献 [9] 提出将重构问题建模为一个总变分（total variation, TV）最优化问题，再利用 GAP 方法进行求解。文献 [10] 提出一种名为 DeSCI 的方法，该方法利用视频或高光谱的非局部相似性和低秩先验构建了一个联合模型，并进一步研究了 SCI 采样的特殊结构，以降低 SCI 重构的计算负担和内存占用。尽管研究者围绕基于分析模型的重构方法进行了广泛的探索，部分方法能够获得不错的重构质量，但由于所采用的人工设计图像先验缺乏足够的精确性，并且往往需要多次迭代求解才能获得较为理想的重构结果，导致重构效率较低，难以满足实时应用场景的要求[11-15]。

随着深度学习技术的快速发展，基于深度学习的 SCI 重构方法得到了学者的广泛关注。基于深度学习的主流 SCI 重构方法可以分为端到端方法、深度展开方法和即插即用（plug-and-play, PnP）方法 3 种类型。文献 [16] 提出构建端到端的全连接网络以进行视频压缩感知重构，并对该网络的实用性进行了讨论。文献 [17] 提出将卷积神经网络用于快照压缩成像重构，有效地利用了视频先验信息。文献 [18] 提出了一种名为 BIRNAT 的端到端网络，该网络首先使用残差块和特征图自注意力机制重构第 1 帧，然后使用双向循环神经网络按序重构后续帧。为了探索更加有效的网络架构，研究者尝试将 Transformer 应用于快照压缩成像重构，通过引入注意力机制对关键部分进行重点处理[19-20]。针对现有端到端方法不能对自适应掩码和输入尺度变化的问题，文献 [21] 提出一种名为 STFormer 的端到端网络，该网络通过时空分解和局部自注意力机制，提高了重构模型对多尺度输入的灵活性。针对高分辨率视频成像面临的计算复杂度较高的问题，文献 [22] 提出了一种名为 EfficientSCI 的端到端网络，该网络通过在残差块中使用密集连接和时空分解机制，有效降低了计算复杂度。然而，该类方法由于缺乏可解释性而饱受诟病，随后研究者提出了基于深度展开的重构方法[23]。文献 [24] 提出一种名为 GAP-net 的网络，该网络将 GAP 方法求解 SCI 问题的每次迭代转化为一层深度网络，并通过构建多层深度网络逐步重构；类似地，文献 [25] 将半二次分裂（half-quadratic splitting, HQS）方法进行展开，并利用 3D 卷积挖掘视频帧之间的时空相关性进行重构。训练好的深度展开网络等效于执行有限次数的迭代求解，增强了深度网络的可解释性。为了进一步提高重构鲁棒性和灵活性，研究者提出基于 PnP 的方法，利用现有去噪器或预训练的去噪网络直接替代分析模型迭代求解中的投

影步骤。文献 [26] 提出了一种名为 PnP-FFDNet 的方法, 该方法将深度去噪器集成到 PnP 框架当中, 缓解了重构效率低以及灵活性差的问题。文献 [27] 提出了一种名为 PnP-CASS 的方法, 该方法使用去噪网络作为高光谱 SCI 的正则化先验, 能够有效权衡重构质量和效率, 并且可以应用于不同编码机制。针对数据分布不匹配的情况, 文献 [28] 提出了一种在线 PnP 方法, 该方法可以在 PnP 迭代内自适应地更新网络参数, 使得去噪网络更适用于具体应用任务。

SCI 重构作为图像逆问题的一种应用, 其本质在于利用观测值中包含的有限信息进行重构。观测值中包含的信息越多、观测过程中丢失的信息越少, 重构的难度越低, 而观测值中包含的信息与掩码和图像纹理 2 方面有关。现有重构方法注重挖掘先验信息以提升重构质量, 忽略了掩码和图像纹理对重构难度的影响, 使得重构质量仍有进一步提高的空间。针对这个问题, 本文提出一种基于重构难度感知的快照压缩成像(reconstruction difficulty perception-based SCI, RdpSCI)方法, 主要创新点有 3 点:

1)基于观测值中包含的信息与掩码和图像纹理 2 方面有关的观察, 提出挖掘掩码和图像纹理与重构难度之间的关联, 引导深度网络进行重构。

2)提出一种改进残差密集连接网络模块 I-ResDNet (improved ResDNet), 通过引入重构难度权重向量引导特征融合, 并创新地在 ResDNet 中引入通道打乱操作, 降低特征融合效果对于通道划分方式的依赖, 能够在不显著增加模型参数量的情况下增强特征融合能力。

3)实验表明, RdpSCI 相比于领域现有最优方法 STFormer 和 EfficientSCI, 在基准灰度数据集和基准彩色数据集上, 重构质量分别有 0.68 dB 和 0.54 dB 的提升。

1 问题描述与动机分析

1.1 视频 SCI 数学模型

如图 1 所示, 假设一个包含 B 帧的视频帧组 $\boldsymbol{X}=\{\boldsymbol{X}_i\}_{i=1}^{B}(\boldsymbol{X}_i\in\mathbb{R}^{w\times h})$, 首先利用一个包含 B 帧的掩码组 $\boldsymbol{M}=\{\boldsymbol{M}_i\}_{i=1}^{B}(\boldsymbol{M}_i\in\mathbb{R}^{w\times h})$对视频帧组进行逐帧采样, 再进行时间维度上的累加运算, 得到观测值$\boldsymbol{Y}\in\mathbb{R}^{w\times h}$:

$$\boldsymbol{Y}=\sum_{i=1}^{B}\boldsymbol{X}_i\odot\boldsymbol{M}_i+\boldsymbol{Z}\,, \tag{1}$$

其中$\boldsymbol{Z}\in\mathbb{R}^{w\times h}$为观测噪声, $\odot$代表对应元素相乘运算, 系统的压缩比为B。为了保证B个掩码具有较低的相关性, 采用高斯随机矩阵构造掩码组$\boldsymbol{M}$。利用向量化操作$Vec(\cdot)$, 式(1)可以进行向量化表示。

$$\boldsymbol{y}=\boldsymbol{H}\boldsymbol{x}+\boldsymbol{z}\,, \tag{2}$$

其中$\boldsymbol{y}=Vec(\boldsymbol{Y})\in\mathbb{R}^{w\times h}$, $\boldsymbol{z}=Vec(\boldsymbol{Z})\in\mathbb{R}^{w\times h}$, $\boldsymbol{H}=(\boldsymbol{D}_1,\boldsymbol{D}_2,\cdots,\boldsymbol{D}_B)\in\mathbb{R}^{w\times h\times w\times h\times B}$, $\boldsymbol{D}_i\in\mathbb{R}^{w\times h\times w\times h}$为由$Vec(\boldsymbol{M}_i)$元素组成的对角阵。$\boldsymbol{x}\in\mathbb{R}^{w\times h\times B}$可以表示为

$$\boldsymbol{x}=(Vec(\boldsymbol{X}_1)^{\mathrm{T}},Vec(\boldsymbol{X}_2)^{\mathrm{T}},\cdots,Vec(\boldsymbol{X}_B)^{\mathrm{T}})^{\mathrm{T}}\,。 \tag{3}$$

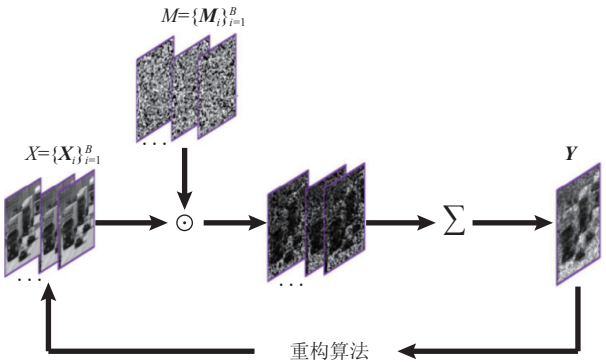
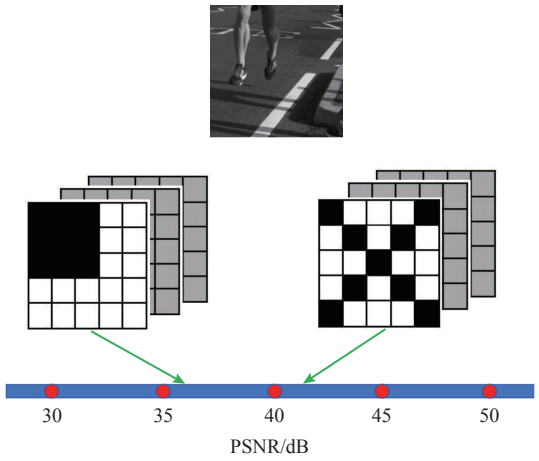


Fig. 1 Framework of video SCI

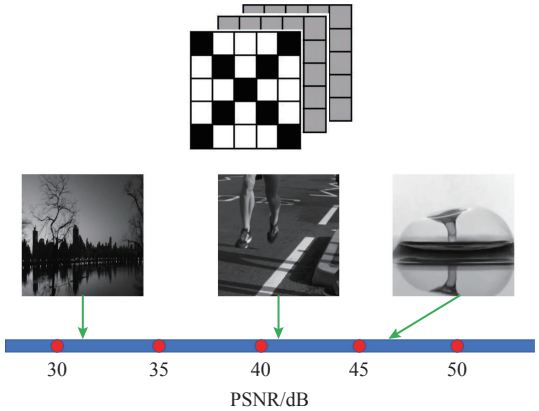
图 1 视频快照压缩成像框架

1.2 研究动机

图 2 对比了采用不同掩码和纹理情况下的平均峰值信噪比(peak signal-to-noise ratio, PSNR), 可见掩码和图像纹理共同影响观测值中包含的信息, 进而影响重构难度。1)观测值由对每帧的采样值进行累加得到, 由于掩码$\boldsymbol{M}_i$由{0,1}元素组成(“1”表示该处进行采样, “0”表示该处不进行采样), 掩码元素的分布决定了所采样信息的分布, 采样信息的分布直接影响重构难度。图 2(a)展示了采用不同掩码情况下, 同一视频序列的重构质量对比。不难观察到, 采用右侧掩码的重构质量明显优于采用左侧掩码的重构质量。其原因在于: 左侧掩码元素分布过于集中, 使得所采样信息的分布过于集中, 易造成集中区域信息的采样冗余和其余区域信息的采样丢失, 导致重构难度增加; 而右侧掩码元素分布相对较为均衡, 使得所采样信息的分布较为均衡, 各区域信息均有采样, 使得重构难度较低。2)图像帧不同区域的纹理并不相同, 纹理不同导致重构难度不同[29-30]。图 2(b)展示了采用相同掩码情况下, 不同视频序列的重构质量对比。不难观察到, 在掩码元素分布较为均衡的情况下, 图像纹理较为平滑的视频序列重构质量明显优于纹理较为复杂的视频序列。其原因在于: 平滑的图像纹理包含信息较少, 重构难度较低; 复杂的图像纹理包含信息较多, 重构难度较高。因此, 掩



(a) 掩码对重构难度的影响



(b) 图像纹理对重构难度的影响

Fig. 2 Observation of factors affecting reconstruction difficulty (reconstruction based on STFormer[21]method)

图 2 影响重构难度的因素观察（基于 STFormer[21] 方法的重构）

码和图像纹理共同影响重构难度，而现有重构方法注重挖掘先验信息以提高重构质量，忽略了这种关联，导致重构质量仍有进一步提高的空间。

在掩码元素分布较为均衡的情况下，图像纹理较为平滑的视频序列重构质量明显优于纹理较为粗糙的视频序列。其原因在于：平滑的图像纹理包含信息较少，重构难度较低；粗糙的图像纹理包含信息较多，重构难度较高。因此，掩码和图像纹理共同影响重构难度，而现有重构方法注重挖掘先验信息以提高重构质量，忽略了这种关联，导致重构质量仍有进一步提高的空间。

2 RdpSCI 介绍

所提出的 RdpSCI 框架如图 3 所示，该框架主要由预处理模块、重构难度感知模块和 I-ResDNet 模块 3 类模块构成。

2.1 预处理模块

相关研究表明，直接利用深度网络对 $\boldsymbol{Y}$ 进行重构并不高效[18]。因此，首先利用预处理模块对 $\boldsymbol{Y}$ 进行处理，通过式(4)获得包含背景和运动轨迹信息的归一化观测值 $\bar{\boldsymbol{Y}} \in \mathbb{R}^{w\times h}$。

$$\bar{\boldsymbol{Y}} = \boldsymbol{Y} / \sum_{i=1}^{B} \boldsymbol{M}_i , \tag{4}$$

其中“/”代表对应元素相除运算。然后，通过式(5)对 $\bar{\boldsymbol{Y}}$ 进行处理获得初始重构帧

$$\begin{aligned} &\boldsymbol{X}^{\mathrm{e}} = \{\boldsymbol{X}_i^{\mathrm{e}}\}_{i=1}^{B} (\boldsymbol{X}_i^{\mathrm{e}} \in \mathbb{R}^{w\times h}) , \\ &\boldsymbol{X}^{\mathrm{e}} = \bar{\boldsymbol{Y}} \odot \boldsymbol{M}_i + \bar{\boldsymbol{Y}} 。 \end{aligned} \tag{5}$$

2.2 重构难度感知模块

基于观测值中包含的信息与掩码和图像纹理 2 方面有关的观察，提出基于注意力机制的重构难度感知模块，挖掘掩码和图像纹理与重构难度之间的关联，引导深度网络进行重构。所提出的重构难度感知模块如图 4 所示。考虑归一化观测值 $\bar{\boldsymbol{Y}}$ 能够一定程度反映图像纹理信息，因此首先分别利用 2 个卷积层(Conv)对 $\bar{\boldsymbol{Y}}$ 和 $\boldsymbol{M}$ 进行特征提取：

$$\boldsymbol{Q} = Conv_{3\times3}(Conv_{3\times3}(\bar{\boldsymbol{Y}})) , \tag{6}$$

$$\boldsymbol{K} = Conv_{3\times3}(Conv_{3\times3}(\boldsymbol{M})) , \tag{7}$$

其中 $\boldsymbol{Q} \in \mathbb{R}^{w\times h\times c}$ 和 $\boldsymbol{K} \in \mathbb{R}^{w\times h\times c}$ 分别为 $\bar{\boldsymbol{Y}}$ 和 $\boldsymbol{M}$ 的特征向量。然后，将 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 进行融合后输入 Softmax 层获得代表重构难度的权重向量 $\boldsymbol{\alpha} \in \mathbb{R}^{w\times h\times c}$：

$$\boldsymbol{\alpha} = Softmax(\boldsymbol{Q} \odot \boldsymbol{K}) 。 \tag{8}$$

代表重构难度的权重向量 $\boldsymbol{\alpha}$ 由掩码和图像纹理共同决定，通过将 $\boldsymbol{\alpha}$ 引入后续 I-ResDNet 模块中，能够实现根据重构难度引导深度网络进行重构。

2.3 I-ResDNet 模块

相关研究表明，对特征不同通道进行划分，并利用密集连接融合处理后的各部分特征，能够有效提取深层特征[22]。借鉴这个思想，文献 [22] 提出 ResDNet 模块。如图 5 所示，ResDNet 采用单向密集连接融合划分后的特征，导致特征融合效果受限于固定通道划分方式。为了解决这个问题，一种最直接的方法是采用双向密集连接融合划分后的特征，但是这种方式会成倍增加模型参数量。为此，本文提出 I-ResDNet，创新地在 ResDNet 中引入通道打乱操作，降低特征融合效果对于通道划分方式的依赖，并引入重构难度权重向量引导特征融合，能够在不显著增加模型参数量的情况下增强特征融合能力。

I-ResDNet 模块如图 5 所示。给定输入特征 $\boldsymbol{X}^{\mathrm{r}} \in \mathbb{R}^{w\times h\times c}$，首先进行通道打乱操作，再按照特征通道

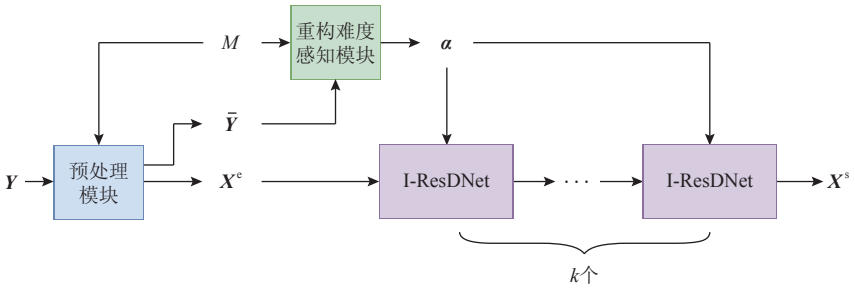


Fig. 3 Framework of RdpSCI

图 3 RdpSCI 框架

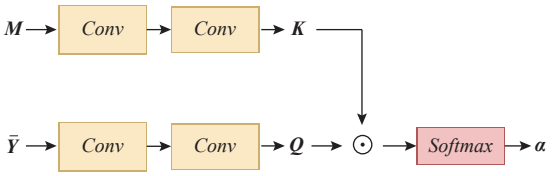


Fig. 4 Reconstruction difficulty perception module

图 4 重构难度感知模块

维度划分为 4 个部分$\{\boldsymbol{X}_i^{\mathrm{r}}\}_{i=1}^4(\boldsymbol{X}_i^{\mathrm{r}}\in\mathbb{R}^{w\times h\times(c/4)})$:

$$(\boldsymbol{X}_1^{\mathrm{r}},\boldsymbol{X}_2^{\mathrm{r}},\cdots,\boldsymbol{X}_4^{\mathrm{r}})=Split(Shuffle(\boldsymbol{X}^{\mathrm{r}}))\,, \tag{9}$$

其中$Split(\cdot)$和$Shuffle(\cdot)$分别为通道划分操作和通道打乱操作。然后, 利用 CFormer 模块[22]对各部分特征进行处理, 充分挖掘时空相关性。具体而言, 对于第i个 CFormer 模块, 首先按照通道维度将第$i-1$个 CFormer 模块的输出特征与第i个 CFormer 模块的输入特征进行串联, 输入第i个 CFormer 模块处理后, 使用1×1卷积来降低其通道维度:

$$\boldsymbol{T}_1=CFormer_1(\boldsymbol{X}_1^{\mathrm{r}})\,, \tag{10}$$

$$\boldsymbol{T}_i=CFormer_i(Conv_{1\times1}(Concat((\boldsymbol{T}_1,\boldsymbol{T}_2,\cdots,\boldsymbol{T}_{i-1},\boldsymbol{X}_i^{\mathrm{r}})))(i=2,3,4)\,, \tag{11}$$

其中$Concat(\cdot)$为通道维度的串联操作。然后利用1×1卷积来融合串联的输出特征, 最后结合残差连接和重构难度感知权重得到 I-ResDNet 模块, 最终输出$\boldsymbol{X}^{\mathrm{s}}\in\mathbb{R}^{w\times h\times c}$:

$$\boldsymbol{X}^{\mathrm{s}}=\boldsymbol{\alpha}\odot(Conv_{1\times1}(Concat((\boldsymbol{T}_1,\boldsymbol{T}_2,\cdots,\boldsymbol{T}_4)))+\boldsymbol{X}^{\mathrm{r}})\,。 \tag{12}$$

除首个 I-ResDNet 模块的输入 c 和最后一个 I-ResDNet 模块的输出 c 设置为 B, 其余 c 均设置为 64。

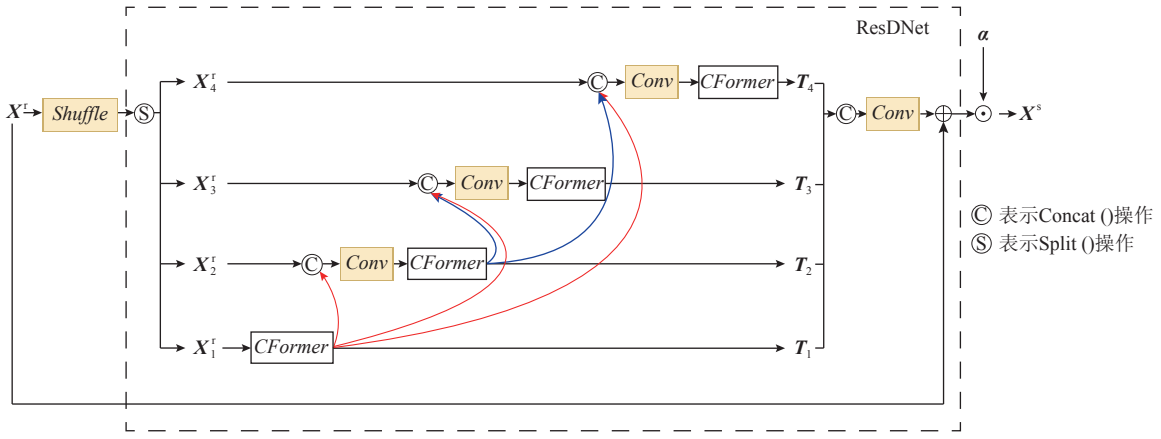


Fig. 5 I-ResDNet module

图 5 I-ResDNet 模块

3 实 验

3.1 数据集

采用 DAVIS 2017[31]数据集作为训练集; 采用基准模拟数据集[22](灰度数据集(256×256×8): Kobe, Traffic, Runner, Drop, Crash, Aerial; 彩色数据集(512×512×3×8): Beauty, Bosphorus, Jockey, Runner, ShakeNDry, Traffic)和基准真实数据集[22]: Domino 和 Hand 作为测试集。为进一步研究所提方法 RdpSCI 在高分辨率场景下的有效性, 增加彩色数据集(1 024×1 920×3×8): Hummingbird 和 Messi 作为测试集。

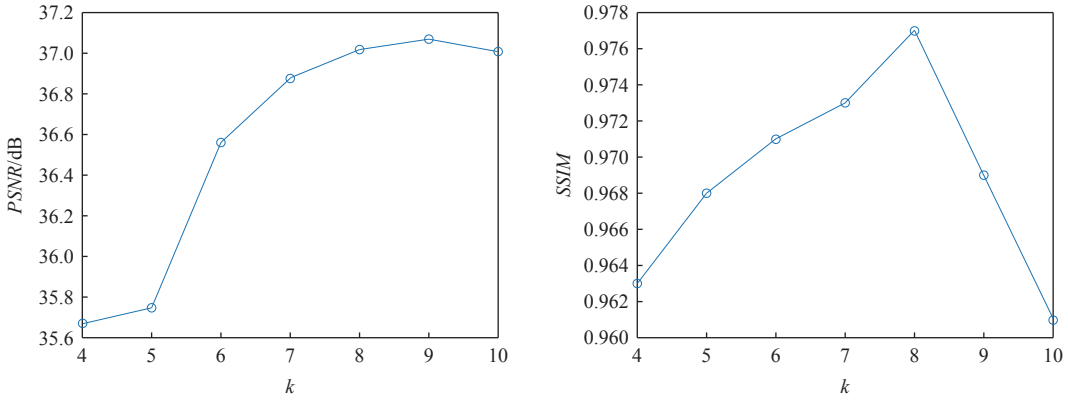
3.2 实施细节

RdpSCI 采用 PyTorch 框架, 利用一块 NVIDIA

RTX 4090 GPU 进行实验；训练集使用随机裁剪、随机缩放和随机翻转进行数据预处理；采用 Adam 优化器[32]进行训练，前 300 个 epoch 学习率设置为 0.000 1，后 100 个 epoch 学习率设置为 0.000 01；batch size 设置为 1；采用均方误差（mean-square error，*MSE*）[33]作为损失函数；采用所有视频组的平均 *PSNR* 和平均结构化相似（structural similarity，*SSIM*）[34]作为衡量重构质量的客观指标；采用模型参数量、每秒浮点运算次数（floating point operations per second，FLOPS）和平均重构时间作为衡量重构计算复杂度的客观指标。

3.3 参数选择

RdpSCI 需要研究 I-ResDNet 模块数量参数 k 对于性能的影响，如图 6 所示，我们进行了不同 k 值的对比实验。可以观察到 RdpSCI 在 $k=8,9,10$ 下的性能优于在 $k=4,5,6,7$ 下的性能，并且与在 $k=8,9,10$ 下具有相似的重构性能。从理论上分析，网络深度的增加会增加网络的容量。然而，更深的网络对学习提出了更大的挑战，这意味着更深的网络需要更长的训练时间和更多的训练技巧。因此，考虑到重构性能和训练难度之间的平衡，我们在后续实验中设 k=8。



(a) *PSNR*随*k*的变化　　(b) *SSIM*随*k*的变化

Fig. 6　Experimental comparison under different k

图 6　不同k下的实验对比

3.4 性能对比

3.4.1 基准模拟灰度数据集的实验结果

为了验证 RdpSCI 的有效性，本节将基准模拟灰度数据集作为测试集，将 RdpSCI 与视频快照压缩成像的基准方法进行对比，其中包括 BIRNAT[18]，RevSCI[35]，DUN-3DUnet[25]，ELP-Unfolding[36]，STFormer[21]，EfficientSCI[22]。

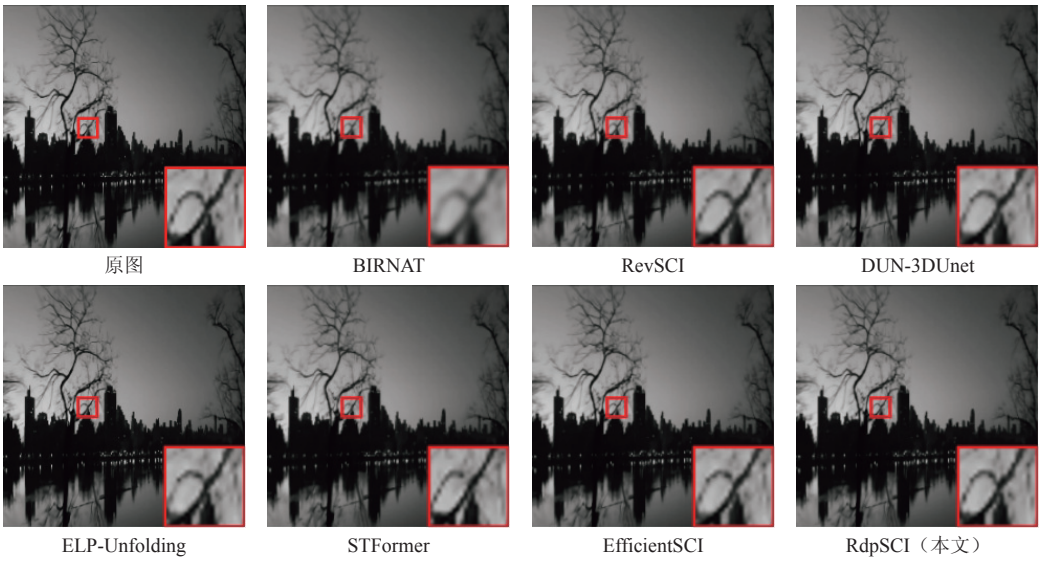
重构质量客观指标对比如表 1 所示，不难观察到，相比于 BIRNAT，RevSCI，DUN-3DUnet，ELP-Unfolding，STFormer，EfficientSCI，在 *PSNR* 方面，RdpSCI 分别有 3.71 dB，3.1 dB，1.76 dB，1.61 dB，0.68 dB，0.54 dB 的提升；在 *SSIM* 方面，RdpSCI 分别有 0.026，0.021，0.009，0.008，0.003，0.002 的提升。特别地，在重构难度最大的 Crash 序列上，相比于 BIRNAT，RevSCI，DUN-3DUnet，ELP-Unfolding，STFormer，EfficientSCI，在 *PSNR* 方面，RdpSCI 分别有 4.17 dB，3.89 dB，2.68 dB，2.36 dB，0.95 dB，0.88 dB 的提升；在 *SSIM* 方面，RdpSCI 分别有 0.048，0.038，0.019，0.016，0.005，0.004 的提升。图 7 展示了重构视觉对比，不难观察到，RdpSCI 的重构结果细节更加清晰。

Table 1　Comparison of Reconstruction Quality of Different Methods on Benchmark Simulated Grayscale Datasets

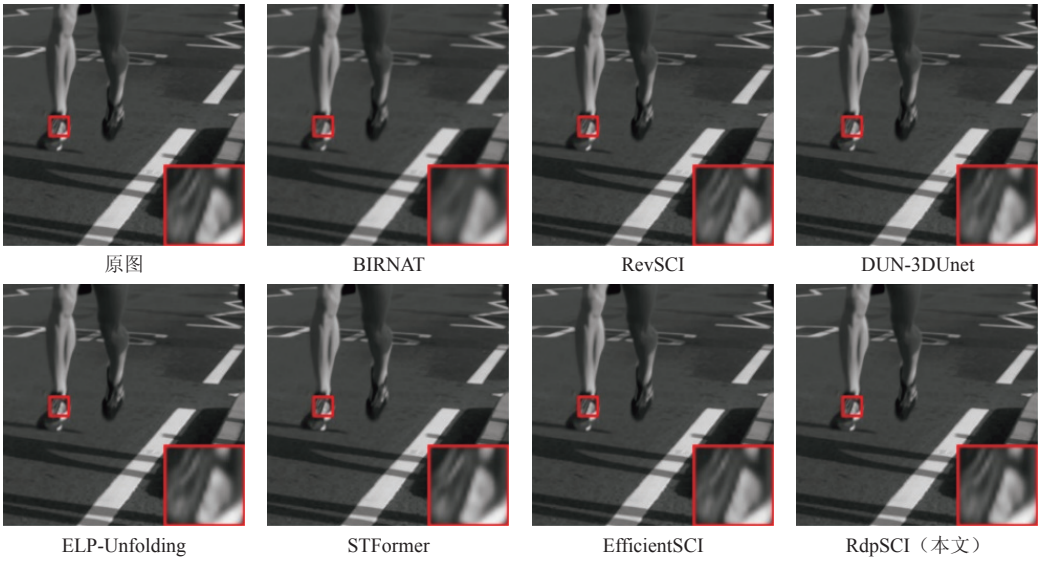
表 1　不同方法在基准模拟灰度数据集上的重构质量对比

方法	性能指标	Kobe	Traffic	Runner	Drop	Crash	Aerial	平均值
BIRNAT	*PSNR*/dB	32.710	29.330	38.700	42.280	27.840	28.990	33.310
	SSIM	0.950	0.942	0.976	0.992	0.927	0.917	0.951
RevSCI	*PSNR*/dB	33.720	30.020	39.400	42.930	28.120	29.350	33.920
	SSIM	0.957	0.949	0.977	0.992	0.937	0.924	0.956
DUN-3DUnet	*PSNR*/dB	35.000	31.760	40.030	44.960	29.330	30.460	35.260
	SSIM	0.969	0.966	0.980	0.995	0.956	0.943	0.968
ELP-Unfolding	*PSNR*/dB	34.410	31.580	41.160	44.990	29.650	30.680	35.410
	SSIM	0.966	0.962	0.986	0.995	0.959	0.944	0.969
STFormer	*PSNR*/dB	35.530	32.150	42.640	45.080	31.060	**31.560**	36.340
	SSIM	0.973	0.967	0.988	0.995	0.970	**0.953**	0.974
Efficient-SCI	*PSNR*/dB	35.760	32.300	**43.050**	**45.180**	31.130	31.500	36.480
	SSIM	0.974	0.968	0.988	0.995	0.971	0.953	0.975
RdpSCI（本文）	*PSNR*/dB	**37.830**	**33.900**	42.450	45.140	**32.010**	30.78	**37.020**
	SSIM	**0.981**	**0.980**	**0.992**	**0.996**	**0.975**	0.936	**0.977**

注：黑体数值表示最优结果。



(a) 示例1



(b) 示例2

Fig. 7 Visual comparison of reconstruction quality of different methods on benchmark simulated grayscale datasets

图 7 不同方法在基准模拟灰度数据集上的重构质量视觉对比

重构计算复杂度客观指标对比如表 2 所示，不难观察到，RdpSCI 与 BIRNAT，RevSCI，EfficientSCI 具有相近的参数量和计算速率，远小于 DUN-3DUnet，ELP-Unfolding 和 STFormer。在重构时间方面，RdpSCI 与其他基准方法具有相近的测试时间，重构时间均为毫秒级。

3.4.2 基准模拟彩色数据集的实验结果

为了验证 RdpSCI 的有效性，综合考虑实验设备显存限制与基准方法性能优越，本节选用领域现有最优方法 STFormer 和 EfficientSCI 在基准模拟彩色数据集上进行对比实验。

重构质量客观指标对比如表 3 所示，不难观察到，相比于 EfficientSCI 和 STFormer，在 *PSNR* 方面，RdpSCI

Table 2 Comparison of Computational Complexity and Reconstruction Time of Different Methods on Benchmark Simulated Grayscale Datasets

表 2 不同方法在基准模拟灰度数据集上的计算复杂度和重构时间对比

方法	参数量	计算速率/GFLOPS	重构时间/s
BIRNAT	4.13×10^6	390.56	0.10
RevSCI	5.66×10^6	766.95	0.19
DUN-3DUnet	61.91×10^6	3 975.83	0.58
ELP-Unfolding	565.73×10^6	4 634.94	0.34
STFormer	19.48×10^6	3 060.75	0.49
EfficientSCI	8.82×10^6	1 426.38	0.31
RdpSCI（本文）	8.86×10^6	1 431.82	0.31

Table 3 Comparison of Reconstruction Quality of Different Methods on Benchmark Simulated Color Datasets (Medium Resolution)

表 3 不同方法在基准模拟彩色数据集（中分辨率）上的重构质量对比

方法	性能指标	Beauty	Jockey	Runner	Bosphorus	ShakeNDry	Traffic	平均值
EfficientSCI	*PSNR*/dB	37.08	38.32	42.22	40.71	35.02	30.06	37.24
	SSIM	0.979	0.968	0.983	0.988	0.955	0.937	0.968
STFormer	*PSNR*/dB	37.37	38.32	42.45	40.39	35.15	30.24	37.32
	SSIM	0.981	0.968	0.985	0.988	0.956	0.939	0.970
RdpSCI（本文）	*PSNR*/dB	**37.54**	**38.53**	**42.69**	**40.90**	**35.19**	**30.25**	**37.52**
	SSIM	**0.981**	**0.970**	**0.985**	**0.989**	**0.957**	**0.941**	**0.971**

注：黑体数值表示最优结果。

分别有 0.28 dB 和 0.2 dB 的提升；在 *SSIM* 方面，RdpSCI 分别有 0.003 和 0.001 的提升。特别地，在重构难度最大的 Traffic 序列上，相比于 EfficientSCI 和 STFormer，在 *PSNR* 方面，RdpSCI 分别有 0.19 dB 和 0.01 dB 的提升；在 *SSIM* 方面，RdpSCI 分别有 0.004 和 0.003 的提升。图 8 展示了重构视觉对比，不难观察到，RdpSCI 的重构结果细节更加清晰。

为了进一步验证 RdpSCI 的有效性，表 4 展示了在高分辨率场景下的实验结果。可以观察到，相比于 EfficientSCI 和 STFormer，RdpSCI 在 PSNR 上分别有 1.6 dB 和 0.71 dB 的提升，提升较为明显。其原因在于随着分辨率的提高，图像纹理更加清晰，产生的重构难度权重向量相比于低分辨率场景更加精准，能够更好地引导重构。图 9 展示了重构视觉对比，不难观察到，RdpSCI 的重构结果细节更加清晰。

3.4.3 真实数据集的实验结果

为了验证 RdpSCI 的有效性，综合考虑实验设备显存限制与基准方法性能优越，本节选用领域现有

Table 4 Comparison of Reconstruction Quality of Different Methods on Benchmark Color Datasets (High Resolution)

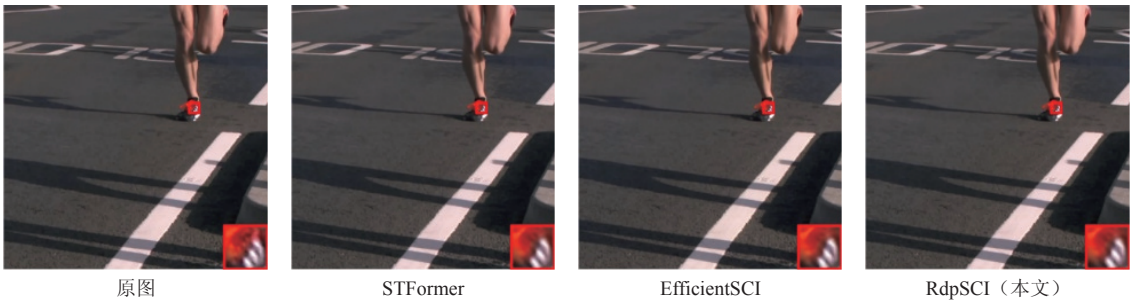
表 4 不同方法在基准彩色数据集（高分辨率）上的重构质量对比

方法	性能指标	Hummingbird	Messi	平均值
EfficientSCI	*PSNR*/dB	35.56	34.41	34.99
	SSIM	0.952	0.973	0.962
STFormer	*PSNR*/dB	38.20	33.55	35.88
	SSIM	0.965	0.964	0.965
RdpSCI（本文）	*PSNR*/dB	**38.41**	**34.77**	**36.59**
	SSIM	**0.965**	**0.981**	**0.973**

注：黑体数值表示最优结果。



(a) 示例1



(b) 示例2

Fig. 8 Visual comparison of reconstruction quality of different methods on benchmark simulated color datasets (medium resolution)

图 8 不同方法在基准模拟彩色数据集（中分辨率）上的重构质量视觉对比

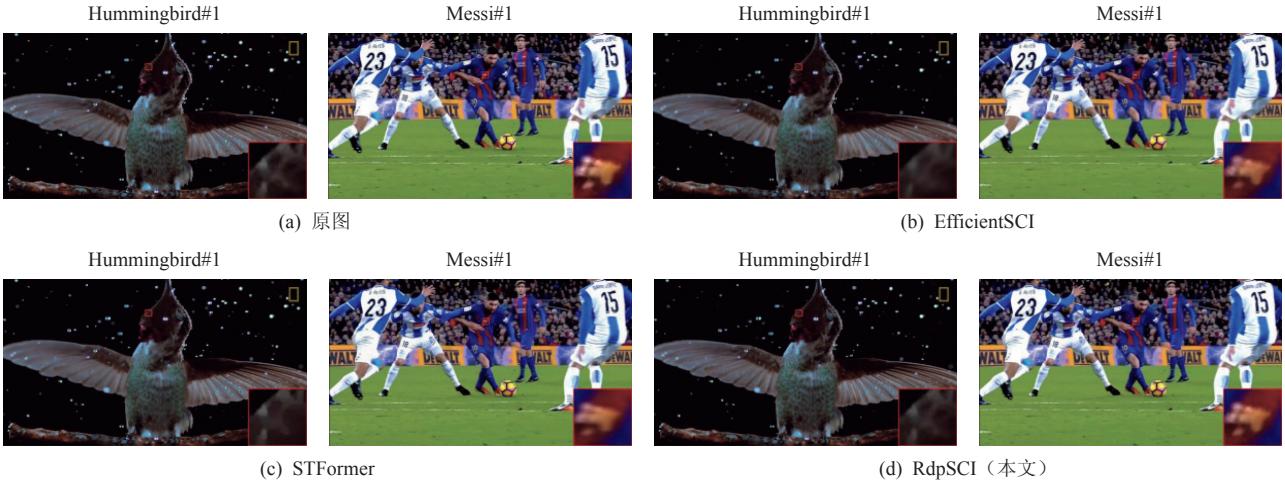


Fig. 9 Visual comparison of reconstruction quality of different methods on benchmark color datasets (high resolution)

图 9 不同方法在基准彩色数据集（高分辨率）上的重构质量视觉对比

最优方法 STFormer 和 EfficientSCI 在基准模拟真实数据集上以更高的压缩比($B=10$)进行对比实验。图 10 展示了不同方法在基准模拟真实数据集上的视觉对比，可以观察到相比于 STFormer 和 EfficientSCI，RdpSCI 即使在 $B=10$ 的情况下，依旧够更好地重构出细节和边缘。

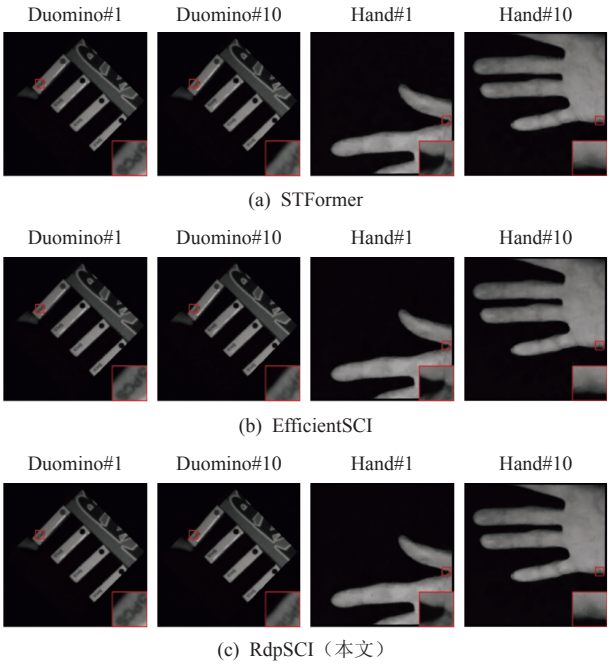


Fig. 10 Visual comparison of reconstruction quality of different methods on real datasets

图 10 不同方法在真实数据集上的重构质量视觉对比

3.5 消融实验

RdpSCI 的核心思想在于挖掘掩码信息和图像纹理与重构难度之间的关联来引导深度网络进行重构，即利用重构难度感知模块生成的权重向量引导 I-ResDNet 模块(在原始 ResDNet 模块上同时增加了通道打乱操作和重构难度加权操作)进行重构。为了验证该核心思想和 I-ResDNet 模块的有效性，在基准模拟灰度数据集上，设置如图 11 所示的消融对照组。1)消融对照组 1 无重构难度感知模块，并且采用原始 ResDNet 模块进行重构，即 EfficientSCI; 2)消融对照组 2 无重构难度感知模块，并且采用 ResDNet-s 模块(仅在原始 ResDNet 模块上单独增加通道打乱操作)进行重构; 3)消融对照组 3 采用重构难度感知模块，并且采用 ResDNet-w 模块(仅在原始 ResDNet 模块上单独增加重构难度加权操作)进行重构; 4)消融对照组 4 采用重构难度感知模块(无掩码分支提取重构难度权重向量)，并采用 I-ResDNet 模块进行重构。

如表 5 所示，相比于消融对照组 1，消融对照组 2 的重构质量有 0.15 dB 的提升，验证了 I-ResDNet 模块的有效性，即通过在 ResDNet 中引入通道打乱操作能够降低特征融合效果对于通道划分方式的依赖，增强特征融合能力; 相比于消融对照组 1，消融对照组 3 的重构质量有 0.25 dB 的提升，验证了 RdpSCI 的核心思想的有效性，即通过在 ResDNet 中引入重构难度加权操作能够引导深度网络进行更好的特征提取。RdpSCI 相比于消融对照组 1、消融对照组 2 和消融对照组 3 的重构质量分别有 0.54 dB、0.39 dB 和 0.29 dB 的提升，验证了通过利用重构难度感知模块生成的模块引导 I-ResDNet 模块进行重构的有效性; 相比于消融对照组 4, RdpSCI 有 0.12 dB 的提升，验证了同时从 $\bar{\boldsymbol{Y}}$ 和 $\boldsymbol{M}$ 中提取重构难度权重向量的有效性。其原因在 $\bar{\boldsymbol{Y}}$ 仅包含了掩码的整体信息 $\sum_{i=1}^{B}\boldsymbol{M}_i$，但是缺乏具

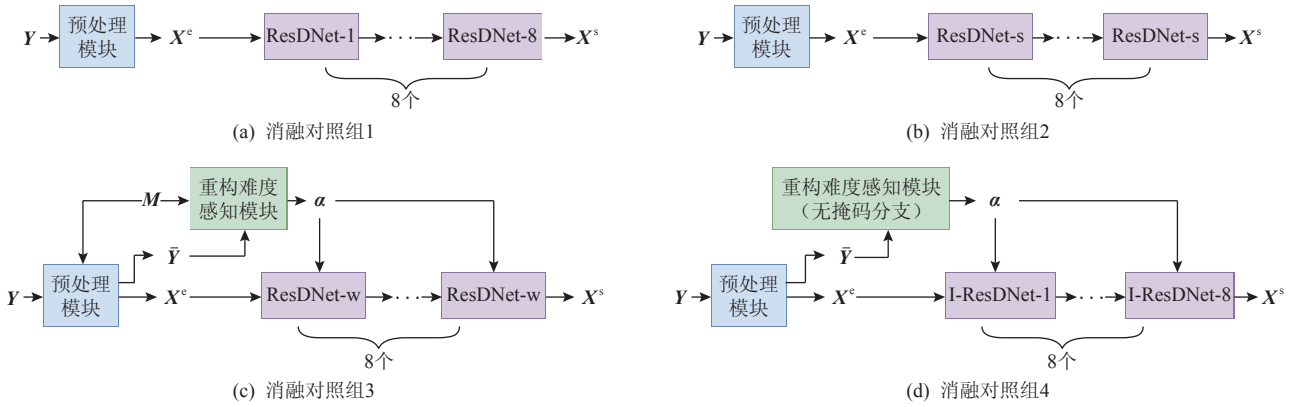


Fig. 11 Frameworks of ablation control groups

图 11 消融对照组框架

体单帧掩码$\boldsymbol{M}_i$的信息。由于每帧的重构难度与图像纹理和对应掩码共同决定，若只利用$\bar{\boldsymbol{Y}}$提取重构难度权重向量，无法反映具体单帧的重构难度。为了更好地阐述重构难度感知模块的机理，图 12 展示了重构难度权重的可视化对比，可以观察到一方面权重向量$\boldsymbol{\alpha}$代表的重构难度与图像纹理和掩码高度相关；另一方面不同帧对应的重构难度不同（例如框内区域），通过引入具体单帧掩码，能够有效地对不同帧的重构难度进行感知。值得注意的是，重构难度感知模块在 Aerial 序列上的效果并不十分理想，其原因在于 Aerial 序列图像纹理最为复杂，但是分辨率较低，重构难度感知模块的输入$\bar{\boldsymbol{Y}}$无法准确反映图像纹理。

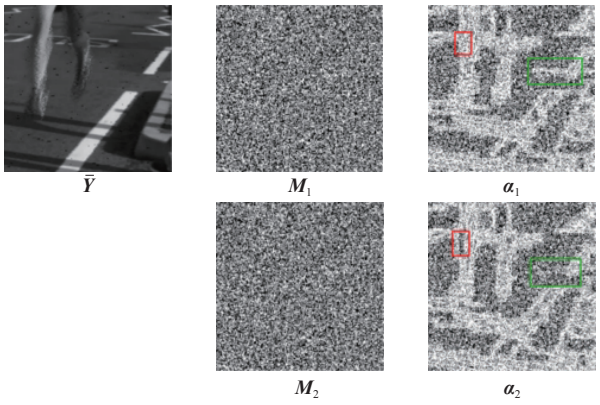


Fig. 12 Visual comparison of reconstruction difficulty weight vectors (Runner)

图 12 重构难度权重向量可视化对比（Runner）

Table 5 Comparison of Ablation Experiments

表 5 消融实验对比

不同方法	性能指标	Kobe	Traffic	Runner	Drop	Crash	Aerial	平均值	时间/s
消融对照组 1	*PSNR*/dB	35.760	32.300	**43.050**	45.180	31.130	31.500	36.480	0.310
	SSIM	0.974	0.968	0.988	0.995	0.971	0.953	0.975	
消融对照组 2	*PSNR*/dB	36.040	32.410	42.190	**45.220**	31.29	**31.630**	36.63	0.310
	SSIM	0.977	0.969	0.989	0.995	**0.978**	**0.961**	**0.978**	
消融对照组 3	*PSNR*/dB	37.210	33.100	42.370	45.050	31.970	30.680	36.730	0.310
	SSIM	0.979	0.977	0.989	0.995	0.975	0.935	0.975	
消融对照组 4	*PSNR*/dB	37.710	33.810	42.330	44.870	31.990	30.700	36.900	0.310
	SSIM	0.981	0.979	0.991	0.995	0.976	0.935	0.976	
RdpSCI（本文）	*PSNR*/dB	**37.830**	**33.900**	42.450	45.140	**32.010**	30.780	**37.020**	0.310
	SSIM	**0.981**	**0.980**	**0.992**	**0.996**	0.975	0.936	0.977	

注：黑体数值表示最优结果。

4 结 论

为了满足高速视频成像的需求，聚焦基于深度学习的视频快照压缩成像重构方法，本文针对现有方法忽略了掩码和图像纹理对重构难度影响的问题，提出一种基于重构难度感知的视频快照压缩成像重构方法 RdpSCI，提出挖掘掩码信息和图像纹理与重

构难度之间的关联来引导深度网络进行重构,并通过大量实验验证了 RdpSCI 的有效性。未来的研究重点将聚焦于如何在不损失重构质量的前提下轻量化网络框架以及如何优化图像纹理,进一步提高重构效率。

作者贡献声明:陈灿主要负责领域调研、方法设计和论文撰写;石鹏飞主要负责代码编写和数据收集;周超、胡俊和张登银主要负责校稿。

参 考 文 献

[1] Yuan Xin, Brady D J, Suo Jinli, et al. Introduction to the special issue on deep learning for high-dimensional sensing[J]. IEEE Journal of Selected Topics in Signal Processing, 2022, 16(4): 603−607

[2] Xue Wanli, Zhang Zhibin, Pei Shenglei, et al. Mixing tokens from target and search regions for visual object tracking[J]. Journal of Computer Research and Development, 2024, 61(2): 460−469 (in Chinese)
(薛万利, 张智彬, 裴生雷, 等. 混合目标与搜索区域令牌的视觉目标跟踪[J]. 计算机研究与发展, 2024, 61(2): 460−469)

[3] Song Xiaogang, Hu Haoyue, Ning Jingyu, et al. Self-supervised monocular depth estimation method for joint semantic segmentation[J]. Journal of Computer Research and Development, 2024, 61(5): 1336−1347 (in Chinese)
(宋霄罡, 胡浩越, 宁靖宇, 等. 联合语义分割的自监督单目深度估计方法[J]. 计算机研究与发展, 2024, 61(5): 1336−1347)

[4] Yuan Xin, Brady D J, Katsaggelos A K. Snapshot compressive imaging: Theory, algorithms, and applications[J]. IEEE Signal Processing Magazine, 2021, 38(2): 65−88

[5] Suo Jinli, Zhang Weihang, Gong Jin, et al. Computational imaging and artificial intelligence: The next revolution of mobile vision[J]. Proceedings of the IEEE, 2023, 111(12): 1607−1639

[6] Tao Chenning. Research on spectral imaging systems and reconstruction algorithms based on compressive sensing[D]. Hangzhou: Zhejiang University, 2021 (in Chinese)
(陶陈凝. 基于压缩感知的光谱成像系统和重构算法研究[D]. 杭州: 浙江大学, 2021)

[7] Meng Ziyi. System design and deep learning-based reconstruction for snapshot compressive imaging[D]. Beijing: Beijing University of Posts and Telecommunications, 2021 (in Chinese)
(孟子艺. 快照压缩成像系统与深度学习重构方法研究[D]. 北京: 北京邮电大学, 2021)

[8] Yuan Xin, Llull P, Liao Xuejun, et al. Low-cost compressive sensing for color video and depth[C]//Proc of the Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2014: 3318−3325

[9] Yuan Xin. Generalized alternating projection based total variation minimization for compressive sensing[C]//Proc of the 2016 IEEE Int Conf on Image Processing (ICIP). Piscataway, NJ: IEEE, 2016: 2539−2543

[10] Liu Yang, Yuan Xin, Suo Jinli, et al. Rank minimization for snapshot compressive imaging[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 41(12): 2990−3006

[11] Lucas A, Iliadis M, Molina R, et al. Using deep neural networks for inverse problems in imaging: Beyond analytical methods[J]. IEEE Signal Processing Magazine, 2018, 35(1): 20−36

[12] Tang Ruidong, Yang Chunling, Xuan Yunyi. Local enhancement reconstruction algorithm based on multi-hypothesis prediction in compressed video sensing[J]. Acta Automatica Sinica, 2022, 48(8): 1984−1993 (in Chinese)
(汤瑞东, 杨春玲, 禤韵怡. 视频压缩感知多假设局部增强重构算法[J]. 自动化学报, 2022, 48(8): 1984−1993)

[13] Yu Lei, Liao Wei, Zhou Youlong, et al. Event camera based synthetic aperture imaging[J]. Acta Automatica Sinica, 2022, 48(8): 1984−1993 (in Chinese)
(余磊, 廖伟, 周游龙, 等. 基于事件相机的合成孔径成像[J]. 自动化学报, 2023, 49(7): 1393−1406)

[14] Yang Xin, Yang Chunling. MAP-based multi-information flow gradient update and aggregation for video compressed sensing reconstruction[J]. Acta Electronica Sinica, 2023, 51(11): 3320−3330 (in Chinese)
(杨鑫, 杨春玲. 基于 MAP 的多信息流梯度更新与聚合视频压缩感知重构算法[J]. 电子学报, 2023, 51(11): 3320−3330)

[15] Wei Zhichao, Yang Chunling. Video compressed sensing reconstruction network based on temporal-attention feature alignment[J]. Acta Electronica Sinica, 2022, 50(11): 2584−2592 (in Chinese)
(魏志超, 杨春玲. 时域注意力特征对齐的视频压缩感知重构网络[J]. 电子学报, 2022, 50(11): 2584−2592)

[16] Iliadis M, Spinoulas L, Katsaggelos A K. Deep fully-connected networks for video compressive sensing[J]. Digital Signal Processing, 2018, 72: 9−18

[17] Qiao Mu, Meng Ziyi, Ma Jiawei, et al. Deep learning for video compressive sensing[J]. Apl Photonics, 2020, 5(3): 030801

[18] Cheng Ziheng, Lu Ruiying, Wang Zhengjue, et al. BIRNAT: Bidirectional recurrent neural networks with adversarial training for video snapshot compressive imaging[C]//Proc of the European Conf on Computer Vision. Berlin: Springer, 2020: 258−275

[19] Cai Yuanhao, Lin Jing, Hu Xiaowan, et al. Coarse-to-fine sparse transformer for hyperspectral image reconstruction[C]//Proc of the European Conf on Computer Vision. Berlin: Springer, 2022: 686−704

[20] Zheng Siming, Yuan Xin. Unfolding framework with prior of convolution-transformer mixture and uncertainty estimation for video snapshot compressive imaging[C]//Proc of the 2023 IEEE/CVF Int Conf on Computer Vision (ICCV). Piscataway, NJ: IEEE, 2023: 12692−12703

[21] Wang Lishun, Cao Miao, Zhong Yong, et al. Spatial-temporal transformer for video snapshot compressive imaging[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(7): 9072−9089

[22] Wang Lishun, Cao Miao, Yuan Xin. EfficientSCI: Densely connected

network with space-time factorization for large-scale video snapshot compressive imaging[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2023: 18477−18486

[23] Meng Ziyi, Yuan Xin, Jalali S. Deep unfolding for snapshot compressive imaging[J]. International Journal of Computer Vision, 2023, 131(11): 2933−2958

[24] Meng Ziyi, Jalali S, Yuan Xin. Gap-net for snapshot compressive imaging[J]. arXiv preprint, arXiv: 2012.08364, 2020

[25] Wu Zhuoyuan, Zhang Jian, Mou Chong. Dense deep unfolding network with 3D-CNN prior for snapshot compressive imaging[J]. arXiv preprint, arXiv: 2109.06548, 2021

[26] Yuan Xin, Liu Yang, Suo Jinli, et al. Plug-and-play algorithms for large-scale snapshot compressive imaging[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 1447−1457

[27] Zheng Siming, Liu Yang, Meng Ziyi, et al. Deep plug-and-play priors for spectral snapshot compressive imaging[J]. Photonics Research, 2021, 9(2): B18−B29

[28] Wu Zongliang, Yang Chengshuai, Su Xiongfei, et al. Adaptive deep pnp algorithm for video snapshot compressive imaging[J]. International Journal of Computer Vision, 2023, 131(7): 1662−1679

[29] Chen Can, Zhou Chao, Zhang Dengyin. Adaptive reconstruction for distributed compressive video sensing based on texture features[J]. Chinese Journal of Sensors and Actuators, 2024, 37(1): 58−63 (in Chinese)
(陈灿, 周超, 张登银. 基于纹理特征的分布式视频压缩感知自适应重构方法[J]. 传感技术报, 2024, 37(1): 58−63)

[30] Chen Can, Liu Pengyuan, Zhou Yuhang, et al. Adaptive reconstruction for block-based image compressed sensing via a texture-gated network[C]//Proc of the 42nd Chinese Control Conf (CCC). Piscataway, NJ: IEEE, 2023: 7542−7546

[31] Pont-Tuset J, Perazzi F, Caelles S, et al. The 2017 DAVIS Challenge on video object segmentation[J]. arXiv preprint, arXiv: 1704.00675, 2017

[32] Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. arXiv preprint, arXiv: 1412.6980, 2014

[33] Willmott C J, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance[J]. Climate Research, 2005, 30(1): 79−82

[34] Setiadi D R I M. PSNR vs SSIM: Imperceptibility quality assessment for image steganography[J]. Multimedia Tools and Applications, 2021, 80(6): 8423−8444

[35] Cheng Ziheng, Chen Bo, Liu Guanliang, et al. Memory-efficient network for large-scale video compressive sensing[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 16246−16255

[36] Yang Chengshuai, Zhang Shiyu, Yuan Xin. Ensemble learning priors driven deep unfolding for scalable video snapshot compressive imaging[C]//Proc of the European Conf on Computer Vision. Berlin: Springer, 2022: 600−618



Chen Can, born in 1993. PhD, lecturer. His main research interests include computational imaging and video reconstruction.
陈　灿, 1993 年生。博士, 讲师。主要研究方向为计算成像、视频重构。



Shi Pengfei, born in 2000. Master candidate. His main research interests include computer vision and video reconstruction.
石鹏飞, 2000 年生。硕士研究生。主要研究方向为计算机视觉、视频重构。



Zhou Chao, born in 1994. PhD, lecturer. His main research interests include video reconstruction and video super-resolution.
周　超, 1994 年生。博士, 讲师。主要研究方向为视频重构、视频超分辨率。



Hu Jun, born in 1985. PhD, senior engineer. His main research interests include signal processing and artificial intelligence.
胡　俊, 1985 年生。博士, 高级工程师。主要研究方向为信号处理、人工智能。



Zhang Dengyin, born in 1964. PhD, professor. His main research interests include signal processing and artificial intelligence.
张登银, 1964 年生。博士, 研究员。主要研究方向为信号处理、人工智能。