

文章编号 1004-924X(2024)14-2311-14

联合图像层级特征的压缩感知迭代重构

刘玉红, 杨 恒*

(兰州交通大学 电子与信息工程学院, 甘肃 兰州 730070)

摘要: 基于卷积神经网络 (Convolutional Neural Networks, CNN) 的图像压缩感知重构算法难以捕捉高分辨率图像的长距离依赖关系, 采用 Transformer 虽能解决该问题, 但网络参数量和图像重构时间成倍增长。基于此, 本文提出了一种联合图像层级特征的压缩感知迭代重构网络 (Combining Image Hierarchical-Feature Network, CHFNet), 在提高图像重构质量的同时减少重构时间。CHFNet 由采样和重构两个子网络组成, 采样子网络通过可学习的采样矩阵为重构过程提供更有效的测量值。在重构子网络中, 设计了一种使用梯度下降操作和特征优化操作的迭代策略, 同时提出一种轻量级 CNN-Transformer 混合架构, 能够建模并优化高细粒度的图像层级特征, 在增强网络感知能力的同时降低计算复杂度。此外, CHFNet 通过联合优化学习采样重构, 实现了完整的端到端训练。实验结果表明, 所提算法在多个公共基准数据集上取得了良好的重构效果。在 Urban100 数据集上, 相较于现有最优算法 CSformer, 平均 PSNR, SSIM 分别提升 0.63 dB 和 0.007 6; 在 0.10 采样率下, 相较 CSformer 在 Set11, BSD68 和 Urban100 数据集上的平均重构时间分别减少了 2.744 7 s, 3.551 0 s 和 4.775 0 s。

关键词: 压缩感知; 图像层级特征; Transformer; 卷积神经网络; 迭代策略; 图像重构

中图分类号: TP394.1; TH691.9 **文献标识码:** A **doi:** 10.37188/OPE.20243214.2311

Iterative reconstruction of compressive sensing combining image hierarchical-feature

LIU Yuhong, YANG Heng*

(School of Electronic and Information Engineering, Lanzhou Jiaotong University,
Lanzhou 730070, China)

* Corresponding author, E-mail: 11220697@stu.lzjtu.edu.cn

Abstract: The compressive sensing image reconstruction algorithms based on Convolutional Neural Networks could not capture long-range dependency of high-resolution images. Although Transformer can address this issue, it significantly increases the number of network parameters and the image reconstruction time. This paper proposed CHFNet, a combining image hierarchical-feature network for compressive sensing iterative-reconstruction to improve image reconstruction quality and reduce reconstruction time. CHFNet consisted of two sub-networks, sampling and reconstruction. The sampling sub-network utilized a learnable sampling matrix to provide more effective measurements for reconstruction phase. In the reconstruction sub-network, we introduced an iterative strategy using gradient descent and feature optimization operations, and proposed a lightweight CNN-Transformer hybrid architecture to model and optimize ex-

收稿日期: 2024-03-01; 修订日期: 2024-04-19.

基金项目: 国家自然科学基金资助项目 (No. 62161016, No. 61661025)

tremely fine-grained image hierarchical-feature, enhancing network's sensing-capability and reducing computation complexity. Moreover, CHFNet achieved complete end-to-end training by jointly optimizing sampling-reconstruction process. The experimental results show that the proposed algorithm obtains satisfactory recovery performance on several public benchmark datasets. On the Urban100 dataset, the method of this paper improves the average PSNR and SSIM metrics by 0.63 dB and 0.007 6 respectively compared to the existing optimal algorithm CSformer. At 0.10 sampling rate, the average reconstruction time of CHFNet decreases 2.744 7 s, 3.551 0 s, and 4.775 0 s compared to CSformer on Set11, BSD68, and Urban100 datasets respectively.

Key words: compressive sensing; image hierarchical-feature; Transformer; convolutional neural networks; iterative strategy; image reconstruction

1 引言

压缩感知(Compressive Sensing, CS)理论^[1]表明,当信号具备稀疏性时,能够以远低于 Nyquist 定理的采样率对其进行采样,并高概率地恢复出原始信号。其核心优势在于对信号同时进行采样和压缩,从而减少采样端的资源浪费,广泛应用于单像素相机^[2]、核磁共振成像^[3]、快照压缩成像^[4]和全息成像^[5]等领域。针对具有更多信息的图像压缩感知(Image Compressive Sensing, ICS),重构算法是研究的重点,其性能优劣直接决定了恢复图像的质量。传统 ICS 重构算法采用迭代优化算法来恢复原始图像,但这种方法严重依赖人工设计的参数,图像重构质量较差。得益于卷积神经网络(Convolutional Neural Networks, CNN)和 Transformer 强大的学习能力,部分学者将其引入 ICS 领域来提升图像重构质量。然而,在处理高分辨率图像时,基于 CNN 的 ICS 重构算法难以捕捉长距离依赖关系;而使用 Transformer 的算法虽能建模该关系,但网络参数量和图像重构时间成倍增长。因此,如何对高分辨率图像进行高质量重构并减少重构时间,仍是一个亟待解决的问题。

针对上述问题,本文提出一种联合图像层级特征的压缩感知迭代重构网络(Combining Image Hierarchical-Feature Network, CHFNet)。CHFNet 包含一个采样子网络和一个重构子网络,二者联合优化学习。采样子网络通过自适应学习采样矩阵保留更多的非冗余信息,为重构过程提供更有效的测量值;重构子网络则由线性的初始重构子网络和增强的混合重构子网络级联

而成,其中混合重构子网络采用迭代策略来逐步提升图像重构质量,每次迭代均由梯度下降操作和特征优化操作组成,前者更新前一阶段重构图像的多通道特征图,获取细粒度更高的图像特征;特征优化操作则设计了一种轻量级 CNN-Transformer 混合架构,在建模图像层级特征的同时进行深度优化,从而增强网络对高分辨率图像的感知能力,并降低算法的计算复杂度。

本文的主要贡献如下:

(1)提出一种联合图像层级特征的压缩感知迭代重构网络(CHFNet),通过迭代使用梯度下降操作和特征优化操作,实现高质量的图像重构;

(2)设计了一种轻量级 CNN-Transformer 混合架构,能够建模图像层级特征并对其进行深度优化,在增强网络感知能力的同时降低计算复杂度;

(3)CHFNet 通过联合优化学习采样重构,建立了端到端的训练模型。在多个基准数据集上的实验显示,本文算法相较于前沿的 ICS 重构算法取得了良好的重构效果,证明了算法有效性。

2 相关研究

2.1 CS 理论及迭代优化算法

传统的 CS 重构算法^[6]通过迭代优化的方式求解稀疏正则化问题,从而得到重构信号。即给定一个原始信号 $x \in R^{N \times 1}$,通过采样矩阵 $\Phi \in R^{M \times N}$ 对信号进行线性采样得到测量值 $y \in R^{M \times 1} (M \ll N)$,表示为:

$$y = \Phi x. \quad (1)$$

根据CS理论,如果信号 x 本身是稀疏的,则可以直接通过测量值 y 来恢复原始信号。然而,由于测量值的维度远小于原始信号的维度,因此信号恢复成为一个欠定的逆问题。为求解该问题,通常需要引入信号先验信息作为约束,并将其转化为一个优化问题:

$$x = \arg \min_x \frac{1}{2} \|\Phi x - y\|_2^2 + \xi \cdot R(x), \quad (2)$$

其中: $\|\Phi x - y\|_2^2$ 表示数据保真项, ξ 表示正则化参数, $R(x)$ 表示正则化项,通常为稀疏约束。

为解决式(2)的优化问题,学者们提出了多种迭代优化算法。其中,交替方向乘子法(The Alternating Direction Method of Multipliers, ADMM)^[7]将一个大的优化问题分解成若干小的子问题,通过交替迭代的方式分别求解每个子问题,并在每次迭代中引入一个乘子来更新解的估计值,从而得到原始问题的解。近似消息传递法(Approximate Message Passing, AMP)^[8]基于概率图模型,在迭代过程中集成多种去噪,确保每次迭代时的信号扰动接近所要去除的高斯白噪声,从而恢复出原始信号。此外,迭代收缩阈值算法(Iterative Shrinkage-thresholding Algorithm, ISTA)^[9]将原问题转化为一阶近似问题,通过梯度下降操作和软阈值操作来不断更新解向量的估计值,从而逐渐逼近原问题的最优解,以第 k 次迭代为例,ISTA算法的操作步骤表示为:

$$r^{(k)} = x^{(k-1)} - \lambda \Phi^T (\Phi x^{(k-1)} - y), \quad (3)$$

$$x^{(k)} = F_{\text{sgn}}(r^{(k)}) \left(|r^{(k)}| - T \right), \quad (4)$$

其中:式(3)表示梯度下降操作, $x^{(k-1)}$ 为第 k 次迭代的输入解向量,也是第 $k-1$ 次迭代的输出解向量, λ 为迭代步长, Φ^T 为采样矩阵的转置;式(4)表示软阈值操作, $x^{(k)}$ 为第 k 次迭代输出解向量, $F_{\text{sgn}}(\cdot)$ 为符号函数, T 为收缩阈值。

2.2 基于CNN的ICS重构算法

基于CNN的ICS重构算法主要分为两类:深度前馈算法和深度展开算法。前者不受任何人工设计参数限制,通过自适应学习进行图像重构。ReconNet^[10]首次将CNN与ICS结合,采用固定的随机高斯矩阵对图像块进行线性采样,并引入BM3D去噪器对重构图像进行优化。Shi等人^[11]提出CSNet系列网络,采用全图重构策略抑制分块采样带来的块效应,并约束采样矩阵以提

高信号传输效率。田金鹏等人^[12]提出的DU-SANet通过自注意力机制学习图像的高频纹理信息,并采用残差连接有效解决网络深度增加所带来的梯度消失问题。这些算法直接使用CNN将测量值映射为图像,可解释性较差。针对该问题,学者们提出深度展开算法,将传统的迭代优化算法映射到深度网络中,从而提高图像重构质量并改善算法的可解释性。Zhang等人^[13]展开AMP算法的迭代去噪过程提出AMPNet,利用采样矩阵和CNN来拟合图像差异,从而表征图像先验信息。在重构过程中,该网络还集成了一个去伪影模块,旨在消除重构图像的块效应。ISTANet+^[14]通过深度网络实现了端到端的ISTA算法,有效解决了与非线性稀疏变换相关的邻近映射问题。在此基础上,ISTANet++^[15]为ICS重构网络引入更多先验知识,通过动态展开策略来处理多采样率下的重构任务,同时采用跨块策略减少块效应并提升图像重构质量。OPINENet+^[16]引入采样矩阵正交约束和二元约束以规范优化问题的正则化项,并通过自适应采样和联合重构优化图像块。陈文俊等人^[17]提出FSOINet,利用近端梯度下降法将图像映射到特征域,并建立迭代优化的信息流动通道来实现图像重构。上述算法相较传统迭代优化算法显著提高了图像重构质量,但仍受CNN架构提取图像局部信息的约束,难以捕提高分辨率图像的长距离依赖关系。

2.3 使用Transformer的ICS重构算法

Transformer^[18]最初是为自然语言处理(Natural Language Processing, NLP)设计的,用于处理序列数据的长距离依赖关系,其内部的自注意力被认为是一种适用于图像处理的非局部滤波^[19]操作。受NLP中Transformer强大性能的启发,ViT^[20]开创性地将该架构扩展到图像领域。Swin Transformer^[21]则设计了一种具有移位机制的多头自注意力,有效降低了计算成本。近年来,Transformer也被应用于ICS领域。TransCS^[22]在重构模块采用迭代的Transformer主干来建模图像的长距离依赖关系,同时引入辅助CNN捕捉图像的局部特征。但它在深度重构网络中同时使用Transformer编码器和解码器,导致网络参数数量和图像重构时间显著增加。CS-

former^[23]则采用多路径前馈方法,将测量值映射到初始化主干、CNN主干和Transformer主干,旨在探索图像的局部特征和全局表示,并通过渐进策略同时计算图像细粒度和长距离依赖关系,有效提升了重构算法性能,但复杂的网络结构也进一步增加了图像重构时间。

2.4 锚点自注意力

自注意力(Self-Attention)对建模图像长距离依赖关系有着良好的可扩展性,通过对图像token进行相似性比较,使得每个像素点的特征向量均包含着其上下文像素的特征信息。该操作通过线性映射将token转换为空间域的查询矩阵 Q 、键矩阵 K 和值矩阵 V ,并通过式(5)来计算所有token的加权和:

$$Y_{SA} = \text{Softmax}\left(\frac{Q \cdot K^T}{\sqrt{d}}\right) \cdot V, \quad (5)$$

其中: $Q = W_Q \cdot I, K = W_K \cdot I, V = W_V \cdot I; I \in \mathbb{R}^{d \times N_u}$ 为自注意力的输入token, $W_Q, W_K, W_V \in \mathbb{R}^{d \times d}$ 为转换矩阵; N_u 为token的数量,也是特征图的尺寸 $H \times W$; d 为每个token的维度。此外, M 表示注意力映射,即 $M = \text{Softmax}\left(\frac{Q \cdot K^T}{\sqrt{d}}\right)$ 。

自注意力的时间复杂度和空间复杂度分别为 $O(N_u^2 d)$, $O(N_u^2)$,随图像分辨率提高呈二次方增长,很容易成为Transformer的计算瓶颈。为解决该问题,文献[24]提出锚点自注意力(Anchored Self-Attention, ASA)。如图1所示,ASA通过平均池化和线性映射对输入token进行降维处理,生成低维度的锚点矩阵 $A \in \mathbb{R}^{d \times M_u}$ ($M_u \ll N_u$)。该矩阵汇聚了特征图信息,通过充当计算相似度的中间环节来降低自注意力的时空复杂度。ASA操作的具体数学表达式为:

$$A = W_A \cdot \text{Avp}_s(I), \quad (6)$$

$$M_d = \text{Softmax}\left(\frac{K^T \cdot A}{\sqrt{d}}\right), \quad (7)$$

$$M_e = \text{Softmax}\left(\frac{A^T \cdot Q}{\sqrt{d}}\right), \quad (8)$$

$$Y_{ASA} = Z \cdot M_e = (V \cdot M_d) \cdot M_e, \quad (9)$$

其中: $\text{Avp}_s(\cdot)$ 表示步长为 S 的平均池化操作, M_d, M_e 分别表示锚点-值对和查询-锚点对之间的注意力映射。通过引入锚点矩阵,自注意力的

时间复杂度降低到 $O(N_u M_u d)$,空间复杂度降低到 $O(N_u M_u)$ 。

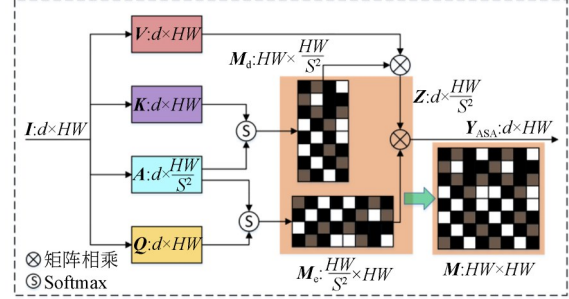


图1 锚点自注意力结构

Fig. 1 Structure of Anchored self-attention

3 本文算法

在ICS领域,利用高斯去噪、非局部相似等先验信息能够显著提升图像重构质量。因此,本文引入图像在空间维度上的先验信息,构建了一种联合图像层级特征的压缩感知迭代重构网络CHNet,整体结构如图2所示,主要包含采样子网络、初始重构子网络和混合重构子网络三个部分。采样子网络通过可学习的采样矩阵对原始图像 X 进行逐块采样得到测量值 Y 。初始重构子网络对测量值进行线性变换得到初始恢复图像 X_{init} 。受ISTA算法启发,本文在混合重构子网络设计了一种迭代策略,通过交替使用梯度下降操作和特征优化操作来逐步提升图像重构质量。在每个迭代阶段,首先构建了一个多尺度特征丰富模块(Multi-scale Feature Enrichment Module, MFEM),对前一阶段的重构图像进行不同尺度的去噪,并增强为多通道特征图。随后采用梯度下降操作对特征图的细粒度信息进行更新。其次,为了增强网络对高分辨率图像的感知能力并降低计算复杂度,构建了一个二次空间信息捕捉模块(Secondary Spatial-information Capture Module, SSCM),通过轻量级CNN-Transformer混合架构建模图像层级特征,并对其深度优化。最后,通过后处理模块(Post-processing Module, PPM)对SSCM优化的特征信息进行非线性增强,并获得本阶段的重构图像。经过 N_{opt} 次迭代后,可以得到最终的重构图像 X_{rec} 。

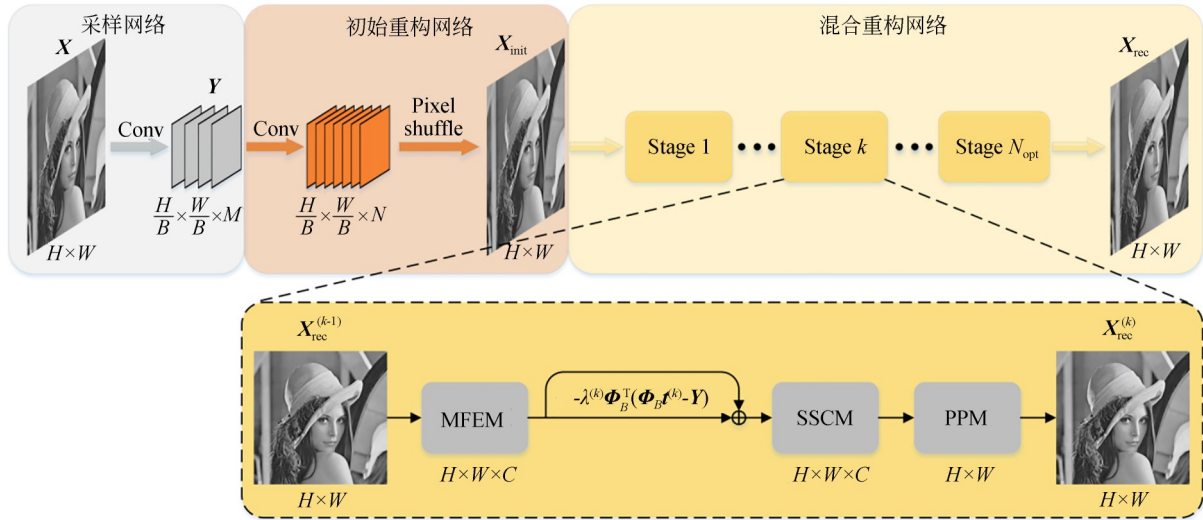


图 2 网络整体结构

Fig. 2 Structure of overall network

3.1 采样子网络

在传统的 ICS 采样中,原始图像 $X \in \mathbb{R}^{H \times W}$ 先被分割为 $B \times B$ 大小的不重叠图像块,并将其矢量化为 $x \in \mathbb{R}^{N \times 1}$ ($N = B \times B$),随后通过固定的采样矩阵 $\Phi_B \in \mathbb{R}^{M \times N}$ ($M = \lfloor \tau \times N \rfloor$) 对每个图像块矢量进行线性采样,从而得到全图的测量值。当采样率 $\tau = 0.1$ 时, $M = \lfloor 0.1 \times B^2 \rfloor$ 。与 CSNet^[11] 类似,本文通过可学习的采样矩阵对图像 X 进行线性采样。具体地,将 Φ_B 重组成 M 个大小为 $B \times B$ 、步长为 B 的不含偏置的卷积核 W_B ,并通过训练数据对矩阵参数进行优化,使测量值保留更多的非冗余信息。因此采样子网络可以表示为:

$$Y = S_B(X) = W_B * X, \quad (10)$$

其中: $S_B(\cdot)$ 表示采样操作, $*$ 表示卷积操作, Y 表示压缩感知测量值。

3.2 初始重构子网络

本文通过采样矩阵的转置 $\Phi_B^T \in \mathbb{R}^{N \times M}$ 来完成从测量值到初始恢复图像的线性变换。与采样子网络类似, Φ_B^T 被重组为 N 个 $1 \times 1 \times M$ 大小的卷积核 W_{B^T} ,随后通过“PixelShuffle”操作得到初始恢复图像 X_{init} 。这一过程表示为:

$$X_{\text{init}} = I_B(Y) = \text{PixelShuffle}(W_{B^T} * Y), \quad (11)$$

其中: $I_B(\cdot)$ 表示初始重构操作。

3.3 混合重构子网络

在梯度下降操作中, TransCS^[22] 等重构算法

对前一阶段的重构图像进行更新,但这种方法限制了网络的感知能力。因此,本文采用自适应步长对前一阶段重构图像的特征图进行梯度下降操作,使得特征优化操作能够处理细粒度更高的图像信息。具体地,本文通过 MFEM 对前一阶段的重构图像进行去噪并获取多通道特征图,随后采用梯度下降操作更新特征图的细粒度信息。以第 k 迭代阶段为例,信息流动过程可以表示为:

$$t^{(k)} = F_{\text{MFEM}}(X_{\text{rec}}^{(k-1)}), \quad (12)$$

$$r^{(k)} = t^{(k)} - \lambda^{(k)} \Phi_B^T (\Phi_B t^{(k)} - Y), \quad (13)$$

其中:式(12)的 $X_{\text{rec}}^{(k-1)}$ 表示第 $k-1$ 迭代阶段得到的重构图像, $F_{\text{MFEM}}(\cdot)$ 表示多尺度特征丰富操作, $t^{(k)} \in \mathbb{R}^{H \times W \times C}$ 表示多通道特征图;式(13)表示梯度下降操作, $\lambda^{(k)}$ 为梯度下降操作的自适应步长,本文通过两个卷积层来实现。

在完成梯度下降操作后,网络得到蕴含高细粒度信息的特征图 $r^{(k)} \in \mathbb{R}^{H \times W \times C}$ 。为进一步提升图像重构质量,需将 $r^{(k)}$ 输入至级联的 SSCM 和 PPM 进行特征优化操作。具体地,SSCM 采用轻量级 CNN-Transformer 混合架构为骨干网络,通过建模并优化图像在空间维度上的层级特征,使网络能够表达更完整的特征信息,进而增强网络对高分辨率图像的感知能力。随后,SSCM 的输出信息被送入到 PPM 进行筛选与融合,在保留重构所需关键信息的同时得到本阶段的重构图像。特征优化操作的数学表达式如式

(14)、式(15)所示:

$$s^{(k)} = F_{SSCM}(r^{(k)}), \quad (14)$$

$$X_{rec}^{(k)} = F_{PPM}(s^{(k)}), \quad (15)$$

其中: $F_{SSCM}(\cdot)$, $F_{PPM}(\cdot)$ 分别表示二次空间信息捕捉操作和后处理操作, $s^{(k)} \in R^{H \times W \times C}$ 表示 SSCM 的输出信息, $X_{rec}^{(k)} \in R^{H \times W}$ 表示本阶段重构图像。

3.3.1 多尺度特征丰富模块

卷积神经网络能够从训练数据中学习到潜在的噪声分布、感知质量等先验信息,但固定大小的卷积核通常只能提取特定尺度的信息。为使梯度下降操作能够更新具有高信噪比的多通道特征图,本文提出了一个多尺度特征丰富模块。MFEM 能够综合不同尺度的先验信息,有效去除前一阶段重构图像的噪声,并生成富含多尺度细节的图像特征。如图 3 所示, MFEM 在各尺度分支均采用双层卷积结构,第一层输出 2C 维特征,第二层则输出 C 维特征,卷积核大小分别为 1×1 , 3×3 , 5×5 和 7×7 。在得到每个尺度分支的输出特征后,通过“Concat”操作和卷积操作进行融合,在得到高信噪比多通道特征图的同时简化后续操作的计算复杂度。MFEM 的信息流动过程可以表示为:

$$F_{MFEM}(X_{rec}^{(k-1)}) = W_{fus} * \text{Concat}(F_{Mconv}(X_{rec}^{(k-1)})), \quad (16)$$

其中: $F_{Mconv}(\cdot)$ 表示多尺度卷积操作, W_{fus} 表示特征融合阶段的卷积核。

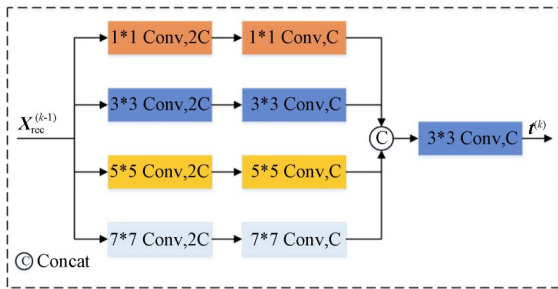


图 3 多尺度特征丰富模块(MFEM)

Fig. 3 Multi-scale feature enrichment module

3.3.2 二次空间信息捕捉模块

图像在空间维度上蕴含着丰富的局部、区域和全局特征,这些层级特征能够有效增强网络对高分辨率图像的感知能力。鉴于此,本文提出了

一个以轻量级 CNN-Transformer 混合架构为主干的二次空间信息捕捉模块。如图 4 所示, SSCM 以梯度下降操作的更新特征 $r^{(k)}$ 为输入,随后通过 N_i 层混合架构建模图像层级特征并进行深度优化,同时采用卷积操作和大型跳跃连接对优化前后的特征进行融合,使重构网络能够更完整地表征图像内容。

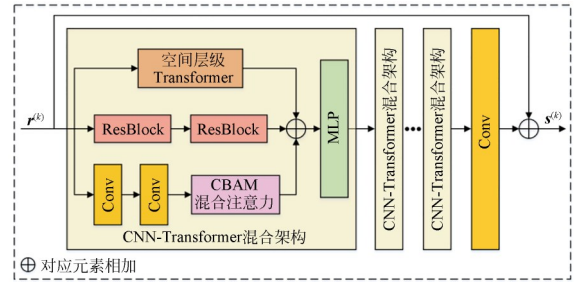


图 4 二次空间信息捕捉模块(SSCM)

Fig. 4 Secondary spatial-information capture module

本文设计的 CNN-Transformer 混合架构旨在建模图像层级特征并对其进行优化。首先,针对跨越几十个像素、以边缘为主要信息的局部特征,大部分网络通常采用具有小感受野的卷积核(例如, 3×3)进行建模。为了增强特征提取能力,局部分支级联两个标准的残差卷积块(ResBlock),通过引入跳跃连接来提升模型性能。其次,对于涉及上百个像素跨度的区域范围,可以使用具有大感受野的卷积核进行建模,但这种设计的特征表达能力并不理想,并且导致参数量较大。为解决该问题,本文在区域分支引入文献[25]的轻量级混合注意力块 CBAM,它集合了通道注意力和空间注意力的优势,能够自适应选择重要的对象和区域进行学习。为了进一步增强 CBAM 块对重要区域的特征感知能力,在执行 CBAM 块之前,采用卷积操作对输入特征进行预处理,以聚合关键信息并促进不同通道特征之间的交互。以第一层 CNN-Transformer 混合架构为例,局部分支和区域分支的特征优化过程分别表示为式(18)、式(19):

$$F_{Res}(x_R) = x_R + W_{R2} * \text{GeLU}(W_{R1} * x_R), \quad (17)$$

$$F_{loc}(r^{(k)}) = F_{Res1}(F_{Res2}(r^{(k)})), \quad (18)$$

$$F_{reg}(r^{(k)}) = F_{CBAM}(W_{r2} * \text{GeLU}(W_{r1} * r^{(k)})), \quad (19)$$

其中:式(17)表示 ResBlock 操作, x_R 为该操作的

输入信息, W_{R1} , W_{R2} 为两层卷积核, $\text{GeLU}(\cdot)$ 为非线性激活函数。式 (19) 的 $F_{\text{CBAM}}(\cdot)$ 表示 CBAM 操作, W_{r1} , W_{r2} 表示区域分支的两层预处理卷积核。

为捕捉图像的全局特征,需要设计一种能够建模长距离依赖关系的网络结构。CNN 为建模图像关系引入强大的归纳偏置,包括局部相关和权值共享等固有属性,使其能够从图像中自适应学习到可扩展的特征表示,但这些特性也限制了它在处理图像长距离依赖关系方面的能力。Transformer 的提出为图像处理带来新的突破,它通过自注意力建模图像的长距离依赖关系,从而更有效地捕捉图像全局特征。然而,随着图像分辨率的提高,自注意力的时空复杂度呈平方增长,这极大地限制了 Transformer 在处理高分辨率图像时的应用。因此,本文提出一种锚点窗口自注意力 (Anchored-Window Self-Attention, AWSA), 将图像分割为 $S_w \times S_w$ 大小的不重叠窗口,并在每个窗口内部独立执行 ASA 操作。AWSA 将每个图像窗口视作一个 token,使其数量 N_u 从 HW 降低至 S_w^2 ,进一步降低了自注意力进行特征建模的计算复杂度。

本文设计的基于 AWSA 的 Transformer 能够有效捕捉图像全局特征,但与常规 Transformer 一样,它们在建模过程中仍面临图像信息表达单调性的挑战,这限制了该架构学习能力的进一步提升。因此,本文在全局分支设计了一种空间层级 Transformer,通过跨窗口策略捕捉多元的非线性全局特征。如图 5 所示,该架构首先利用基于小窗口 AWSA 的 Transformer 对图像的全局特征进行建模,随后级联一个使用大窗口 AWSA 的 Transformer,并通过移位技术促进不同窗口间的信息交互,从而得到异于单层 Transformer 的全局特征。值得注意的是,与常规 Transformer 相比,它的时空复杂度随图像分辨率提高仍呈线性增长。因此,全局分支特征优化过程 $F_{\text{glo}}(\cdot)$ 的具体操作可定义为:

$$a^{(l)} = F_{S\text{-AWSA}}(\text{LN}(z^{(l-1)})) + z^{(l-1)}, \quad (20)$$

$$z^{(l)} = \text{MLP}(\text{LN}(a^{(l)})) + a^{(l)}, \quad (21)$$

$$a^{(l+1)} = F_{L\text{-AWSA}}(\text{LN}(z^{(l)})) + z^{(l)}, \quad (22)$$

$$z^{(l+1)} = \text{MLP}(\text{LN}(a^{(l+1)})) + a^{(l+1)}, \quad (23)$$

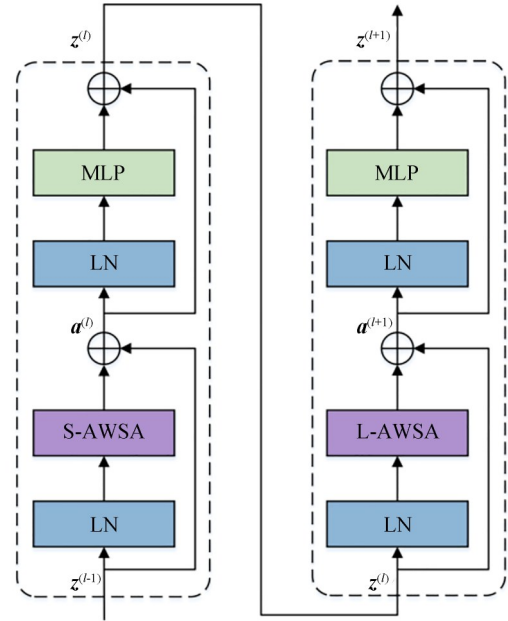


图 5 空间层级 Transformer 结构

Fig. 5 Structure of spatial-hierarchy Transformer

其中:式 (20)、式 (21) 表示基于小窗口 AWSA 的 Transformer 操作,式 (22)、式 (23) 表示基于大窗口 AWSA 的 Transformer 操作, $z^{(l)}$, $z^{(l+1)}$ 分别表示相对应的两个 Transformer 操作的输出信息, $F_{S\text{-AWSA}}(\cdot)$, $F_{L\text{-AWSA}}(\cdot)$ 分别表示采用小窗口和采用大窗口的 AWSA 操作, $z^{(l-1)}$ 为空间层级 Transformer 的输入信息,在第一层 CNN-Transformer 混合架构中即为 SSCM 的输入信息 $r^{(k)}$, $\text{MLP}(\cdot)$ 为一个标准的多层感知机, $\text{LN}(\cdot)$ 为 LayerNorm 归一化操作。

在获取全局分支的输出后,将其他两个分支的输出信息通过逐元素相加的方式进行特征融合,然后输入至标准的 MLP 进行处理,从而得到单层 CNN-Transformer 混合架构的优化输出特征。以第 k 迭代阶段的第一层混合架构为例,特征流动过程可表述为:

$$F_{\text{CT1}}^{(k)} = \text{MLP}\left(\Sigma\left(F_{\text{loc}}(r^{(k)}), F_{\text{reg}}(r^{(k)}), F_{\text{glo}}(r^{(k)})\right)\right), \quad (24)$$

其中, $\Sigma(\cdot)$ 为逐元素相加操作。

3.3.3 后处理模块

由 SSCM 深度优化的特征信息 $s^{(k)}$ 仍具有多通道特性。为保留重构所需的关键信息,并得到本阶段的重构图像 $X_{\text{rec}}^{(k)}$,本文设计了一个后处理模块。PPM 级联了一个标准 ResBlock 块和两层

卷积,对特征信息进行空间上的变换、筛选与融合,从而确保重构图像在结构、细节等方面尽可能地接近原始图像。具体表示为:

$$F_{\text{PPM}}(s^{(k)}) = W_{P2} * \text{ReLU}(W_{P1} * F_{\text{Res}}(s^{(k)})), \quad (25)$$

其中: $\text{ReLU}(\cdot)$ 为激活函数, W_{P1} , W_{P2} 分别表示级联的两个卷积层,后者为不带偏置的 1×1 卷积核。

3.4 损失函数

本文所提算法以原始图像 $X \in R^{H \times W}$ 为输入,首先通过采样子网络生成测量值 Y ,随后通过重构子网络得到最终的重构图像 X_{rec} 。为实现端到端训练,采用以下损失函数:

$$\text{Loss} = \frac{1}{HW} \|X_{\text{rec}} - X\|_2^2 + \gamma \frac{1}{M^2} \|\Phi_B \Phi_B^T - E\|_2^2, \quad (26)$$

其中:第一项为像素损失,用于衡量原始图像和重构图像之间的差异。第二项为采样矩阵约束损失,旨在规范矩阵中行与行的相关性,确保测量值存在更少的冗余, E 为单位矩阵, M 为测量值的数量, γ 为正则化参数,在实验中被设置为 0.01。

4 实验结果及分析

4.1 实验参数设置

为了得到更公平的实验结果,本文采用与文献[11]一致的训练集,包括 BSD500 数据集^[26]中训练集和测试集共 400 幅图像。在训练阶段,首先将图像随机裁剪成 96×96 像素大小的子图像,随后通过数据增强和灰度化得到 89 600 张子图像作为训练数据。为全面评估算法性能,在多个广泛使用的基准数据集上进行了测试,包括 Set11^[27], BSD68^[28] 和 Urban100^[29]。训练优化器采用 Adam^[30],前三个 epoch 作为 warm-up^[31] 阶段,学习率从 0 线性增长至 2×10^{-4} ,随后 97 个 epoch 将学习率按照余弦退火方式降至 5×10^{-5} ,批量大小 $\text{batch size} = 32$ 。在网络模型中,设置采样块大小 $B = 32$;混合重构子网络的特征通道数 $C = 16$;对于全局分支的空间层级 Transformer, AWSA 的窗口大小 $S_w \times S_w$ 分别设置为 16×16 , 32×32 ;迭代阶段数 $N_{\text{opt}} = 6$;SSCM 的 CNN-Transformer 混合架构层数 $N_l = 1$;未特殊

说明的卷积核大小均为 3×3 。

所有实验均在搭载 Intel Xeon Platinum 8375C 处理器和 RTX 4090 GPU 的 PyTorch 平台上完成。评价标准为主观的视觉效果展示以及客观的峰值信噪比(Peak Signal to Noise Ratio, PSNR)、结构相似性(Structural Similarity Index Measure, SSIM)^[32]、网络参数量和平均重构时间。

4.2 与现有算法的重构性能对比

本文选取了六种前沿的 ICS 重构算法作为对比算法,包括 CSNet+^[11], AMPNet^[13], ISTANet++^[15], OPINENet+^[16], TransCS^[22] 和 CSformer^[23]。其中,CSNet+ 属于深度前馈算法,而 AMPNet, ISTANet++ 和 OPINENet+ 属于深度展开算法,TransCS 和 CSformer 则是使用 Transformer 的 ICS 重构算法。为了更全面地比较,选取从 0.01 到 0.50 中的五个采样率进行实验,重构性能最优的用粗体标记,次优的用下划线标记。

表 1 对比了不同算法在 Set11 数据集上的图像重构性能。当采样率为 0.10 时,AMPNet 由于模拟了迭代算法的去噪过程,图像重构性能可以与使用 Transformer 的重构算法相媲美。然而,随着采样率的降低,使用 Transformer 的算法因其卓越的全局特征建模能力,表现出远超 CNN 算法的图像重构性能,在 0.01 采样率下,CSformer 在 PSNR 和 SSIM 上相较 AMPNet 分别提升 1.38 dB 和 0.049 4。本文所提算法通过高效整合 CNN 和 Transformer 的优势,对图像层级特征进行建模和优化,使得 CHFNet 在低采样率下的重构性能均优于对比算法。其中,在 0.04 和 0.10 采样率下,PSNR/SSIM 相较次优算法分别提升 0.12 dB/0.006 3, 0.29 dB/0.004 5。

为了进一步验证所提算法的泛化性能,本文在 BSD68 和 Urban100 上与其他算法进行了比较。如表 2 所示,本文算法重构图像的平均 PSNR 和 SSIM 均优于其他算法,相较 CSformer 分别平均提升 0.11 dB/0.008 5 和 0.63 dB/0.007 6。在低采样率 0.10 下,所提算法在 BSD68 数据集上,相较 CNN 算法中性能最优的 AMPNet 提升 PSNR 值为 0.50 dB;在 Urban100 数据集上,对比其他算法中性能最优的 Tran-

表 1 各采样率不同算法在 Set11 数据集上重构图像的 PSNR(dB)/SSIM 对比

Tab. 1 Comparison of PSNR(dB)/SSIM of reconstructed images on Set11 dataset with different algorithms for each sampling rate

算法	采样率					
	0.01	0.04	0.10	0.25	0.50	平均
CSNet+	20.16/0.517 4	24.59/0.760 4	28.29/0.873 9	33.16/0.942 6	38.92/0.978 9	29.02/0.814 6
AMPNet	20.20/0.558 1	25.26/0.772 2	29.40/0.877 9	34.63/0.948 1	40.33/0.980 4	29.96/0.827 3
ISTANet++	17.92/0.455 4	22.88/0.672 3	28.34/0.853 1	32.77/0.927 4	38.73/0.972 7	28.13/0.776 2
OPINENet+	20.02/0.536 2	25.52/0.787 9	29.81/0.890 4	34.81/0.951 4	40.19/0.980 0	30.07/0.829 2
TransCS	20.23/0.504 9	25.41/0.788 3	29.54/0.887 7	35.06/0.954 8	40.49/0.981 5	30.15/0.823 4
CSformer	21.58/0.607 5	26.28/0.806 1	29.69/0.888 2	34.80/0.952 4	40.72/0.982 4	30.61/0.847 3
CHFNet	21.60/0.608 2	26.40/0.812 4	30.10/0.894 9	35.33/0.956 1	40.62/0.981 8	30.81/0.850 7

表 2 各采样率不同算法在不同数据集上重构图像的 PSNR(dB)/SSIM 对比

Tab. 2 Comparison of PSNR(dB)/SSIM of reconstructed images on different datasets with different algorithms for each sampling rate

数据集	算法	采样率					
		0.01	0.04	0.10	0.25	0.50	平均
BSD68	CSNet+	21.76/0.509 0	24.89/0.676 8	27.39/0.802 5	31.03/0.908 3	35.87/0.967 6	28.19/0.772 8
	AMPNet	22.28/0.538 7	25.26/0.676 0	27.86/0.792 6	31.74/0.904 8	36.82/0.968 0	28.79/0.776 0
	ISTANet++	19.89/0.462 1	23.39/0.601 1	26.39/0.741 6	29.79/0.862 0	34.92/0.951 3	26.88/0.723 6
	OPINENet+	21.88/0.516 2	25.16/0.684 1	27.81/0.804 0	31.50/0.906 2	36.32/0.965 8	28.53/0.775 3
	TransCS	21.55/0.493 3	25.28/0.688 1	27.86/0.808 5	31.74/0.912 1	36.81/0.969 9	28.65/0.774 4
	CSformer	22.81/0.556 6	25.73/0.695 4	28.05/0.803 5	31.80/0.910 6	37.13/0.971 6	29.10/0.787 5
	CHFNet	22.94/0.566 1	25.85/0.702 1	28.36/0.827 5	31.96/0.913 9	36.93/0.970 3	29.21/0.796 0
Urban100	CSNet+	18.47/0.393 2	21.51/0.630 6	24.46/0.798 0	28.75/0.914 6	34.05/0.970 3	25.45/0.741 3
	AMPNet	18.85/0.434 5	21.93/0.641 2	25.28/0.804 9	30.39/0.925 8	36.45/0.976 9	26.58/0.756 7
	ISTANet++	16.79/0.347 2	20.04/0.543 4	24.48/0.773 2	28.83/0.894 9	35.99/0.972 7	25.23/0.706 3
	OPINENet+	18.62/0.418 9	22.07/0.666 5	25.77/0.829 7	30.96/0.935 1	36.93/0.979 7	26.87/0.766 0
	TransCS	18.32/0.379 5	22.20/0.675 6	25.87/0.834 7	31.14/0.938 7	37.23/0.981 4	26.95/0.762 0
	CSformer	19.43/0.465 3	22.62/0.682 3	25.80/0.826 7	30.49/0.932 9	36.45/0.979 3	26.96/0.777 3
	CHFNet	19.50/0.466 9	22.46/0.681 7	26.70/0.850 2	31.80/0.941 8	37.48/0.984 1	27.59/0.784 9

sCS, 重构图像的 PSNR 和 SSIM 分别提升 0.83 dB 和 0.015 5。Urban100 数据集包含高分辨率的复杂城市图像,这使得基于 CNN 的 ICS 重构算法难以实现高质量的图像重构;然而,这些图像中的规则性纹理为使用自注意力进行相似性比较的 Transformer 提供了更多的建模机会,从而实现了更为显著的重构性能提升。在 0.01 采样率下,CHFNet 相较 OPINENet+ 和 TransCS 分别提升 PSNR/SSIM 为 0.88 dB/0.048 0 和 1.18 dB/0.087 4。值得注意的是,在

0.50 采样率下,本文算法在 Urban100 数据集的 PSNR 重构性能相较 CSformer 提升 0.98 dB,仅在 Set11 和 BSD68 数据集上的重构效果稍显不足,但 PSNR 降幅均不超过 0.2 dB。一个原因是 CSformer 通过采用更大尺寸的训练图像扩展了数据规模,从而提升了算法性能。另一方面,CSformer 在重构阶段对 CNN 主干采用 128 维通道作为特征表达空间,有效挖掘了高采样率测量值所蕴含的局部特征,实现了对较低分辨率图像的高保真度重构;相比之下,本文算法聚焦于网络

结构的轻量化设计和高分辨率图像重构,在重构时仅采用 16 维通道作为局部特征表达空间,在一定程度上影响了算法对高采样率下较低分辨率图像的重构性能。

随着采样率的降低,ICS 重构算法在捕捉图像信息时面临更大的挑战,这将增加图像重构的难度。图 6 展示了不同算法在低采样率下重构图像的视觉效果,从上到下采样率分别为 0.01, 0.10 和 0.25。明显看出,在极低采样率 0.01 下,CSNet+ 和 TransCS 的重构图像中均出现了明显的块伪影;而 CSformer 和本文所提算法则更加

出色地捕捉了图像的长距离依赖关系,重构的图像质量也更为饱满。在 0.10 采样率的 Barbara 图像中,TransCS 和 CSformer 未能准确还原出围巾的纹理,细节信息严重丢失;而 CHFNet 通过深入建模图像的局部特征和区域特征,成功重构出包含丰富信息的图像。此外,在更高采样率的重构结果中,尽管 TransCS 和 CSformer 重构出的图像在整体效果上优于 CNN 算法,但在相似性细节上表现不理想;而本文算法以高细粒度的形式还原了图像的相似性结构,使得重构图像具有更少的噪声和更清晰的纹理。

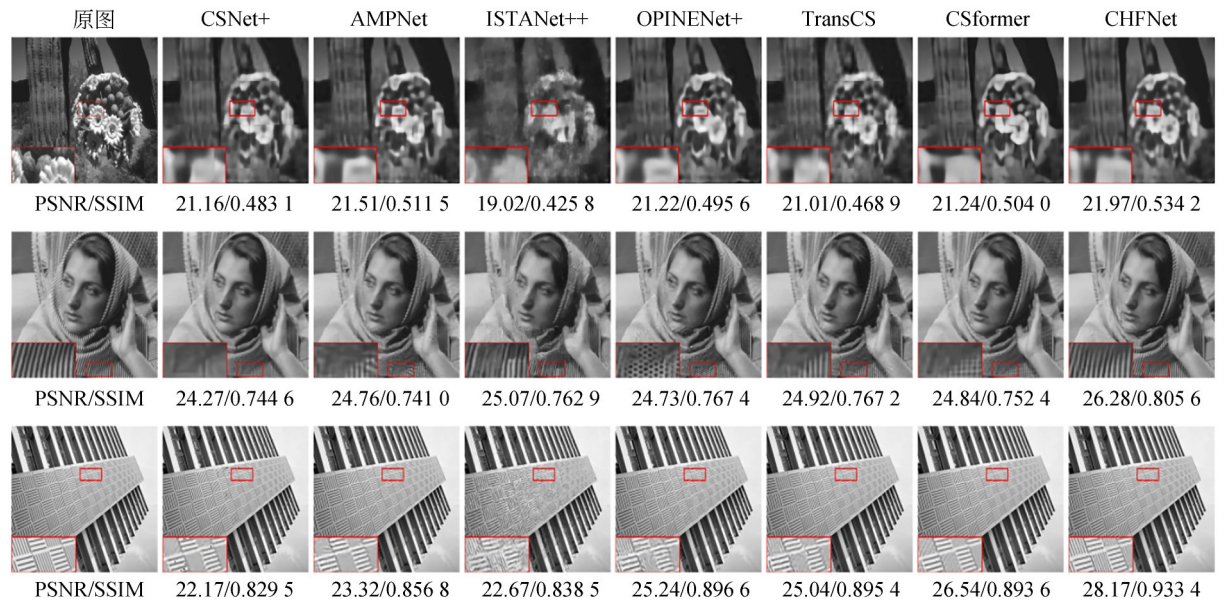


图 6 不同算法在低采样率下重构图像的视觉效果对比

Fig. 6 Comparison of visual effects of different algorithms for reconstructing images at low sampling rates

为了评估使用 Transformer 的 ICS 重构算法的计算复杂度,本文对比了 0.10 和 0.50 采样率下 TransCS,CSformer 和本文算法的网络参数量和平均重构时间,所有数据均在相同环境中测试得到。如表 3 所示,CSformer 在重构网络中使用了大量的特征通道,导致其网络参数量和平均重构时间相较于 TransCS 成倍增长。相比之下,本文所提算法通过优化 Transformer 中自注意力的时空复杂度和混合重构子网络的特征通道数,大幅减少了网络参数量。在 0.10 采样率下,CHFNet 的参数量仅为 CSformer 的 9.15%,在三个数据集上的平均重构时间分别减少了 2.744 7 s, 3.551 0 s 和 4.775 0 s。

表 3 不同算法的复杂度对比

Tab. 3 Comparison of the complexity of different algorithms

采样率	算法	网络参数量	平均重构时间(GPU)/s		
			Set11	BSD68	Urban100
0.10	TransCS	1.444 M	0.706 2	0.669 8	0.726 7
	CSformer	6.647 M	2.772 0	3.582 3	4.804 1
	CHFNet	0.608 M	0.027 3	0.031 3	0.029 1
0.50	TransCS	2.284 M	0.380 0	0.705 8	0.711 9
	CSformer	6.713 M	2.545 6	3.773 3	4.632 6
	CHFNet	1.028 M	0.027 8	0.027 8	0.030 5

4.3 消融实验与分析

4.3.1 策略消融

本节在 0.50 采样率下对每个迭代阶段的不同策略进行了消融研究。如表 4 所示,SSCM 较 MFEM 为所提算法提供了更多的性能提升。在三个测试集上,SSCM 使重构图像的 PSNR 分别提升 0.82 dB,1.37 dB 和 0.88 dB,而 MFEM 的

PSNR 提升分别为 0.28 dB,0.27 dB 和 0.17 dB。在 SSCM 中,全局分支因其强大的捕捉长距离依赖关系的能力,在高分辨率图像重构中表现出更优越的性能,相较于区域分支,为 BSD68 和 Urban100 分别带来 0.25 dB 和 0.15 dB 的 PSNR 提升。综上所述,本文所提的不同策略均为 ICS 重构带来了一定的性能提升。

表 4 0.50 采样率下不同策略的消融实验

Tab. 4 Ablation experiments with different strategies at 0.50 sampling rate

MFEM	重构策略			测试集的平均 PSNR/dB		
	SSCM			Set11	BSD68	Urban100
	全模块	区域分支	全局分支			
×	✓	—	—	40.34	36.66	37.31
✓	—	✓	×	40.44	36.49	37.24
✓	—	×	×	40.28	36.30	37.15
✓	×	—	—	39.80	35.56	36.60
✓	✓	✓	✓	40.62	36.93	37.48

4.3.2 迭代阶段数和混合架构层数

由于混合重构子网络的每个迭代阶段对应传统算法的一次迭代,并且每个阶段中 SSCM 的 CNN-Transformer 混合架构层数也是可变的,预期随着迭代阶段数 N_{opt} 和混合架构层数 N_l 的增加,图像重构性能也会随之提高。图 7 给出了 0.10 采样率下由不同的迭代阶段数和混合架构层数训练得到的 CHFNet 在 Set11 数据集上图像

重构的 PSNR。当迭代阶段数相同时,每增加一层混合架构,图像重构的 PSNR 约提升 0.15 dB;混合架构层数不变时,随着迭代阶段数的增加,网络的重构性能也随之提高,这表明更多的迭代阶段数和混合架构层数能带来更高的重构性能。但需要注意的是,更多的迭代阶段数和混合架构层数也会增加网络模型的时空复杂度。综合考虑模型性能与复杂度,选取 $N_{opt}=6, N_l=1$ 作为实验的默认参数。

4.3.3 收敛性分析

随着采样率的降低,网络采样所捕获的测量值信息显著减少,导致 ICS 重构任务在利用有限测量值恢复原始图像时面临更大的挑战,进而对算法收敛性产生不利影响。具体地,网络在训练过程中 Loss 曲线的收敛速度会随采样率降低明显放缓,且收敛值更难达到高采样率下的收敛水平。本文所提算法属于深度学习范畴,因此本节在图 8 给出了 CHFNet 在不同采样率下 Loss 在训练 epoch 的变化曲线,其中每个 epoch 在训练时迭代 2 800 次。实验结果表明,得益于更优的初始学习率参数,本文算法在训练初期展现出较快的 Loss 下降速度。随着训练 epoch 数的增加,Loss 曲线逐渐趋于稳定,表现出良好的收敛性。值得注意的是,Loss 曲线仅在 0.50 采样率下的

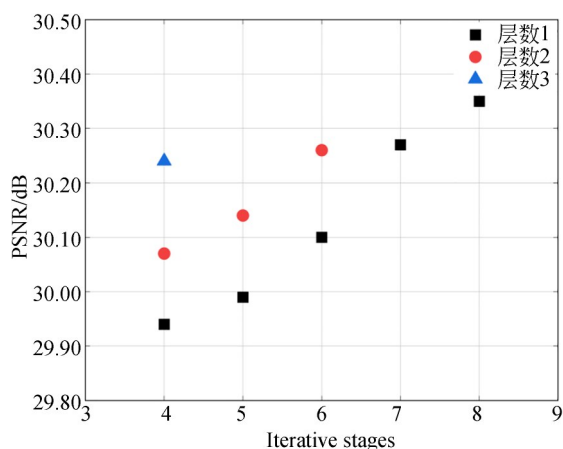


图 7 0.10 采样率不同迭代阶段数和混合架构层数重构性能

Fig. 7 Reconfiguration performance with different iterative stages and mixed architectural layers at 0.10 sampling rate

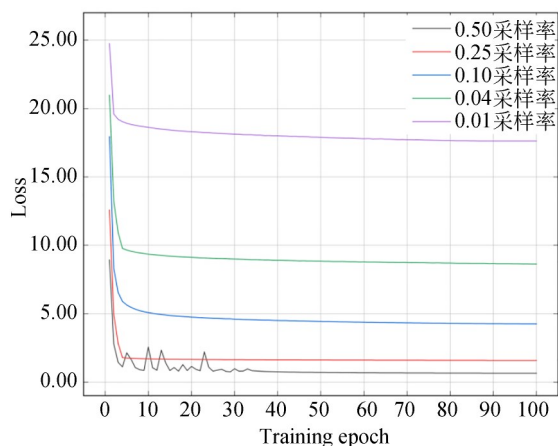


图8 不同采样率下 Loss 随训练 epoch 的变化曲线

Fig. 8 Variation curves of Loss with training epoch at different sampling rates

第 3~25 个 epoch 期间出现轻微振荡,但 Loss 值始终保持在 2.50 以下。此外,一个显著现象是:每当采样率减半, Loss 收敛值约为上一采样率的

2 倍,这进一步验证了采样率对重构任务及算法收敛性的影响程度。

4.3.4 特征分析

为了深入探究 SSCM 中局部、区域和全局分支在特征建模上的差异,本文在 0.25 采样率下,对不同迭代阶段中 SSCM 的三个分支所捕捉的层级特征进行了可视化。如图 9 所示,局部分支在浅层迭代优化时的图像建模能力相对较弱,造成这种情况的一个可能原因是:纯 CNN 架构的感受野较小,限制了图像建模的广度。区域分支通过使用 CBAM 块高效提取图像的关键特征,从而捕捉到更广泛的像素关系,特别是图像的凸出边缘部分。相比之下,全局分支得益于 Transformer 的长距离依赖关系建模能力,使其在浅层迭代优化的建模范围显著优于前两个分支。此外,通过自注意力的相似性比较,全局分支使网络能够学习到更丰富的纹理细节,使其在深层迭代优化中表现出更高的细粒度增益。

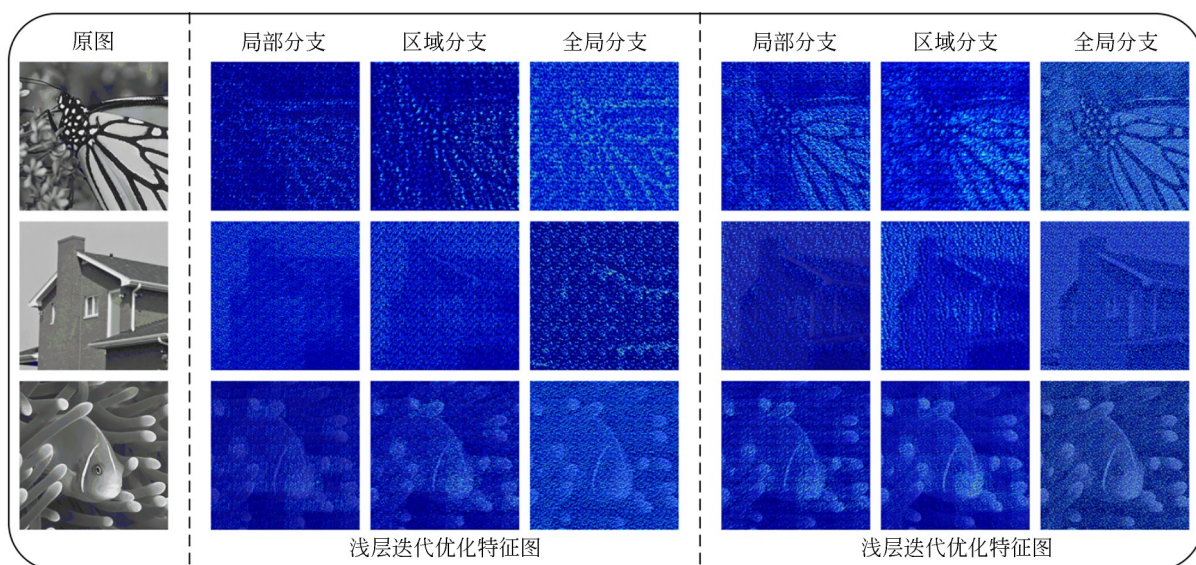


图9 0.25 采样率下不同迭代阶段中 SSCM 的层级特征图

Fig. 9 Hierarchical-feature map of SSCM in different iterative stages at 0.25 sampling rate

5 结 论

本文提出了一种联合图像层级特征的压缩感知迭代重构算法 CHFNet,在提高图像重构质量的同时减少了重构时间。采样子网络,通过自适应学习采样矩阵有效减少了冗余测量值。初

始重构子网络通过线性变换得到初始恢复图像;混合重构子网络则采用迭代使用梯度下降操作和特征优化操作的策略来逐步提升图像重构质量,并进一步整合了 CNN 与 Transformer 的优势,通过建模和优化图像层级特征,增强了网络对高分辨率图像的感知能力,同时降低了模型复

杂度。实验结果表明,所提算法在重构高分辨率图像时表现出更大的增益,在 Urban100 数据集

上重构图像的平均 PSNR 相较 CSformer 提升了 0.63 dB,平均重构时间减少了 4.775 0 s。

参考文献:

- [1] DONOHO D L. Compressed sensing [J]. *IEEE Transactions on Information Theory*, 2006, 52(4): 1289-1306.
- [2] DUARTE M F, DAVENPORT M A, TAKHAR D, *et al.* Single-pixel imaging via compressive sampling [J]. *IEEE Signal Processing Magazine*, 2008, 25(2): 83-91.
- [3] ZHU L, WU X, SUN Z Y, *et al.* Compressed-sensing accelerated 3-dimensional magnetic resonance cholangiopancreatography: application in suspected pancreatic diseases[J]. *Investigative Radiology*, 2018, 53(3): 150-157.
- [4] 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.
- [5] KUMAR S, MAHADEVAPPA M, DUTTA P K. Compressive holography from poisson noise plagued holograms using expectation-maximization [J]. *IEEE Transactions on Computational Imaging*, 2020, 6: 857-867.
- [6] TROCEN M, TRAMEL E W, FOWLER J E, *et al.* Compressed-sensing recovery of multiview image and video sequences using signal prediction[J]. *Multimedia Tools and Applications*, 2014, 72(1): 95-121.
- [7] AFONSO M V, BIOUCAS-DIAS J M, FIGUEIREDO M A T. An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems[J]. *IEEE Transactions on Image Processing*, 2011, 20(3): 681-695.
- [8] METZLER C A, MALEKI A, BARANIUK R G. From denoising to compressed sensing [J]. *IEEE Transactions on Information Theory*, 2016, 62(9): 5117-5144.
- [9] BAYRAM İ. On the convergence of the iterative shrinkage/thresholding algorithm with a weakly convex penalty[J]. *IEEE Transactions on Signal Processing*, 2016, 64(6): 1597-1608.
- [10] KULKARNI K, LOHIT S, TURAGA P, *et al.* ReconNet: non-iterative reconstruction of images from compressively sensed measurements [C]. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA. IEEE, 2016: 449-458.
- [11] SHI W Z, JIANG F, LIU S H, *et al.* Image compressed sensing using convolutional neural network [J]. *IEEE Transactions on Image Processing*, 2020, 29: 375-388.
- [12] 田金鹏,侯保军. 基于深度展开自注意力网络的压缩感知图像重构[J/OL]. 吉林大学学报(工学版), 1-9. (2023-04-10) [2024-02-27]. <http://kns.cnki.net/kcms/detail/22.1341.t.20230407.1455.005.html>.
- [13] TIAN J P, HOU B J. Compressive sensing image reconstruction based on deep unfolding self-attention network [J/OL]. *Journal of Jilin University (Engineering and Technology Edition)*, 1-9. (2023-04-10) [2024-02-27]. <http://kns.cnki.net/kcms/detail/22.1341.t.20230407.1455.005.html>. (in Chinese)
- [14] ZHANG Z H, LIU Y P, LIU J N, *et al.* AMP-net: denoising-based deep unfolding for compressive image sensing [J]. *IEEE Transactions on Image Processing*, 2021, 30: 1487-1500.
- [15] ZHANG J, GHANEM B. ISTA-net: interpretable optimization-inspired deep network for image compressive sensing [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA. IEEE, 2018: 1828-1837.
- [16] YOU D, XIE J F, ZHANG J. ISTA-NET: flexible deep unfolding network for compressive sensing [C]. 2021 *IEEE International Conference on Multimedia and Expo (ICME)*. Shenzhen, China. IEEE, 2021: 1-6.
- [17] ZHANG J, ZHAO C, GAO W. Optimization-inspired compact deep compressive sensing [J]. *IEEE Journal of Selected Topics in Signal Processing*, 2020, 14(4): 765-774.
- [18] 陈文俊,杨春玲. 图像压缩感知的特征域优化及自注意力增强神经网络重构算法[J]. 电子学报, 2022, 50(11): 2629-2637.
- [19] CHEN W J, YANG C L. Feature-space optimization-inspired and self-attention enhanced neural network reconstruction algorithm for image com-

- pressive sensing [J]. *Acta Electronica Sinica*, 2022, 50(11): 2629-2637. (in Chinese)
- [18] VASWANI A, SHAZEER N, PARMAR N, *et al.* Attention is all you need[EB/OL]. 2017: *arXiv*: 1706.03762. <http://arxiv.org/abs/1706.03762>
- [19] WANG X L, GIRSHICK R, GUPTA A, *et al.* Non-local neural networks[C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA. IEEE, 2018: 7794-7803.
- [20] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, *et al.* An image is worth 16x16 words: transformers for image recognition at scale[EB/OL]. 2020: *arXiv*: 2010.11929. <http://arxiv.org/abs/2010.11929>
- [21] LIU Z, LIN Y T, CAO Y, *et al.* Swin Transformer: hierarchical Vision Transformer using Shifted Windows[C]. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*. Montreal, QC, Canada. IEEE, 2021: 9992-10002.
- [22] SHEN M H, GAN H P, NING C, *et al.* TransCS: a transformer-based hybrid architecture for image compressed sensing[J]. *IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society*, 2022, 31: 6991-7005.
- [23] YE D J, NI Z K, WANG H L, *et al.* CSformer: bridging convolution and transformer for compressive sensing[J]. *IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society*, 2023, 32: 2827-2842.
- [24] LI Y W, FAN Y C, XIANG X Y, *et al.* Efficient and explicit modelling of image hierarchies for image restoration[C]. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vancouver, BC, Canada. IEEE, 2023: 18278-18289.
- [25] GUO M H, XU T X, LIU J J, *et al.* Attention mechanisms in computer vision: a survey [J]. *Computational Visual Media*, 2022, 8 (3): 331-368.
- [26] ARBELÁEZ P, MAIRE M, FOWLKES C, *et al.* Contour detection and hierarchical image segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(5): 898-916.
- [27] ZHANG J, ZHAO D B, GAO W. Group-based sparse representation for image restoration [J]. *IEEE Transactions on Image Processing*, 2014, 23 (8): 3336-3351.
- [28] MARTIN D, FOWLKES C, TAL D, *et al.* A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics[C]. *Proceedings Eighth IEEE International Conference on Computer Vision*. ICCV. Vancouver, BC, Canada. IEEE, 2001: 416-423.
- [29] HUANG J B, SINGH A, AHUJA N. Single image super-resolution from transformed self-exemplars[C]. 2015 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, MA, USA. IEEE, 2015: 5197-5206.
- [30] KINGMA D P, BA J. Adam: a method for stochastic optimization [J]. *ArXiv e-Prints*, 2014: arXiv: 1412.6980.
- [31] LOSHCILLOV I, HUTTER F. SGDR: stochastic gradient descent with warm restarts[J]. *ArXiv e-Prints*, 2016: arXiv: 1608.03983.
- [32] 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: a Publication of the IEEE Signal Processing Society*, 2004, 13(4): 600-612.

作者简介:



刘玉红(1975—),女,黑龙江鸡西人,硕士,副教授,2005年于兰州交通大学获得硕士学位,主要从事压缩感知、语音/图像信号处理、深度学习等方面的研究。E-mail: liuyuhong@mail.lzjtu.cn

通讯作者:



杨恒(1999—),男,甘肃天水人,硕士研究生,2022年于兰州城市学院获得学士学位,主要从事压缩感知、图像处理等方面的研究。E-mail: 11220697@stu.lzjtu.edu.cn