

# 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

陈文俊, 杨春玲

(华南理工大学电子与信息学院, 广东广州 510640)

**摘 要:** 现有的图像压缩感知(Image Compressive Sensing, ICS)优化启发网络沿用了传统算法的像素域优化思想, 构建了像素域的图像信息流动通道, 而没有充分利用卷积神经网络所提取的图像特征中的信息. 对此, 本文提出了在特征域构建信息流的思想, 并设计了一种特征域优化启发 ICS 网络(Feature-Space Optimization-Inspired Network, FSOINet)以实现该思想. 考虑到卷积操作感受野较小, 本文通过将自注意力模块引入 FSOINet 以更高效地利用图像非局部自相似性, 进一步提高重构质量, 我们将其命名为 FSOINet<sup>+</sup>. 此外, 本文还提出把迁移学习策略应用于不同采样率图像压缩感知重构网络训练中, 提高网络学习效率与重构质量. 仿真实验表明, 本文所提出的网络在峰值信噪比(Peak Signal to Noise Ratio, PSNR)、结构相似性(Structural Similarity Index Measure, SSIM)与视觉效果上都优于现有的最优 ICS 重构方法, FSOINet 与 FSOINet<sup>+</sup> 在 Set11 数据集上与 OPINet<sup>+</sup> 相比重构图像 PSNR 分别平均提升了 1.04 dB 和 1.27 dB.

**关键词:** 图像压缩感知; 深度学习; 卷积神经网络; 自注意力; 图像重构; 迁移学习

中图分类号: TN919.8

文献标识码: A

文章编号: 0372-2112(2022)11-2629-09

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20220155

## Feature-Space Optimization-Inspired and Self-Attention Enhanced Neural Network Reconstruction Algorithm for Image Compressive Sensing

CHEN Wen-jun, YANG Chun-ling

(School of Electronic and Information Engineering, South China University of Technology, Guangzhou, Guangdong 510640, China)

**Abstract:** The existing optimization-inspired networks for image compressive sensing(ICS) implement information optimization and flow in the pixel domain following the traditional algorithms, which does not make full use of the information in the image feature maps extracted by the convolutional neural network. This paper proposes the idea of constructing information flow in the feature domain. A feature-space optimization-inspired network(FSOINet) is designed to implement this idea. Considering the small receptive field of the convolution operation, this paper introduces the self-attention module into FSOINet to efficiently utilize the non-local self-similarity of images to further improve the reconstruction quality, which is named FSOINet<sup>+</sup>. In addition, this paper proposes a training strategy that applies transfer learning to the ICS reconstruction network training for different sampling rates to improve the network learning efficiency and reconstruction quality. Experimental results show that the proposed method is superior to the existing state-of-the-art ICS methods in peak signal to noise ratio(PSNR), structural similarity index measure(SSIM) and the visual effect. Compared with OPINet<sup>+</sup> on the Set11 dataset, FSOINet and FSOINet<sup>+</sup> have an average PSNR improvement of 1.04 dB/1.27 dB respectively.

**Key words:** image compressive sensing; deep learning; convolutional neural networks; self-attention; image reconstruction; transfer learning

### 1 引言

压缩感知理论<sup>[1]</sup>旨在对信号进行采样的同时完成

对信号的压缩, 以减少信号采集时的采样端资源浪费, 成为传统采样技术在采样端资源受限条件下的一个补

充方案,在单像素相机<sup>[2]</sup>、核磁共振成像<sup>[3]</sup>、快照压缩成像<sup>[4]</sup>、视频压缩<sup>[5]</sup>等信息采集任务中得以成功地应用。

压缩感知理论指出,对于可稀疏表示的信号,可以以远低于奈奎斯特采样定理的采样率对其进行采样并高概率地恢复出原信号<sup>[1]</sup>。对于图像信号,直接对全图进行压缩感知采样会带来较大的计算负担,因此文献[6]提出了基于块的图像压缩感知算法(Block-based Compressive Sensing, BCS),其对图像分块处理,同时降低了采样端和重构端的计算压力,成为了后续图像压缩感知(Image Compressive Sensing, ICS)算法研究的基础。

传统 ICS 算法<sup>[7,8]</sup>基于图像先验知识人工设计参数,通过迭代优化算法求解优化问题得到重构图像,但人工设计的参数限制了重构图像的质量,且重构时间受迭代算法的高计算复杂度限制。得益于深度学习能从数据集中自适应学习图像先验知识的优点,众多有着更高的重构质量,更快的重构速度的基于神经网络的 ICS 算法被相继提出。Kulkarni 等人<sup>[9]</sup>首次使用卷积神经网络完成对图像的分块重构。Shi 等人<sup>[10]</sup>采用分块采样、全图重构的策略以抑制分块采样带来的块效应,同时将采样矩阵加入网络训练过程,设计了重构网络 CSNet,随后作者对其进行了改进,先后提出了 SCS-Net<sup>[11]</sup>,CSNet<sup>[12]</sup>。然而,这些方法仅在初始重构时使用观测值,没有充分利用观测值信息,使得重构图像纹理细节较为缺失,因此文献[13]引入残差重构思想,多次利用观测值以补充重构图像的纹理细节。文献[14]对不同图像块自适应地分配不同采样率以增强重构图像的纹理细节。上述方法均将神经网络视作“黑盒子”,直接将观测值通过复杂的网络映射为图像,可解释性较差。针对该问题,文献[15]提出了 ISTANet,将传统迭代优化算法展开为多层卷积神经网络,为压缩感知重构网络引入了传统优化算法中的数学先验知识,在提高重构质量的同时提高了网络的可解释性。随后,众多优化启发网络被相继提出,文献[16~18]将采样矩阵加入网络训练过程,用全图重构替代 ISTANet 的分块重构,深度展开不同的传统算法,提升了重构图像质量。文献[19]在优化启发网络结构上引入长短期记忆模块加强网络信息交互能力,进一步提高了重构效果。此类优化启发网络<sup>[15-19]</sup>使用网络单元实现近端梯度法的近端算子,在特征域完成对当前阶段图像的去噪过程,保留了近端梯度法在像素域的梯度下降操作,提升了图像压缩感知重构算法的整体效率,但其仍存在着几点不足:(1)其保留了传统算法中信息在像素域中传递的特点,即在像素域做更新操作,优化阶段之间只传递像素域图像信息,导致每个阶段网络所提取的特征信息无法被充分利用。(2)其仅将近端映射操作映射为网络结构

而沿用了传统迭代方法在像素域的梯度下降操作,没有充分发挥神经网络强大的学习能力挖掘观测值所包含的信息。(3)卷积神经网络为深度学习引入了包括局部相关性、权重共享的归纳偏置。这些归纳偏置推动了卷积神经网络在计算机视觉任务上的发展,但也限制了网络模型捕捉长距离依赖关系的能力<sup>[20]</sup>。现有的 ICS 优化启发网络仅通过基本的卷积单元构成,而单纯的卷积操作无法有效地利用图像的非局部自相似性。

基于以上分析,本文将传统优化算法逐步优化图像的思想转换为逐步优化图像特征,设计了一个特征域优化启发网络(Feature-Space Optimization-Inspired Network, FSOINet),在特征域实现信息更新。每个网络单元都对对应着一次优化操作,各单元之间传递的是网络提取的特征图而非像素域图像。网络构建了特征域的信息流动通道,使得网络在引入优化迭代思想的同时更充分地发挥卷积神经网络学习得到的特征处理能力。在 FSOINet 基础上,我们还引入自注意力操作捕捉特征向量间的非局部自相关性以增强图像特征,我们将其命名为 FSOINet<sup>+</sup>。

## 2 特征域优化及自注意力增强图像压缩感知重构网络

迭代优化算法通过将图像重构问题转化为优化问题以求解原图像。即给定一个原始信号  $\mathbf{x}_{N \times 1}$ ,压缩感知通过观测矩阵  $\Phi_{M \times N}$ ,其中  $M$  远小于  $N$ ,对信号线性采样得到观测值  $\mathbf{y}_{M \times 1}$ 。而病态的重构问题可以转化为优化问题:

$$\min_{\mathbf{x}} \frac{1}{2} \|\Phi \mathbf{x} - \mathbf{y}\|_2^2 + \lambda \psi(\mathbf{x}) \quad (1)$$

其中  $\psi(\mathbf{x})$  为正则项,由信号先验知识得到, $\lambda$  为正则化参数。传统压缩感知算法一般基于近端梯度下降法对该优化问题进行迭代求解,每次迭代均由梯度下降操作和近端投影组成:

$$\mathbf{r}^{(k)} = \mathbf{x}^{(k-1)} - \rho \Phi^T (\Phi \mathbf{x}^{(k-1)} - \mathbf{y}) \quad (2)$$

$$\mathbf{x}^{(k)} = \text{prox}_{\lambda, \psi}(\mathbf{r}^{(k)}) = \underset{\mathbf{x}}{\text{argmin}} \frac{1}{2} \|\mathbf{x} - \mathbf{r}\|_2^2 + \lambda \psi(\mathbf{x}) \quad (3)$$

其中  $k$  表示迭代次数, $\rho$  表示更新步长, $T$  表示矩阵的转置。

受近端梯度下降法启发,本文构建的压缩感知重构神经网络整体结构如图 1(a)所示,不同颜色示出了所包含的三个子网络:采样子网络,初始重构子网络和特征域深度重构子网络。采样子网络对应着采样操作,在采样的同时实现了数据的压缩。初始重构子网络通过对观测值的线性变换得到图像的初始估计,随后特征域深度重构子网络通过逐步优化图像特征图最终完成图像的非线性重构。在每个优化阶段中,我

们构建了一个特征域信息补充模块 (Feature-Space Information supplementing Module, FSIM) 利用观测值优化所提取的特征图, 补充纹理细节信息. 同时构建了一个双尺度去噪模块 (Dual scale Denoising Module, DDM) 在两个不同的尺度上对特征图进行去噪处理. 为了更高效地利用图像非局部自相似性, 我们进一步

引入窗口自注意力增强模块 (Window Self-attention Enhancement Module, WSEM) 通过自注意力操作捕捉特征向量间的非局部自相关性以增强图像特征. 此外, 本文还提出把迁移学习策略应用于不同采样率图像压缩感知重构网络训练中, 提高网络学习效率与重构质量.

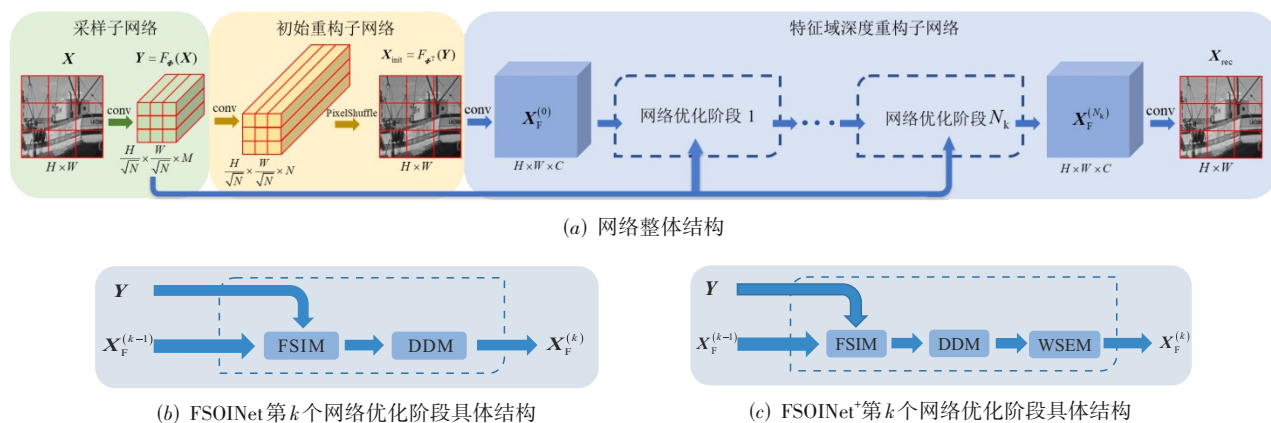


图 1 网络整体结构框架

## 2.1 采样与初始重构子网络

在 BCS 算法中, 一张图像  $X_{H \times W}$  先被分成大小为  $\sqrt{N} \times \sqrt{N}$  的不重叠的图像块, 每个图像块被展开为向量  $\mathbf{x}_{N \times 1}$  并通过采样矩阵完成线性采样. 当采样率为  $r$  时, 采样矩阵  $\Phi_{M \times N}$ , 其中  $M = \lfloor r \times N \rfloor$ . 与 CSNet<sup>[10]</sup> 相同, 我们使用可学习的采样矩阵来完成采样, 并通过不含偏置项的卷积来模拟分块线性采样的过程. 我们将  $\Phi$  设置为可学习的网络参数, 并将其重组成  $M$  个大小为  $\sqrt{N} \times \sqrt{N}$ , 步长为  $\sqrt{N} \times \sqrt{N}$  的卷积核  $W_\Phi$ , 因此我们的采样子网络表示为

$$Y = F_\Phi(X) = W_\Phi * X \quad (4)$$

其中  $F_\Phi(\bullet)$  表示采样子网络,  $*$  表示卷积操作.

为了从观测值  $Y$  中获得每个图像块的合理初始估计, 同时不引入更多网络学习参数, 我们使用采样矩阵的转置来完成观测域到图像域的线性变换, 实现初始重构. 与采样网络类似, 我们将  $\Phi^T$  重组成  $N$  个大小为  $1 \times 1 \times M$  的卷积核  $W_{\Phi^T}$ , 并辅以 PixelShuffle 操作以得到初始重构图, 整个初始重构子网络表示为下式:

$$X_{init} = F_{\Phi^T}(Y) = \text{PixelShuffle}(W_{\Phi^T} * Y) \quad (5)$$

其中  $F_{\Phi^T}(\bullet)$  表示初始重构子网络.

## 2.2 FSOINet 网络结构

现有的基于优化启发 ICS 网络均沿用了传统优化算法中对像素域图像逐步优化的思想, 每个优化阶段都通过梯度下降操作在像素域补充细节信息, 并通过神经网络在特征域完成去噪过程, 随后将去噪后的像

素域图像传递给下一阶段. 这种在每个阶段之间传递像素域信息的做法, 无法充分利用卷积神经网络在每个阶段所提取的特征信息, 限制了卷积神经网络强大的特征处理能力. 在本文中, 我们提出对特征进行优化的思想, 以充分利用卷积神经网络的特征表征能力, 同时加强阶段间的信息交互能力.

我们构建了 FSOINet 来实现上述思想, 其特征域优化深度重构子网络如图 1(a) 和图 1(b) 所示, 在得到初始重构图像后, 先通过大小  $3 \times 3$  的单层卷积从初始重构图中提取特征  $X_F^{(0)}$ , 此时的特征图缺少大量细节信息含有较多的噪声, 因此深度重构子网络通过多个网络优化阶段逐步优化所提取的特征图. 本节提出了 FSIM 与 DDM 分别将近端梯度下降法中梯度下降操作与近端投影操作展开至特征域实现, 每个优化阶段均对特征进行优化, 表示为式 (6):

$$R_F^{(k)} = F_{\text{FSIM}}(X_F^{(k-1)}, Y) \quad (6)$$

$$X_F^{(k)} = F_{\text{DDM}}(R_F^{(k)}) \quad (7)$$

经过  $N_k$  个优化阶段后, 我们从优化后的特征图  $X_F^{(N_k)}$  中通过大小  $1 \times 1$  卷积层还原出最终重构图像  $X_{rec}$ .

### 2.2.1 特征域信息补充模块

传统优化迭代算法通过梯度下降操作在重构过程中反复使用观测值, 以约束重构图像符合  $\Phi x = y$  的降质过程, 为了在引入迭代优化思想的同时发挥卷积神经网络的特征处理能力, 我们提出了通过观测值在特征域而不是像素域补充信息的思想, 构建了一个特征域信息补充模块通过观测值  $Y$  补充图像特征信息. 如图

2(a)和图2(c)所示,我们首先通过 $3 \times 3$ 卷积降低特征 $X_F^{(k-1)}$ 通道数,随后使用采样子网络将图像投影至观测域并与真实的观测值 $Y$ 作差,得到观测域的残差信息并通过初始重构子网络将其映射至图像域,最后通过 $3 \times 3$ 卷积与一个残差卷积单元(ResBlock)得到需要补充的特征域信息与 $X_F^{(k-1)}$ 相加得到 $R_F^{(k)}$ . 在每个优化阶段都通过观测域残差信息校准特征图,从而保证网络在通过潜在的图像先验知识优化特征的同时兼顾特征在观测域的映射与观测值 $Y$ 的一致性,通过ResBlock引入非线性提升了算法从观测值提取信息的能力.

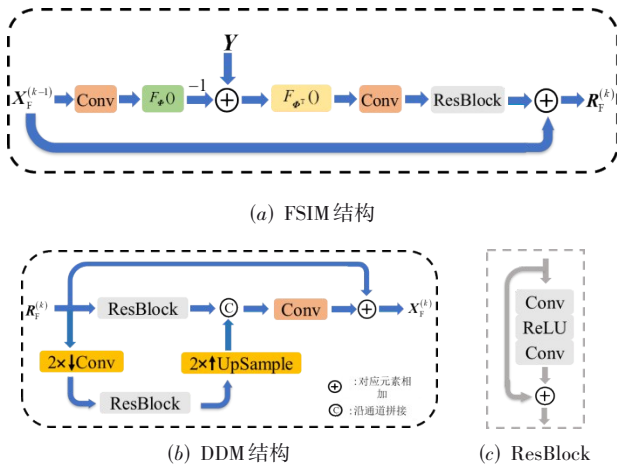


图2 特征域信息补充模块与双尺度去噪模块实现细节

### 2.2.2 双尺度去噪模块

得益于深度学习数据驱动的特性,我们可以从训练集中学习到图像的潜在先验知识对图像进行去噪. 多尺度网络可以综合多尺度的特征信息去除不同尺度的噪声以提高去噪效果,然而多尺度的结构会引入过多的参数量与计算量. 同时考虑到去噪效果与计算复杂度,本文设计了一个双尺度去噪模块在两个尺度上对特征进行去噪以提升去噪性能,同时每个尺度均使用简单的模块进行去噪以降低复杂度. 如图2(b)和图2(c)所示,DDM并行处理两个尺度的特征并通过 $1 \times 1$ 卷积加以融合以完成对特征的优化过程. 考虑到复杂度,在每个DDM中我们只使用两个ResBlock分别对高分辨率特征与低分辨率特征进行处理. 其中高分辨率特征分辨率与原图像一致,而低分辨率特征通过步长为2的卷积从高分辨率特征下采得到,通道数为高分辨率特征的两倍.

FSIM与DDM均包含一个大的跳跃连接,因此,两个模块均可看做是对特征的信息补充或优化操作. 通过堆叠这两个模块,我们构建的特征域优化启发网络形成了特征域的信息流动通道,使得网络在引入优化迭代思想的同时更充分地发挥卷积神经网络的特征处

理能力.

### 2.3 FSOINet<sup>+</sup>网络结构

卷积神经网络为深度学习引入了强大的归纳偏置,包括局部相关性、权重共享,推动了计算机视觉的发展. 但卷积的权重共享使其响应不能自适应于输入,卷积的局部性使其无法有效捕捉长距离依赖关系. 近年来,自注意力模型在计算机视觉任务中大放异彩,其依靠图像块或像素之间的相关性建模,高效地捕捉图像内部的相关性,取得了媲美卷积模型的成效. 卷积操作通过不断聚合局部信息得到更深层次的特征,每个像素点所对应的特征向量均包含着其周围像素位置的信息,相似的特征向量所对应的周围信息也彼此相似,通过注意力操作聚合所有特征向量之间的信息,能有效地对图像非局部自相似性进行建模. 如图1(c)所示,我们在FSOINet的基础上添加了一个窗口自注意力增强模块以实现上述思想,因此优化过程可以表示为下式:

$$R_F^{(k)} = F_{\text{FSIM}}(X_F^{(k-1)}, Y) \quad (8)$$

$$X_F^{(k)'} = F_{\text{DDM}}(R_F^{(k)}) \quad (9)$$

$$X_F^{(k)} = F_{\text{WSEM}}(X_F^{(k)'}) \quad (10)$$

我们将由式(8)~(10)堆叠而成的网络命名为FSOINet<sup>+</sup>. 其中式(8)与式(9)所含模块与2.2节所述一致,本节后半部分将主要阐述WSEM的具体实现方式.

全局像素级自注意力操作的时间空间复杂度随着输入图像分辨率的增长而二次增长,为了缓解该问题,我们将特征块不重叠地分成大小为 $S_w \times S_w$ 的窗口,独立地在每个窗口内做自注意力操作,同时为了在一定程度上加强窗口之间的联系并提高感受野,在执行自注意力操作之前,我们使用了一个 $7 \times 7$ 的逐通道卷积对特征进行预处理以聚合局部信息,同时起到交互不同窗口间信息的作用. 在每个窗口内,我们将每个像素位置所对应的特征向量投影至嵌入空间计算相关性以执行自注意力操作,具体表示为式(11):

$$\hat{x}_i = \sum_{v_j} \frac{e^{x_i^T W_\theta^T W_\phi x_j}}{\sum_{v_j} e^{x_i^T W_\theta^T W_\phi x_j}} x_j \quad (11)$$

其中 $x_i$ 指窗口内 $i$ 位置所对应的特征向量, $W_\theta$ 与 $W_\phi$ 为投影矩阵. 整个模块的具体实现如图3所示,我们在WSEM中同样使用了跳跃连接,因此没有破坏网络在特征域的信息流动通道,同时模块头的逐通道卷积还承担着将特征映射至残差域的作用.

### 2.4 迁移学习策略

图像压缩感知网络<sup>[12-19]</sup>对于不同采样率往往需要独立训练多个模型以适应不同采样率的重构,需要大量的训练资源. 考虑到不同采样率的重构任务具有高度相关性,我们将迁移学习的思想引入压缩感知网络

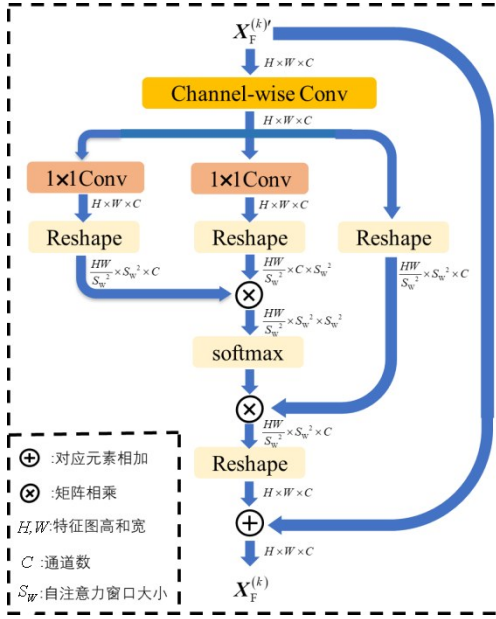


图3 WSEM结构

模型的训练,在完成一个高采样率网络模型的训练后,对于低采样率网络模型,我们将深度重构子网络的所有参数从高采样率网络模型迁移至低采样率网络模型,将高采样率采样矩阵  $\Phi_{M \times N}$  的前  $M'$  行作为低采样率采样矩阵  $\Phi_{M' \times N}$  的初始参数,从而低采样率网络只需要使用同样的训练集微调深度重构子网络和采样矩阵的参数,以此提高了学习效率,大大节约了训练资源.同时高采样率模型学习得到的更为丰富的图像潜在先验知识能指导低采样率模型的训练,提升低采样率模型的重构性能.

### 2.5 损失函数

我们的网络以原始图像  $X_{H \times W}$  为输入,首先通过采样子网络得到压缩感知观测值  $Y$ ,随后通过重构子网络得到最终的重构图像  $X_{rec}$ . 我们通过以下损失函数端到端优化我们的网络:

$$L = L_{mse}(X_{rec}, X) + \gamma L_{orth}(\Phi) \quad (12)$$

其中  $L_{mse}$  是原始图像  $X$  和重构图像  $X_{rec}$  的均方误差 (Mean Square Error, MSE), 具体表示为

$$L_{mse}(X_{rec}, X) = \frac{1}{HW} \|X_{rec} - X\|_F^2 \quad (13)$$

其中,  $\|\cdot\|_F^2$  表示矩阵的 Frobenius 范数的平方,即矩阵的所有元素的平方和.

$L_{orth}$  是文献[16]提出的采样矩阵正交约束,用以约束采样矩阵行与行之间的相关性,以尽可能保证观测值之间存在更少冗余,具体表示为式(14):

$$L_{orth}(\Phi) = \frac{1}{M^2} \|\Phi\Phi^T - I\|_F^2 \quad (14)$$

其中,  $I$  表示单位矩阵,  $M$  表示观测值的数量. 在我们的

实验中,式(12)的正则化参数  $\gamma$  被设置为 0.01.

## 3 仿真实验结果及分析

### 3.1 网络训练细节与数据集

在训练中,为了得到更公平的比较结果,我们使用与文献[12]一致的训练集,其包括 BSD500 数据集<sup>[21]</sup>中训练集(200 张图像)和测试集(200 张图像)共 400 张图像. 我们先将训练图像随机裁剪成 96 pixel×96 pixel 大小的子图像并使用旋转、翻转增强训练集,最终对数据增强后的图像进行灰度化处理得到 89600 个子图像并将其作为网络的训练图像. 我们使用 Adam<sup>[22]</sup> 优化器对网络进行训练. 对于 0.5 采样率的网络训练,前三个 epoch 作为 warm-up<sup>[23]</sup> 阶段,学习率从 0 线性增长为  $2 \times 10^{-4}$ ,随后 97 个 epoch 将学习率余弦下降至  $5 \times 10^{-5}$ ;对于其他低采样率的网络训练,我们使用 2.4 节的迁移学习策略,将学习率设为  $1 \times 10^{-4}$  微调 40 个 epoch. 在我们的网络模型中,采样块大小  $\sqrt{N} = 32$ ,特征通道数  $C = 16$ ,优化阶段数  $N_{opt} = 16$ ,未特殊说明的卷积核大小均为  $3 \times 3$ ,FSOINet 的 batchsize 为 32,FSOINet<sup>+</sup> 的 batchsize 为 6.

所有实验均在搭载 Intel Core i5-6500 处理器和 GTX2080ti GPU 的 PyTorch 平台上完成. 实验所使用的测试集为三个在图像压缩感知中广泛使用的基准数据集,包括 Set11<sup>[9]</sup>、BSDS68<sup>[24]</sup> 和 Urban100<sup>[25]</sup>,彩色图像均在亮度通道上评估重构结果,并采用峰值信噪比 (Peak Signal to Noise Ratio, PSNR) 和结构相似度 (Structural Similarity Index Measure, SSIM) 来评价重构结果.

### 3.2 与现有算法重构性能对比

我们选取了传统 ICS 方法中重构质量最优的 GSR<sup>[8]</sup> 与七种前沿的使用参数可学习采样矩阵的 ICS 方法进行了比较,包括 CSNet<sup>+[12]</sup>、SCSNet<sup>[11]</sup>、SPLNet<sup>[16]</sup>、AMPNet<sup>[17]</sup>、OPINENet<sup>+[18]</sup>、BCSNet<sup>[14]</sup>、MADUN<sup>[19]</sup> 和我们网络的初步版本<sup>[26]</sup>,我们将其命名为 FSOINet<sup>\*</sup>. 其中,CSNet<sup>+</sup>、SCSNet、BCSNet 为普通的压缩感知网络,SPLNet、AMPNet、OPINENet<sup>+</sup> 和 MADUN 为基于优化启发的压缩感知网络. 为了更全面地比较,我们选取了从 0.01 到 0.5 中的五个采样率进行仿真实验.

由表 1 可见,在 0.01 采样率下,基于深度学习的 ICS 方法的重构质量主要取决于网络在特征域的推理能力,普通的压缩感知网络得益于复杂的特征处理模块和特征域信息流动通道取得了优于现有的基于优化启发的压缩感知网络的重构质量. 然而随着采样率的升高,观测值所含的信息量增加,GSR 和基于优化启发的压缩感知网络通过优化步骤不断使用观测值优化重构图像,重构质量逐渐高过普通的压缩感知网络. 而我们的方法在特征域展开传统优化算法,构建了特征域信息流动通道同时能够多次利用观测值优化特征,无论

在低采样率或高采样率下均优于上述方法,FSOINet 重  
构图像平均 PSNR/SSIM 较 OPINENet<sup>+</sup>提高了 1.09 dB/

0.0202. FSOINet<sup>+</sup>在 Set11 数据集上平均重构 PSNR/SSIM  
较 FSOINet 再提升了 0.19 dB/0.0028.

表 1 Set11 数据集中各采样率不同算法重构图像 PSNR(dB)/SSIM 对比

方法	采样率					
	0.01	0.05	0.1	0.3	0.5	平均
GSR	16.78/0.4520	22.79/0.7155	26.64/0.8562	34.77/0.9466	38.76/0.9721	27.95/0.7885
SCSNet	21.04/0.5562	25.85/0.7839	28.52/0.8616	34.64/0.9511	39.01/0.9769	29.81/0.8259
CSNet <sup>+</sup>	21.02/0.5566	25.86/0.7846	28.34/0.8508	34.30/0.9490	38.52/0.9749	29.61/0.8232
SPLNet	21.22/0.5552	26.59/0.8177	29.49/0.8874	35.79/0.9603	40.27/0.9815	30.67/0.8404
OPINENet <sup>+</sup>	20.02/0.5362	26.36/0.8186	29.81/0.8904	36.04/0.9600	40.19/0.9800	30.48/0.8370
BCSNet	20.81/0.5427	26.50/0.7893	29.36/0.8650	35.40/0.9527	—	—
AMP-Net	20.20/0.5425	26.17/0.8128	29.40/0.8876	36.03/0.9623	40.34/0.9821	30.43/0.8375
MADUN	—	—	29.91/0.8986	36.94/0.9676	40.77/0.9832	—
FSOINet <sup>+</sup>	21.73/0.5937	27.36/0.8415	30.44/0.9018	37.00/0.9665	41.08/0.9832	31.52/0.8573
FSOINet	<u>21.88/0.5958</u>	<u>27.30/0.8387</u>	<u>30.57/0.9020</u>	<u>37.00/0.9664</u>	<u>41.10/0.9833</u>	<u>31.57/0.8572</u>
FSOINet <sup>+</sup>	<b>21.91/0.5984</b>	<b>27.47/0.8437</b>	<b>30.81/0.9056</b>	<b>37.30/0.9678</b>	<b>41.29/0.9837</b>	<b>31.76/0.8600</b>

为了进一步验证本文网络的泛化性能,我们将 FSOI-  
Net 和 FSOINet<sup>+</sup>与 CSNet<sup>+</sup>、SCSNet、SPLNet、AMPNet、  
OPINENet<sup>+</sup>在 BSDS68 和 Urban100 上进行比较,如表 2 所  
示,我们的方法重构图像 PSNR 和 SSIM 均明显优于其他  
五种方法. 在 BSDS68 数据集上我们的 FSOINet 和 FSOI-  
Net<sup>+</sup>较五种方法中性能最好的 SPLNet 分别平均提升  
0.46 dB 与 0.54 dB, 在 Urban100 数据集上我们的 FSOI-  
Net 较五种方法中性能最好的 OPINENet<sup>+</sup>平均提升  
1.04 dB, 由于 Urban100 数据集中图像的规律性纹理较  
多,自注意力操作所能带来的性能提升也更高,FSOINet<sup>+</sup>  
较 FSOINet 进一步提升了 0.30 dB. 图 4、图 5 给出了不同

采样率下不同重构方法的重构图像的视觉效果对比.

如图 4 所示,对于 0.1 采样率下的 Barbara 图像,CS-  
Net<sup>+</sup>与 SCSNet 均无法还原出围巾的纹理,细节信息损失  
严重;而 SPLNet 与 AMPNet 引入了迭代优化的思想,重  
建出了部分纹理,但纹理处有着明显的振铃效应;  
OPINENet<sup>+</sup>重建出错误的纹理;而 FSOINet 重建出了较大  
一部分的纹理且条纹边缘更为清晰,而 FSOINet<sup>+</sup>重构图  
像纹理细节最为饱满. 在图 5 更高采样率的重构结果  
中,可以明显看出,前五种方法重建的左窗户都有着不  
同程度的模糊和人工效应,而 FSOINet 与 FSOINet<sup>+</sup>重建  
出的石砖与窗户具有更少的噪声与更为清晰的纹理.



图 4 0.1 采样率下图像 Barbara(Set11)的重构图像视觉效果对比

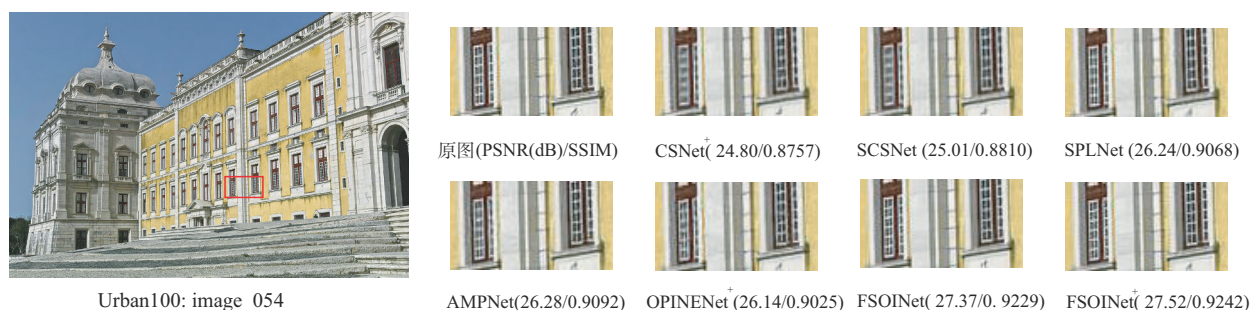


图 5 0.3 采样率下图像 image\_054(Urban100)的重构图像视觉效果对比

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

数据集	方法	采样率					平均
		0.01	0.05	0.1	0.3	0.5	
BSDS68	CSNet <sup>+</sup>	21.71/0.5249	25.04/0.6845	26.89/0.7756	31.66/0.9152	35.42/0.9614	28.14/0.7723
	SCSNet	21.88/0.5250	24.98/0.6843	27.13/0.7785	31.76/0.9173	35.67/0.9640	28.28/0.7738
	SPLNet	22.33/0.5242	25.87/0.7198	27.85/0.8094	32.77/0.9303	36.86/0.9708	29.13/0.7907
	AMP-Net	22.28/0.5315	25.77/0.7204	27.85/0.8113	32.84/0.9321	36.82/0.9715	29.11/0.7934
	OPINENet <sup>+</sup>	21.88/0.5162	25.66/0.7136	27.81/0.8040	32.50/0.9236	36.32/0.9658	28.83/0.7846
	FSOINet	<u>22.80/0.5435</u>	<u>26.24/0.7328</u>	<u>28.28/0.8185</u>	<u>33.28/0.9345</u>	<u>37.36/0.9728</u>	<u>29.59/0.8004</u>
	FSOINet <sup>+</sup>	<b>22.83/0.5441</b>	<b>26.27/0.7340</b>	<b>28.39/0.8210</b>	<b>33.37/0.9352</b>	<b>37.47/0.9732</b>	<b>29.67/0.8015</b>
Urban100	CSNet <sup>+</sup>	19.27/0.4812	22.63/0.6792	24.64/0.7741	29.90/0.9162	33.55/0.9572	26.00/0.7616
	SCSNet	19.28/0.4798	22.63/0.6774	24.93/0.7827	30.12/0.9193	33.92/0.9601	26.18/0.7639
	SPLNet	19.55/0.4873	23.55/0.7301	26.19/0.8290	32.11/0.9405	36.41/0.9737	27.56/0.7921
	AMP-Net	19.62/0.4969	23.45/0.7290	26.04/0.8283	32.19/0.9418	36.33/0.9737	27.53/0.7939
	OPINENet <sup>+</sup>	19.38/0.4872	23.70/0.7363	26.61/0.8362	32.58/0.9414	36.62/0.9727	27.78/0.7948
	FSOINet	<u>20.05/0.5257</u>	<u>24.66/0.7761</u>	<u>27.62/0.8623</u>	<u>33.88/0.9541</u>	<u>37.91/0.9788</u>	<u>28.82/0.8194</u>
	FSOINet <sup>+</sup>	<b>20.14/0.5331</b>	<b>24.80/0.7805</b>	<b>28.05/0.8718</b>	<b>34.29/0.9569</b>	<b>38.31/0.9800</b>	<b>29.12/0.8245</b>

为了比较不同方法的时间与空间复杂度,本文给出了 0.5 和 0.1 采样率下五个优化启发网络的参数量与平均运行时间对比. 对比数据均使用作者提供的源码在相同环境下对 Set11 与 BSD68 数据集进行测试得到. 如表 3 所示,由于使用的重构网络结构大小相同,同一方法不同采样率模型的运行时间相近. 本文网络参数量均少于其他三种优化启发方法,同时在不同的测试集上均拥有中等的运行时间.

表 3 不同采样率不同算法在 Set11 与 BSDS68 上的复杂度对比

方法	参数量	0.5 采样率		0.1 采样率	
		Set11	BSDS68	Set11	BSDS68
		平均运行时间/s			
SPLNet	1.388 M	0.0061	0.0076	0.0090	0.0089
AMPNet	1.529 M	0.0562	0.0671	0.0564	0.0649
OPINENet <sup>+</sup>	1.095 M	0.0087	0.0126	0.0134	0.0132
FSOINet	1.061 M	0.0198	0.0184	0.0215	0.0190
FSOINet <sup>+</sup>	1.086 M	0.0294	0.0264	0.0283	0.0258

### 3.3 消融实验与分析

#### 3.3.1 网络单元数量

由于网络每一个优化阶段对应于传统迭代算法中的一次迭代,可以预期随着网络优化阶段数的增多,重构性能也会随之提高,图 6 给出了在 0.1 采样率情况下,不同优化阶段数  $N_{opt}$  训练得到的 FSOINet 在 Set11 数据集的重构结果,可以看出随着优化阶段数  $N_{opt}$  的增加,网络的重构性能也随之增加,但时间空间复杂度也会随之增加. 综合考虑模型性能与复杂度,选取  $N_{opt} = 16$  为我们的默认优化阶段数.

#### 3.3.2 模块消融

本节在 0.5 采样率下对三个模块进行了消融实验.

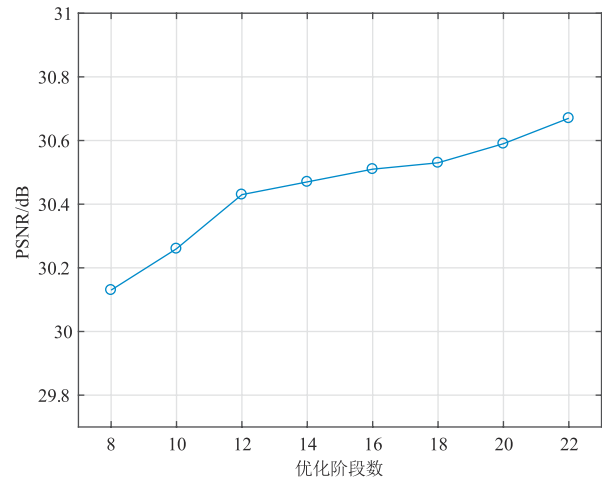


图 6 0.1 采样率不同优化阶段数在 Set11 上的平均重构结果

如表 4 所示,在高采样率下由于观测值所含信息量更为丰富,FSIM 较 DDM 给予 FSOINet 更多的性能提升,在三个数据集下为重构图像分别带来了 1.45 dB, 1.00 dB 和 1.60 dB 的 PSNR 提升. WSEM 作为增强网络捕捉非局部自相似性能力的模块,进一步带来了 0.19 dB, 0.11 dB 和 0.40 dB 的 PSNR 性能提升. 综上所述,本文所提的三个功能模块均给压缩感知重构图像带来了一定的性能提升.

表 4 0.5 采样率下不同模块的消融实验

模型设置			测试集 PSNR/dB		
FSIM	DDM	WSEM	Set11	BSDS68	Urban100
√	×	×	40.22	36.82	36.31
×	√	×	39.65	36.36	35.38
√	√	×	41.10	37.36	37.91
√	√	√	41.29	37.47	38.31

### 3.3.3 迁移学习策略

图7给出了随机初始化参数进行训练与使用0.5采样率模型迁移获得的初始参数进行训练的0.1采样率FSOINet的loss曲线图。迁移训练由于使用更优秀的初始化参数,loss下降速度更快,在训练中期loss明显低于随机初始参数的模型,为了进一步给出不同训练方法所得模型的泛化性能,表5给出了两种方式训练得到的三个不同采样率的模型在Set11数据集上的仿真结果。当只训练40个epoch时,随机初始化训练的网络模型还未收敛,重构质量较低,而使用了迁移学习策略的网络模型拥有着更为合适的初始化参数,训练时更不容易陷入局部最优点,能够在较短的训练周期中完成收敛,以节约大量的训练时间;而单独将从头随机初始化训练的网络模型训练100个epoch后,尽管此时模型的loss已低于迁移训练所得模型,但由于高采样率模型学习得到的知识对低采样率模型训练所起的指导作用,使得迁移学习得到的网络模型具有更好的泛化性能,较随机初始化训练的网络模型有着平均0.04~0.07 dB的重构性能提升。

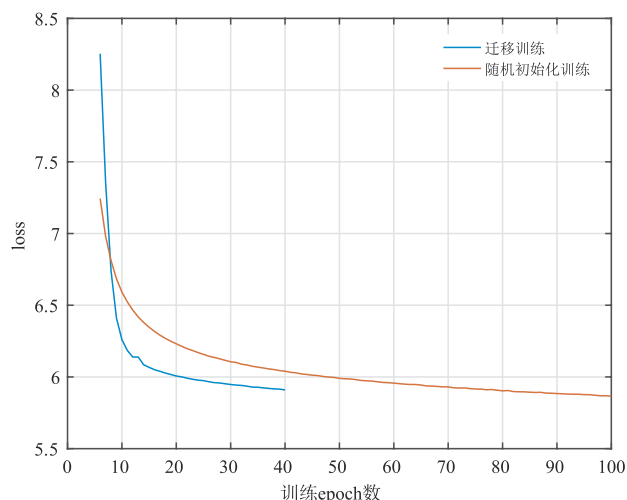


图7 0.1采样率下不同参数初始化策略的训练loss曲线图

表5 不同训练策略的FSOINet在Set11上的重构图像PSNR(dB)/SSIM

训练策略	0.01 采样率	0.1 采样率	0.3 采样率
随机初始化训练 40 epoch	21.77/0.5920	30.31/0.8995	36.74/0.9651
随机初始化训练 100 epoch	21.81/0.5940	30.51/0.9021	36.96/0.9660
迁移学习 40 epoch	21.88/0.5958	30.57/0.9020	37.00/0.9664

## 4 结论

本文提出了特征域优化启发网络FSOINet与自注意力增强特征域优化启发网络FSOINet<sup>+</sup>,通过逐步优化图像特征得到了高质量的重构结果。同时,本文还将迁

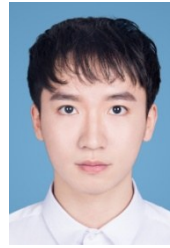
移学习策略应用于不同采样率图像压缩感知重构网络训练中,提高了网络学习效率与重构质量。未来工作将考虑将特征域优化思想延伸至视频重建任务中。

### 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, 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] 禰韵怡, 杨春玲. 基于帧间组稀疏的两阶段递归增强视频压缩感知重构网络[J]. 电子学报, 2021, 49(3): 435-442. XUAN Yun-yi, YANG Chun-ling. Two-stage recursive enhancement reconstruction based on video inter-frame group sparse representation in compressed video sensing [J]. Acta Electronica Sinica, 2021, 49(3): 435-442. (in Chinese)
- [6] GAN L. Block compressed sensing of natural images[C]// 15th International Conference on Digital Signal Processing. Cardiff: IEEE, 2007: 403-406.
- [7] CHEN C, TRAMEL E W, FOWLER J E. Compressed-sensing recovery of images and video using multihypothesis predictions[C]//2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers. Pacific Grove: IEEE, 2011: 1193-1198.
- [8] ZHANG J, ZHAO D, GAO W. Group-based sparse representation for image restoration[J]. IEEE Transactions on Image Processing, 2014, 23(8): 3336-3351.
- [9] 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. Las Vegas: IEEE, 2016: 449-458.
- [10] SHI W, JIANG F, ZHANG S, et al. Deep networks for compressed image sensing[C]//2017 IEEE International Conference on Multimedia and Expo. Hong Kong: IEEE, 2017: 877-882.
- [11] SHI W, JIANG F, LIU S, et al. Scalable convolutional neural network for image compressed sensing[C]//2019

- IEEE Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 12290-12299.
- [12] SHI W, JIANG F, LIU S, et al. Image compressed sensing using convolutional neural network[J]. IEEE Transactions on Image Processing, 2020, 29: 375-388.
- [13] CHEN J, SUN Y, LIU Q, et al. Learning memory augmented cascading network for compressed sensing of images[C]//2020 European Conference on Computer Vision. Glasgow: Springer, 2020: 513-529.
- [14] ZHOU S, HE Y, LIU Y, et al. Multi-channel deep networks for block-based image compressive sensing[J]. IEEE Transactions on Multimedia, 2021, 23: 2627-2640.
- [15] ZHANG J, GHANEM B. ISTA-Net: interpretable optimization-inspired deep network for image compressive sensing[C]//2018 IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 1828-1837.
- [16] 裴翰奇, 杨春玲, 魏志超, 曹燕. 基于 SPL 迭代思想的图像压缩感知重构神经网络[J]. 电子学报, 2021, 49(6): 1195-1203.  
PEI Han-qi, YANG Chun-ling, WEI Zhi-chao, CAO Yan. Image compressive sensing reconstruction network based on iterative SPL theory[J]. Acta Electronica Sinica, 2021, 49(6): 1195-1203. (in Chinese)
- [17] ZHANG Z, LIU Y, LIU J, et al. AMP-Net: denoising-based deep unfolding for compressive image sensing[J]. IEEE Transactions on Image Processing, 2021, 30: 1487-1500.
- [18] 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.
- [19] SONG J, CHEN B, ZHANG J. Memory-augmented deep unfolding network for compressive sensing[C]//2021 ACM International Conference on Multimedia. Chengdu: ACM, 2021: 4249-4258.
- [20] D' ASCOLI S, Touvron H, Leavitt M L, et al. Convit: improving vision transformers with soft convolutional inductive biases[C]//2021 International Conference on Machine Learning. Virtual Only: ACM, 2021: 2286-2296.
- [21] ARBEL' AEZ 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.
- [22] KINGMA D P, BA J. Adam: A method for stochastic optimization[EB/OL]. [2022-04-11]. <https://arxiv.org/abs/1412.6980>.
- [23] LOSHCHILOV I, HUTTER F. Sgdr: Stochastic gradient descent with warm restarts[EB/OL]. [2022-04-11]. <https://arxiv.org/abs/1608.03983>.
- [24] 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]//2001 IEEE International Conference on Computer Vision. Vancouver: IEEE, 2001: 416-423.
- [25] 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. Boston: IEEE, 2015: 5197-5206.
- [26] CHEN W, YANG C, YANG X. FSOINET: feature-space optimization-inspired network for image compressive sensing[EB/OL]. [2022-04-11]. <https://arxiv.org/abs/2204.05503>.

#### 作者简介



陈文俊 男, 1997年出生, 福建福州人. 现为华南理工大学电子与信息学院硕士研究生. 主要研究方向为图像/视频压缩感知.  
E-mail: eecwjun@mail.scut.edu.cn



杨春玲(通讯作者) 女, 1970年出生, 河南新乡人. 现为华南理工大学电子与信息学院博士生导师. 主要研究方向为图像/视频压缩编码、图像质量评价.  
E-mail: eeclyang@scut.edu.cn