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

禤韵怡,杨春玲
(华南理工大学电子与信息学院,广东广州 510640)

摘 要: 基于迭代优化的传统视频压缩感知重构算法运行时间长,参数的自适应性较低,限制了其实用性和泛化能力.利用神经网络强大的计算能力和运行速度快、参数可学习的优点,本文首先提出了帧间组稀疏网络(VGSR-Net),用神经网络将图像块组映射到更高维的稀疏表示域中,并利用可学习的阈值提取帧间相关特征.在此基础上,提出了两阶段混合递归增强重构网络(2sRER-VGSR-Net).首先,利用 VGSR-Net 对初始重构结果进行初步增强;然后,引入 STMC-Net 实现运动估计,并利用残差重构网络进一步重构当前帧丢失的信息,得到更高质量的重构结果.在第二阶段重构过程中采用混合递归结构,充分利用已有的高质量重构帧信息.仿真结果表明,所提算法与现有最优迭代优化重构算法 SSIM-InterF-GSR 相比重构性能提升了 1.99dB;和基于深度学习的重构网络 CSVideoNet 相比,性能提升了 4.60dB.

关键词: 视频压缩感知; 深度学习; 帧间组稀疏表示; 混合递归网络; 运动估计; 增强重构
中图分类号: TN919.8 **文献标识码:** A **文章编号:** 0372-2112 (2021)03-0435-08
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.12263/DZXB.20200272

Two-Stage Recursive Enhancement Reconstruction Based on Video Inter-frame Group Sparse Representation in Compressed Video Sensing

XUAN Yun-yi, YANG Chun-ling
(*School of Electronic and Information Engineering, South China University of Technology, Guangzhou, Guangdong* 510640, *China*)

Abstract: The traditional iterative optimized based video compression sensing algorithms are limited by long running time and low adaptability of parameters, resulting in low practicability and generalization. Taking advantage of the powerful computing power, fast speed and learnable parameters of neural networks, this paper first proposes a group sparse representation network (VGSR-Net), which maps the image block group to a higher-dimensional sparse domain through convolution, and uses a learnable threshold to denoise and extract inter-frame correlation. On this basis, a two-stage recursive enhance reconstruction network (2sRER-VGSR-Net) is proposed. First, we perform VGSR-Net to preliminarily enhance the initial reconstruction and then introduce STMC-Net as motion estimation, and the compensated frames are fed into the residual reconstruction network to further extract the missing detail and enhance the current frame. The second stage of reconstruction adopts a hybrid recursive structure with the aim of making full use of the existing better quality reconstructed frames. The simulation results show that the proposed algorithm improves the PSNR (Peak Signal to Noise Ratio) by 1.99dB compared with the existing state-of-art traditional compressed video sensing reconstruction algorithms SSIM-InterF-GSR, while improves the PSNR by 4.60dB with the comparation of the network-based algorithm CSVideoNet.

Key words: compressed video sensing; deep learning; group based sparse representation; hybrid recursive network; motion estimation; enhancement reconstruction

1 引言

压缩感知理论指出,对于可压缩或在某个变换域中可稀疏的信号,可通过适当的重构算法从少量的观测值中恢复出原信号.压缩感知重构算法的研究是压缩感知理论应用于实际的关键问题之一.

在图像压缩感知重构算法中,Gan 等人提出了分块压缩感知算法(Block-based Compressed Sensing, BCS)[2],

收稿日期: 2020-03-17;修回日期: 2020-08-20;责任编辑:覃怀银
基金项目: 广东省自然科学基金-重点(No. 2017A030311028);广东省自然科学基金(No. 2019A1515011949)

以图像块为单位对图像进行采样和重构,极大地缓解了编码端的存储和传输压力,成为图像/视频压缩感知重构算法的基本处理框架[3~5]之一. 文献[5]提出了组稀疏重构算法(Group based Sparse Representation,GSR),通过相似块组的稀疏表示,挖掘图像的局部和非局部相似性,是目前性能表现较优的传统图像压缩感知重构算法.

与图像不同,视频信号在时间维度上的冗余程度更高,充分利用视频帧间相关性成为提高视频压缩感知(Compressed Video Sensing,CVS)重构性能的关键. Kang 等人提出了分布式视频压缩感知框架(Distributed Compressed Video Sensing,DCVS)[6],将视频序列划分为图像组(Groups Of Pictures,GOP). DCVS 框架被广泛应用于传统 CVS 算法[7~11]中,具有较好的重构效果. 但传统的基于迭代优化的 CS 重构算法,运行时间较长,且参数的设置自适应性较低,在低采样率下重构效果不佳,限制了其实用性和泛化能力. 近年来,由于深度学习算法的发展,卷积网络被应用到图像压缩感知重构中[12~15]中,表现出较好的重构性能和较高的灵活性. 文献[15]提出一个非迭代的阈值收缩网络(Interpretable Shrinkage Threshold Deep Network,ISTA-Net),通过 CNN 网络实现较好的图像 CS 重构. 与图像相比,视频信号采用更低的采样率采集数据,在低采样率下兼顾重构质量和重构速度,成为 CVS 重构算法中一大难点. 文献[16]提出了一个基于 LSTM (Long Short-Term Memory)框架的 CVS 重构网络 CSVideoNet,通过 LSTM 传递帧间信息来提取整个 GOP 内时间相关性,是现有的唯一的实现 DCVS 重构的深度学习算法. 但由于缺少运动估计,该算法对存在复杂运动的视频序列不能得到很好的重构效果. 近年来,基于卷积神经网络的光流法被广泛地应用于视频信号处理领域[17~19]中,均取得较好的效果和较快的实现速度. 文献[20]利用基于光流法的运动补偿卷积网络(Spatial Transformer Motion Compensation,STMC-Net)来有效提取帧间相关性,实现了压缩视频的多帧质量增强. 本文借鉴了文献[20]的方法,利用 STMC-Net 进行运动补偿. 本文的主要创新点有:(1)提出一个视频帧间组稀疏表示网络(Video inter-frame Group Sparse Representation based on Network, VGSR-Net),引入视频帧间组稀疏的概念,将迭代自适应字典学习映射为非迭代的自适应稀疏字典,并通过可学习的阈值提取组内相似结构特征;(2)基于 VGSR-Net 提出了一个混合递归结构的两阶段递归增强重构网络(Two-Stage Recursive Enhancement Reconstruction based on VGSR-Net,2sRER-VGSR-Net),充分利用现有的已重构视频帧信息;(3)引入光流算法卷积神经网络(STMC-Net)作为运动估计网络,通过非迭代的方式获取视频帧间最佳相似块组. 基于视频帧间组稀疏表示的增强重构算法[21]的基本思路及少量仿真结果已被 ICME2020 国际会议录用,本文对所提思想做了进一步理论分析和实验证明,并提出了混合递归结构的重构网络.

2 ISTA-Net$^+$算法与 SSIM-InterF-GSR 算法

本文的 2sRER-VGSR-Net 算法是基于 ISTA-Net$^+$算法和 SSIM-InterF-GSR 算法提出的,在该小节先简单介绍这两种算法.

2.1 ISTA-Net$^+$

假设 $\boldsymbol{x}\in R^N$ 表示原信号,$\boldsymbol{\Phi}\in R^{M\times N}$ 表示观测矩阵,CS 观测值 $\boldsymbol{y}\in R^M$ 可通过 $\boldsymbol{y}=\boldsymbol{\Phi x}$ 得到. CS 采样率可定义为 M/N,其中 $M\ll N$. 当原信号 $\boldsymbol{x}$ 在某个变换域 $\boldsymbol{\Psi}$ 是稀疏的, CS 的重构问题即为求解式(1)的优化问题:

$$\min_{\boldsymbol{x}}\frac{1}{2}\|\boldsymbol{\Phi x}-\boldsymbol{y}\|_2^2+\lambda\ \|\boldsymbol{\Psi x}\|_1 \tag{1}$$

其中,λ 是正则化参数. ISTA-Net 算法[15]将迭代 ISTA 算法[22]映射为非迭代卷积网络,与 ISTA 算法类似,式(1)可以通过迭代下面两式求解:

$$\boldsymbol{r}^{(k)}=\boldsymbol{x}^{(k-1)}-\rho^{(k)}\boldsymbol{\Phi}^{\mathrm{T}}(\boldsymbol{\Phi x}^{(k-1)}-\boldsymbol{y}) \tag{2}$$

$$\boldsymbol{x}^{(k)}=\arg\min_{\boldsymbol{x}}\frac{1}{2}\|\boldsymbol{x}-\boldsymbol{r}^{(k)}\|_2^2+\lambda\ \|\boldsymbol{F}(\boldsymbol{x})\|_1 \tag{3}$$

ISTA-Net 将式(2)和式(3)的迭代映射为几个阶段,其中,k 表示阶段数,ρ 为步长,是可训练的. $F(\cdot)$ 表示由 CNN 实现的图像稀疏化表示. 文献[15]在 ISTA-Net 的基础上又提出了一个残差网络 ISTA-Net$^+$,其中 $F(\cdot)$ 被定义为 $F(\cdot)=H\circ D$. 同理,式(3)可以改写为:

$$\boldsymbol{x}^{(k)}=\arg\min_{\boldsymbol{x}}\frac{1}{2}\|\boldsymbol{x}-\boldsymbol{r}^{(k)}\|_2^2+\lambda\ \|H(D(\boldsymbol{x}))\|_1 \tag{4}$$

定义 $\tilde{H}(\cdot)$ 为 $H(\cdot)$ 的左逆,需要满足对称性约束 $\tilde{H}(\cdot)\circ H(\cdot)=\boldsymbol{I}$. ISTA-Net$^+$的损失函数定义为:

$L(\boldsymbol{\Theta})=L_{\text{discrepancy}}+\gamma L_{\text{constraint}}$,

$$\text{where}\begin{cases}L_{\text{discrepancy}}=\dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\|\hat{\boldsymbol{x}}_i-\boldsymbol{x}_i\|_2^2\\ L_{\text{constraint}}=\dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\sum_{k=1}^{N_r}\|\tilde{H}^{(k)}(H^{(k)}(\hat{\boldsymbol{x}}_i^{(k)}))-\boldsymbol{x}_i\|_2^2\end{cases} \tag{5}$$

其中,$\hat{\boldsymbol{x}}_i$,$\boldsymbol{x}_i$,N,N_r和 γ 分别表示重构的图像块、真实图像块、训练集大小、ISTA-Net$^+$的阶段数和正则化参数. $\boldsymbol{\Theta}$ 表示整个网络可训练参数集合. ISTA-Net$^+$是基于单个图像块的重构网络.

2.2 SSIM-InterF-GSR

SSIM-InterF-GSR 算法将图像组稀疏拓展到 CVS 重构中,提出了视频帧间组稀疏的概念,实现 CVS 高质量重构. 在 SSIM-InterF-GSR 中,为了给当前帧的重构提供更多的相关信息,相似块组通常包含较多的匹配块,增

加了计算复杂度和不必要的噪声. 此外,SSIM-InterF-GSR 是基于迭代的重构算法,在每次迭代中,均选用 SSIM 作为匹配准则,计算复杂度较高,运行时间较长,实用性低.

本文利用帧间组稀疏表示的高质量特性和卷积网络的计算高效性,在上述两个工作的基础上,提出了 VGSR-Net 网络,用卷积网络实现图像块组的稀疏表示.

3　两阶段递归增强视频压缩感知重构网络(2sRER-VGSR-Net)

传统 CVS 算法通过迭代获取最优相似块组,存在运行时间长、在低采样率下重构效果不佳的缺点. 针对上述问题,本文提出一种两阶段递归增强神经网络 2sRER-VGSR-Net,网络结构如图 1 所示.

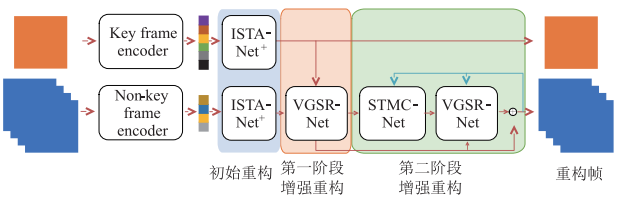

图1　2sRER-VGSR-Net算法框架

在编码端,视频序列被划分为图像组(GOP),每个 GOP 的第一帧为关键帧,采样率较高,其余帧为非关键帧,采样率相对较低. 按照 BCS 的思想,对每个视频帧独立采样. 为了满足 RIP 约束(Restricted Isometry Property)[23],本文采用随机高斯观测矩阵对信号进行采样.

在解码端,本文基于组稀疏的概念,设计了帧间组稀疏表示子网络(VGSR-Net),并基于该子网络,提出了两阶段递归增强重构网络(2sRER-VGSR-Net),包括第一阶段增强和含有运动估计的第二阶段增强.

3.1　视频帧间组稀疏表示子网络(VGSR-Net)

(1)组稀疏的概念

对于一个非关键帧,按照编码端的方式将其分成 n 个不重叠的图像块,对每个图像块构建一个视频帧间相似块组 $\boldsymbol{x}_{\mathrm{VG}_{ij}}, t=1,2,\cdots,T; j=1,2,\cdots,n$,表示第 t 帧中第 j 个图像块的相似块组. 该相似块组利用自适应学习字典 $\boldsymbol{D}_{\mathrm{VG}_{ij}}$ 可被稀疏表示为

$$\boldsymbol{x}_{\mathrm{VG}_{ij}}=\boldsymbol{D}_{\mathrm{VG}_{ij}}\boldsymbol{\alpha}_{\mathrm{VG}_{ij}} \tag{6}$$

其中,$\boldsymbol{\alpha}_{\mathrm{VG}_{ij}}$ 为稀疏系数. 因此,视频帧的重构,为求解下式的优化问题:

$$\min_{\boldsymbol{x}}\frac{1}{2}\|\boldsymbol{\Phi}\boldsymbol{D}_{\mathrm{VG}_{ij}}\boldsymbol{\alpha}_{\mathrm{VG}_{ij}}-\boldsymbol{y}\|_2^2+\lambda\|\boldsymbol{\alpha}_{\mathrm{VG}_{ij}}\|_1 \tag{7}$$

本文提出一种视频帧间组稀疏表示子网络(VGSR-Net)来解决式(7)的优化问题,通过优化卷积网络获得自适应稀疏变换字典,利用该字典将低维信号映射到更高维的空间中,使得相似块组中的相似结构特征与非相似特征区分更明显. 稀疏域中,较大的系数为相似块组的主要结构特征,并通过可学习的阈值提取主要相似特征.

受 ISTA-Net+[15] 启发,本文将稀疏字典定义为 $F(\cdot)=H\circ D$,则稀疏系数为:

$$\boldsymbol{\alpha}_{\mathrm{VG}_{ij}}=H(D(\boldsymbol{x}_{\mathrm{VG}_{ij}})) \tag{8}$$

通过的 N_p 次迭代以下两式求解式(7):

$$\boldsymbol{r}_{\mathrm{VG}_{ij}}^{(k)}=\boldsymbol{x}_{\mathrm{VG}_{ij}}^{(k-1)}-\rho^{(k)}\boldsymbol{\Phi}^{\mathrm{T}}(\boldsymbol{\Phi}\boldsymbol{x}_{\mathrm{VG}_{ij}}^{(k-1)}-\boldsymbol{y}_{ij}) \tag{9}$$

$$\boldsymbol{x}_{\mathrm{VG}_{ij}}^{(k)}=\arg\min_{\boldsymbol{x}_{\mathrm{VG}_{ij}}}\frac{1}{2}\|\boldsymbol{x}_{\mathrm{VG}_{ij}}-\boldsymbol{r}_{\mathrm{VG}_{ij}}^{(k)}\|_2^2+\lambda\|H^{(k)}(D^{(k)}(\boldsymbol{x}_{\mathrm{VG}_{ij}}))\|_0 \tag{10}$$

文献[15]中证明了式(10)可以得到如下的闭式解:

$$\boldsymbol{x}_{\mathrm{VG}_{ij}}^{(k)}=G^{(k)}(\tilde{H}^{(k)}(\mathrm{soft}(H^{(k)}(D^{(k)}(\boldsymbol{r}_{\mathrm{VG}_{ij}}^{(k)})),\theta)))+\boldsymbol{r}_{\mathrm{VG}_{ij}}^{(k)} \tag{11}$$

其中,θ 为可学习的阈值.

(2)组稀疏表示的卷积神经网络实现(VGSR-Net)

本文对 ISTA-Net+ 进行了扩展,提出了基于全局相关信息的视频帧间相似块组稀疏表示卷积网络(VGSR-Net),结构如图 2 所示.

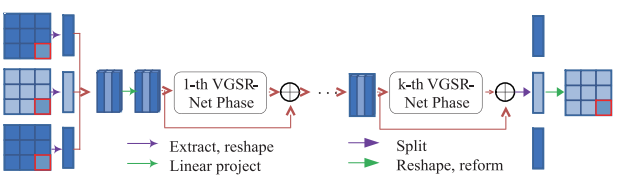

图2　VGSR-Net框架

在每个阶段中,利用卷积 $H,\tilde{H},D$ 和 G 实现自适应稀疏变换,卷积层的设置如表 1 所示. 针对稀疏系数 $\boldsymbol{\alpha}_{\mathrm{VG}_{ij}}$,通过软阈值滤波得到组内主要的相似特征 $\hat{\boldsymbol{\alpha}}_{\mathrm{VG}_{ij}}$.

表 1　VGSR-Net 各个阶段的卷积层参数设置

操作	D	H	$\tilde{H}$	G
层数	1	2	2	1
卷积核大小	3×3	3×3	3×3	3×3
卷积核个数	32	32	32	n_f
激活函数	无	ReLU	ReLU	无

VGSR-Net 以优化整个 GOP 为目标,损失函数表示为:

$$L(\boldsymbol{\Theta}_{\mathrm{VGSR}})=\sum_{t=1}^{T-1}L_{\mathrm{discrepancy}_t}+\gamma L_{\mathrm{constraint}_t} \tag{12}$$

其中,

$$L_{\mathrm{discrepancy}_t}=\frac{1}{N}\sum_{i=1}^{N}\|\hat{\boldsymbol{x}}_{ti}-\boldsymbol{x}_{ti}\|_2^2 \tag{13}$$

$$L_{\mathrm{constraint}_t}=\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N_p}\|\tilde{H}^{(k)}(H^{(k)}(\hat{\boldsymbol{x}}_{ti}^{(k)}))-\boldsymbol{x}_{ti}\|_2^2 \tag{14}$$

其中,$t=1,2,\cdots,T$,T 为整个 GOP 的大小,N 为批尺寸

(batch size)的大小. $\hat{\boldsymbol{x}}_{ti}$表示第 i 个训练样本中第 t 帧的重构结果,而 $\boldsymbol{x}_{ti}$表示真实的视频帧. $\boldsymbol{\Theta}_{\mathrm{VGSR}}$表示整个 VGSR-Net 可训练的参数集合.

3.2 两阶段递归增强重构网络(2sRER-VGSR-Net)

基于 VGSR-Net,本文提出两阶段递归增强网络(2sRER-VGSR-Net)实现视频的压缩感知重构. 重构过程主要包括两个部分:利用单帧信息的初始重构和利用帧间相关性的针对非关键帧的两阶段增强重构. 由于关键帧具有较高的采样率,具有较好的初始重构结果,初始重构即为最终的重构结果,并用其为非关键帧的重构提供高质量的相关信息. 帧间重构网络分两步对非关键帧进行增强重构,包含了两个 VGSR-Net 网络运动补偿网络 STMC-Net,整体框架如图 1 所示.

3.2.1 第一阶段增强重构网络的意义及实现

非关键帧的采样率较低,初始重构质量较差. 本文所提网络首先通过 VGSR-Net,利用重构质量较高关键帧作为参考帧,对初始重构结果进行第一次增强,为下一步的运动补偿提供更准确的信息. 在第一阶段增强重构中,选用与当前帧最相邻的关键帧作为参考帧. 损失函数与 VGSR-Net 一致,即

$$L_1 = L_{\mathrm{VGSR}} \tag{15}$$

3.2.2 第二阶段增强重构网络

视频信号由于运动的存在,视频帧间存在差异. 第一阶段增强重构实际上是在未对齐的关键帧的对应位置上提取匹配块,在当前帧与关键帧距离较远时,当前帧的重构质量提升并不明显. 因此,第二阶段增强重构网络首先引入了运动补偿网络提高组内图像块的相似性,再通过残差重构网络来挖掘当前帧丢失的信息.

(1)运动估计网络及参考帧的选择

(a)运动估计网络

Lucas-Kanade 金字塔光流法是常用的传统光流计算方法. 基于亮度恒定和相邻像素的运动相似两个假设建立目标方程,在多尺度上计算光流. 由于目标方程较为复杂,且需要通过多次迭代求解,实时性较差. 为了提高运算速度,本文引入 STMC-Net 网络作为运动补偿网络. 类似地,STMC-Net 采用金字塔结构,卷积网络的设置与文献[20]一致.

(b) 参考帧的选择

为了尽可能地利用已有的高质量重构帧,本文在第二阶段增强网络中提出了一种混合递归网络.

如图 3 所示,非关键帧采用从两端向中间靠拢的重构顺序. 对于前半个 GOP,采用向前递归网络. 实线表示把第二阶段增强重构结果作为运动补偿对象,虚线表示将第一阶段增强结果运动补偿对象. 反之,后半个 GOP 采用向后递归,中间帧采用双向递归. 训练时,本文用真实的视频帧的运动补偿帧 $\boldsymbol{x}_{t-1}^{\mathrm{MC}}$ 和 $\boldsymbol{x}_{t+1}^{\mathrm{MC}}$ 来计算损失函数,损失函数为:

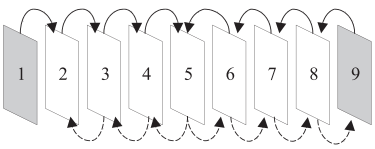

图3 重构顺序及参考帧的选择

$$L_{\mathrm{MC}}(\boldsymbol{\Theta}_{\mathrm{MC}}) = \frac{1}{N}\sum_{t=1}^{T-1} \| \boldsymbol{x}_{t-1}^{\mathrm{MC}} - \boldsymbol{x}_t \|_2^2 + \| \boldsymbol{x}_{t+1}^{\mathrm{MC}} - \boldsymbol{x}_t \|_2^2 \tag{16}$$

(2)残差重构网络

第二阶段增强重构中采用残差网络结构,从参考帧中提取当前帧丢失的细节信息. 本文选用两个运动补偿帧、与当前帧相邻的两个重构帧作为参考帧,通过 VGSR-Net 重构出图像残差 $\boldsymbol{d}_i$. 最后,由式(17)得到二次增强图像 $\hat{\hat{\boldsymbol{x}}}_t$.

$$\hat{\hat{\boldsymbol{x}}}_t = \hat{\boldsymbol{x}}_t + \boldsymbol{d}_i \tag{17}$$

同理,残差重构网络的损失函数与 VGSR-Net 网络的损失函数一致,即

$$L_2 = L_{\mathrm{VGSR}} \tag{18}$$

3.2.3 损失函数

2sRER-VGSR-Net 以视频帧的观测值 $\{\boldsymbol{y}_t\}_{t=1}^{T+1}$ 作为输入,输出一系列重构帧 $\{\hat{\hat{\boldsymbol{x}}}_t\}_{t=1}^{T}$. 给定真实视频帧 $\{\boldsymbol{x}_t\}_{t=1}^{T}$,本文利用 MSE(Mean Square Error)形式的损失函数对网络进行优化. 特别地,本文采用 ISTA-Net$^+$ 对视频帧进行独立的初始重构. 本文主要讨论非关键帧的帧间 CVS 重构,帧间重构网络的损失函数为:

$$L(\boldsymbol{\Theta}_{\mathrm{inter}}) = L_1 + L_2 + \beta L_{\mathrm{MC}} \tag{19}$$

其中,β 为一常数,本文实验中,$\beta = 0.01$,L_1、L_2 权重相同.

4 仿真结果及分析

与 CSVideoNet[16] 相同,本文使用 UCF-101 数据集训练网络,提取 499760 个 GOP 作为训练集和测试集. 每帧图像从中心提取 96×96 大小的图像,仅保留亮度通道,每个图像分成大小为 32×32 的不重叠块. VGSR-Net 的阶段数 N_p 设置为 6,卷积层的设置参数如表 1 所示. 使用 Adam[24] 方法训练网络,训练率为 0.0001. 在本文中,将预训练好的初始重构网络固定下来,不再作进一步的训练和优化. 为了加速收敛速度,我们首先按顺序训练每个子网络,最后将所有子网络组合起来进行联合训练.

4.1 与传统 CVS 重构算法结果对比及分析

本小节将 2sRER-VGSR-Net 算法与几个开源的传统 DCVS 重构算法进行对比,即 MH-BCS-SPL[7]、RRS[8] 和 SSIM-InterF-GSR[9]. 选取 25YUV 数据集中 QCIF 格

式的 hall、mother-daughter、football、soccer、tennis、ice、stefan 和 mobile 的前 96 帧作为测试序列. 关键帧采样率为 0.7,非关键帧采样率分别为 0.1、0.05、0.01,GOP 大小为 8. 用峰值信噪比(Peak Signal to Noise Ratio,PSNR)和 SSIM 作为重构视频帧质量评判标准. 各采样率下 4 种算法的重构质量如表 2 所示.

表 2　各采样率下算法重构质量对比

算法	hall	mother-daughter	football	soccer	tennis	ice	stefan	mobile
key subrate = 0.7, non-key subrate = 0.1, GOP = 8, PSNR(dB)/SSIM								
MH-BCS-SPL[7]	32.80/0.95	37.73/0.95	26.70/0.71	29.45/0.82	27.43/0.74	30.00/0.90	22.24/0.69	22.25/0.69
RRS[8]	29.06/0.92	39.97/0.97	26.97/0.71	30.84/0.85	27.19/0.75	32.40/0.95	20.56/0.64	20.56/0.56
SSIM-InterF-GSR[9]	36.16/0.97	39.73/0.96	27.73/0.76	30.90/**0.88**	27.42/0.77	32.43/**0.96**	23.03/0.70	25.50/0.81
2sRER-VGSR-Net	**38.28/0.98**	**40.42/0.97**	**28.72/0.79**	**32.46**/0.86	**28.96/0.78**	**35.21/0.96**	**24.15/0.81**	**26.87/0.88**
key subrate = 0.7, non-key subrate = 0.05, GOP = 8, PSNR(dB)/SSIM								
MH-BCS-SPL	30.95/0.93	35.37/0.92	25.11/0.63	27.40/0.75	26.16/0.69	28.26/0.87	21.01/0.60	20.63/0.57
RRS	22.47/0.79	32.73/0.88	24.11/0.62	27.37/0.75	22.67/0.62	26.67/0.87	20.03/0.54	19.37/0.43
SSIM-InterF-GSR	34.62/0.97	38.07/0.95	26.10/0.68	28.15/0.80	25.44/0.69	29.44/0.92	22.12/0.60	23.90/0.74
2sRER-VGSR-Net	**37.10/0.98**	**39.63/0.97**	**27.03/0.71**	**30.13/0.81**	**27.75/0.74**	**33.07/0.95**	**23.35/0.76**	**25.80/0.85**
key subrate = 0.7, non-key subrate = 0.01, GOP = 8, PSNR(dB)/SSIM								
MH-BCS-SPL	24.64/0.80	29.50/0.78	22.35/0.51	23.50/0.60	23.26/0.58	23.80/0.76	19.23/0.44	18.13/0.35
RRS	18.39/0.48	24.40/0.63	21.74/0.49	21.98/0.55	20.70/0.50	21.40/0.65	16.78/0.37	15.66/0.25
SSIM-InterF-GSR	29.41/0.93	31.48/0.86	23.49/0.55	23.98/0.63	23.51/0.59	25.25/0.83	21.01/0.45	22.65/0.65
2sRER-VGSR-Net	**33.86/0.96**	**37.33/0.95**	**24.15/0.56**	**25.55/0.66**	**25.04/0.65**	**28.43/0.88**	**22.00/0.65**	**24.08/0.80**

由表 2 可见,2sRER-VGSR-Net 在各采样率下与其他传统的 DCVS 重构算法相比,重构性能均有较大的提升. 在非关键帧采样率为 0.1、0.05、0.01 下,ice 序列分别比传统的性能较好的 SSIM-InterF-GSR 算法提高了 2.78dB、3.63dB、3.18dB. 图 4 展示了 Soccer 序列第 42 帧在 0.05 采样率下,各算法的重构视觉效果. MH-BCS-SPL 存在严重的块效应且出现许多的噪点;RRS 出现了过度平滑的情况,运动块(人物)信息几乎丢失. SSIM-InterF-GSR 能重构出图像的主要信息,但是对于图中运动较快的部分(图中红色框部分),重构质量较差. 相比

(*a*)原图　(*b*)MH-BCS-SPL(26.09dB)　(*c*)RRS(23.09dB)

(*d*)SSIM-InterF-GSR(26.57dB)　(*e*)2sRER-VGSR-Net(30.17dB)

图4　soccer序列第42帧重构效果图

之下,2sRER-VGSR-Net 重构出来的图像较为清晰,运动部分和细节部分(如草坪)均有较好的重构效果.

与传统的重构算法不同,2RER-VGSR-Net 首先对低质的初始重构结果进行了初步的增强,降低了错误匹配的可能性,使算法在低采样率下仍然具有较好的重构效果. 第二阶段引入了全局范围的运动估计,提高了参考帧的相似性,更有利于帧间相关信息的提取.

为了对比这几种算法的运行速度,表 3 列出了各算法的运行时间,其中,传统的基于迭代的重构算法仿真在配置了 CPU Intel Core i5 3.20GHz 的台式计算机运行,2sRER-VGSR-Net 在搭载了 CPU Intel Xeon Silver 4210 2.20Hz 和 GPU 2080ti 的服务器上运行. 由表 3 可见,2sRER-VGSR-Net 与传统的基于迭代的算法相比,运行时间明显缩短,更具实时性.

表 3 各采样率下算法运行时间(s)

算法	hall	mother-daughter	football	soccer	tennis	ice	stefan	mobile
key subrate = 0.7, non-key subrate = 0.1, GOP = 8								
MH-BCS-SPL[7]	38.3	38.3	35.8	25.8	39.9	40.1	41.8	35.5
RRS[8]	150	161	156	175	163	152	159	165
SSIM-InterF-GSR[9]	95.4	99.8	128	129	128	113	127	133
2sRER-VGSR-Net	**0.03**	**0.03**	**0.03**	**0.03**	**0.03**	**0.03**	**0.03**	**0.03**
key subrate = 0.7, non-key subrate = 0.05, GOP = 8								
MH-BCS-SPL	36.7	38.0	35.8	37.0	35.9	37.4	42.8	35.4
RRS	168	164	175	168	172	171	179	178
SSIM-InterF-GSR	102	116	124	124	140	126	135	137
2sRER-VGSR-Net	**0.03**	**0.03**	**0.03**	**0.03**	**0.03**	**0.03**	**0.03**	**0.03**
key subrate = 0.7, non-key subrate = 0.01, GOP = 8								
MH-BCS-SPL	40.4	38.5	37.0	38.8	38.5	40.4	35.4	35.4
RRS	223	213	220	211	215	224	226	229
SSIM-InterF-GSR	99	103	118	112	128	101	136	132
2sRER-VGSR-Net	**0.03**	**0.03**	**0.03**	**0.03**	**0.03**	**0.03**	**0.03**	**0.03**

4.2 与基于深度学习的 CVS 重构算法结果对比及分析

本小节将 2sRER-Video-Net 与 CSVideoNet[16] 对比,测试集设置与 CSVideoNet 一致. CSVideoNet 的实验结果均取自原文献,关键帧采样率为 0.2,非关键帧采样率分别为 0.037,0.018,0.009. 各采样率下的重构质量如表 4 所示.

由表 4 可见,2sRER-VGSR-Net 相比于 CSVideoNet 有明显的性能提升,在各采样率下 PSNR 有 4dB 左右的提升,而 SSIM 也有 0.08 ~ 0.09 的提升. CVS 的解码过程是一个欠定性问题,与 CSVideoNet 相比,2sRER-VGSR-Net 每个阶段都引入了观测值 $\boldsymbol{y}_t$ 限制其解空间,使重构图像与真实图像不会产生较大的偏差. 此外,2sRER-VGSR-Net 加入了运动估计,可以更充分利用帧间相关信息.

为验证本文所提算法在其他测试集下的泛化能力,表 5 展示了 2sRER-VGSR-Net 和 2sER-VGSR-Net[21] 在几个常用的测试集 UCF-101、VID4 和 DAVIS 上的重构性能. 其中,从 DAVIS 数据集中挑选 10 个视频序列组成测试集,记为 DAVIS-10. 从每个视频帧的中心提取大小为 160 × 160 的图像,仅保留亮度分量.

表 4 各采样率下算法重构质量对比(PSNR(dB)/SSIM)

非关键采样率	CSVideoNet[16]	2sRER-VGSR-Net
0.037	26.87/0.81	**31.52/0.89**
0.018	25.09/0.77	**29.87/0.86**
0.009	24.23/0.74	**28.60/0.83**

表 5 2sRER-VGSR-Net 和 2sER-VGSR-Net 算法在不同测试集下重构质量对比(PSNR(dB)/SSIM)

非关键采样率	2sER-VGSR-Net[21]	2sRER-VGSR-Net
UCF-101		
0.037	31.23/**0.89**	**31.52/0.89**
0.018	29.49/0.85	**29.87/0.86**
0.009	28.29/0.82	**28.60/0.83**
VID4		
0.037	25.21/0.69	**25.62/0.72**
0.018	24.31/0.63	**24.48/0.65**
0.009	23.51/0.58	**23.59/0.60**
DAVIS-10		
0.037	28.32/0.80	**28.70/0.82**
0.018	26.19/0.73	**26.44/0.74**
0.009	24.46/0.67	**24.74/0.68**

从表5可以看出,2sRER-VGSR-Net 在其他数据集仍有较好的重构性能,平均 PSNR 和 SSIM 均高于 2sER-VGSR-Net. 2sRER-VGSR-Net 的混合递归结构,可以充分地利用质量较好的已重构好的视频帧,减少了噪声的引入.

4.3　运动估计和第二阶段增强对重构质量的改善

为了验证 2sRER-VGSR-Net 框架的有效性,本文设计了一组对比实验,证明运动估计和第二阶段增强的必要性. 将只进行一个阶段的增强重构网络记为 VGSR-Net,而 2sRER-nomc 表示第二阶段重构中去掉运动估计,仅用与当前帧相邻的两个帧作为参考帧进行组稀疏重构. 实验设置与第 4.1 小节一致,三种算法的重构结果如表6所示. 从表中数据可见,2sRER-VGSR-Net 比 VGSR-Net 和 2sRER-nomc 均有较大的提升,其中运动序列 football、soccer、tennis、ice、stefan 和 mobile 提升更为明显. 可见,在快速运动序列中,运动估计对重构质量的提升尤为重要. 运动估计使相邻帧中的有效信息更好地参与到当前帧的重构中,提高了视频帧间图像块组的相似性,从而提高重构质量.

表6　VGSR-Net 和 2sRER-nomc 较 2sRER-VGSR-Net 的 PSNR 对比情况

算法	hall	mother-daughter	football	soccer	tennis	ice	stefan	mobile
key subrate = 0.7, non-key subrate = 0.1, GOP = 8, PSNR(dB)								
2sRER-VGSR-Net	38.28	40.42	28.72	32.46	28.96	35.21	24.15	26.87
2sRER-nomc	−0.61	−0.33	−1.81	−1.81	−1.76	−1.82	−1.86	−0.85
VGSR-Net	−1.01	−0.42	−0.86	−0.8	−1.07	−1.45	−1.04	−1.11
key subrate = 0.7, non-key subrate = 0.05, GOP = 8, PSNR(dB)								
2sRER-VGSR-Net	37.10	39.63	27.03	30.13	27.75	33.07	23.35	25.80
2sRER-nomc	−0.44	−0.28	−1.25	−1.57	−1.25	−1.4	−0.99	−0.79
VGSR-Net	−1.53	−0.89	−1.08	−1.26	−1.36	−1.81	−1.07	−1.3
key subrate = 0.7, non-key subrate = 0.01, GOP = 8, PSNR(dB)								
2sRER-VGSR-Net	33.86	37.33	24.15	25.55	25.04	28.43	22.00	24.08
2sRER-nomc	−0.38	−0.23	−0.9	−1.91	−0.68	−1.66	−0.26	−0.44
VGSR-Net	−1.53	−1.04	−0.77	−1.29	−0.89	−1.95	−0.51	−0.98

5　结论

本文利用组稀疏的概念,提出了视频帧间组稀疏网络(VGSR-Net),在高维空间中获得更加合理的图像块组的稀疏表示来提取视频帧间相关性. 在此基础上,本文提出了两阶段递归增强重构网络(2sRER-VGSR-Net),分两步实现 CVS 重构:(1)为了为运动估计提供更准确的信息,在初始重构的基础上,首先对非关键帧进行初步增强重构;(2)通过卷积网络实现运动估计,提高图像块组的相似性,并在运动补偿帧和第一阶段增强重构结果的基础上进行残差重构,对非关键帧进行进一步增强重构. 此外,为了充分利用已有的重构信息,2sRER-VGSR-Net 采用混合递归结构,尽可能为当前帧提供质量较好的参考帧以减少噪声的引入. 仿真结果表明,所提算法和其他传统的 DCVS 重构算法相比具有较为明显的质量提升;与现有的唯一的基于深度网络的 DCVS 重构算法 CSVideoNet 相比,所提算法也有 4dB 左右的 PSNR 提升.

参考文献

[1] Donoho D L. Compressed sensing[J]. IEEE Transactions on Information Theory,2006,52(4):1289 − 1306.

[2] Gan L. Block compressed sensing of natural images[A]. 15th International Conference on Digital Signal Processing [C]. Cardiff,UK:IEEE,2007. 403 − 406

[3] Mun S,Fowler J E. Block compressed sensing of images using directional transforms[A]. 16th IEEE International Conference on Image Processing (ICIP)[C]. Cairo,Egypt:IEEE,2009. 3021 − 3024.

[4] Chen R,Tong Y,Yang J,et al. Compressed video sensing with multi-hypothesis prediction[A]. International Conference on Emerging Internetworking,Data & Web Technologies[C]. China:Springer,2017. 489 − 496.

[5] Zhang J,Zhao D,Gao W. Group-based sparse representation for image restoration[J]. IEEE Transactions on Image Processing,2014,23(8):3336 − 3351.

[6] Kang L,Lu C. Distributed compressive video sensing[A]. International Conference on Acoustics,Speech and Signal Processing[C]. Taipei,China:IEEE,2009. 1169 − 1172.

[7] Tramel E W,Fowler J E. Video Compressed Sensing with Multihypothesis [A]. Data Compression Conference (DCC)[C]. Snowbird,USA:IEEE,2011. 193 − 202.

[8] Zhao C,Ma S,Zhang J,et al. Video compressive sensing reconstruction via reweighted residual sparsity[J]. IEEE Transactions on Circuits and Systems for Video Technology,2017,27(6):1182 − 1195.

[9] 和志杰,杨春玲,汤瑞东. 视频压缩感知中基于结构相似的帧间组稀疏表示重构算法研究[J]. 电子学报,2018,46(3):544 − 553.

HE Zhi-jie, YANG Chun-ling, TANG Rui-dong. Research on structural similarity based inter-frame group sparse representation for compressed video sensing[J]. Acta Electronica Sinica, 2018, 46(3): 544 – 553. (in Chinese)

[10] 杨春玲,郑学炜. CVS 中基于多维度参考帧的双稀疏重构算法[J]. 华南理工大学学报(自然科学版),2018,46(08):1 – 10.
YANG Chun-ling, ZHENG Xue-wei. Dual sparse reconstruction algorithm based on multi-dimensional reference frames in CVS [J]. Journal of South China University of Technology (Natural Science Edition), 2018, 46(08): 1 – 10. (in Chinese)

[11] 汤瑞东,杨春玲,褚韵怡. 视频压缩感知多假设局部增强重构算法[J/OL]. 自动化学报, https://doi.org/10.16383/j.aas.c190408, 2019-11-15.
TANG Rui-dong, YANG Chun-ling, XUAN Yun-Yi. Localenhancement recovery algorithm based on multi-hypothesis prediction in compressed video sensing [J/OL]. Acta Automatica Sinica, https://doi.org/10.16383/j.aas.c190408. (in Chinese)

[12] Kulkarni K, Lohit S, Turaga P, et al. Reconnet: Non-iterative reconstruction of images from compressively sensed measurements[A]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. Las Vegas, USA: IEEE, 2016. 449 – 458.

[13] Shi W, Jiang F, Zhang S, Zhao D. Deep networks for compressed image sensing[A]. IEEE International Conference on Multimedia and Expo (ICME)[C]. Hong Kong, China: IEEE, 2017. 877 – 882.

[14] 练秋生,富利鹏,陈书贞,石保顺. 基于多尺度残差网络的压缩感知重构算法[J]. 自动化学报,2019,45(11):2082 – 2091.
LIAN Qiu-sheng, FU Li-peng, CHEN Shu-zhen, SHI Baoshun. A compressed sensing algorithm based on multi-scale residual reconstruction network[J]. Acta Automatica Sinica, 2019, 45(11): 2082 – 2091. (in Chinese)

[15] Zhang J, Ghanem B. ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing [A]. IEEE/CVF Conference on Computer Vision and Pattern Recognition [C]. Salt Lake City, USA: IEEE, 2018. 1828 – 1837.

[16] Xu K, Ren F. CSVideoNet: A real-time end-to-end learning framework for high-frame-rate video compressive sensing[A]. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV)[C]. Lake Tahoe, USA: IEEE, 2018. 1680 – 1688.

[17] M S M Sajjadi, R Vemulapalli, M Brown. Frame-recurrent video super-resolution[A]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)[C]. Salt Lake City, USA: IEEE, 2018. 6626 – 6634.

[18] Gu D, Wen Z, Cui W, Wang R, Jiang F, Liu S. Continuous bidirectional optical flow for video frame sequence interpolation[A]. 2019 IEEE International Conference on Multimedia and Expo (ICME)[C]. Shanghai, China: IEEE, 2019. 1768 – 1773.

[19] Kim T H, Sajjadi M S M, Hirsch M, Scholkopf B. Spatio-temporal transformer network for video restoration[A]. The European Conference on Computer Vision (ECCV)[C]. Munich, Germany: Springer, 2018. 106 – 122.

[20] Yang R, Xu M, Wang Z, Li T. Multi-frame quality enhancement for compressed video[A]. Conference on Computer Vision and Pattern Recognition (CVPR)[C]. Salt Lake City, USA: IEEE, 2018. 6664 – 6673.

[21] Xuan Y, Yang C. 2sER-VGSR-Net: A two-stage enhancement reconstruction based on video group sparse representation network for compressed video sensing[A]. 2020 IEEE International Conference on Multimedia and Expo (ICME)[C]. London, United Kingdom: IEEE, 2020. 1 – 6.

[22] M V Afonso, J M Bioucas-Dias, M A T Figueiredo. Fast image recovery using variable splitting and constrained optimization[J]. IEEETransactions on Image Processing, 2010, 19(9): 2345 – 2356.

[23] E J Candes. The restricted isometry property and its implications for compressed sensing [J]. Comptes Rendus Mathematique, 2008, 346(9-10): 589 – 592.

[24] D P Kingma, J Ba Adam: A Method for Stochastic Optimization[DB/OL]. https://arxiv.org/abs/1412.6980, 2017-01-30.

作者简介

褚韵怡 女,1995 年生于广东佛山,华南理工大学电子与信息学院研究生. 研究方向:视频压缩感知.
E-mail: eeyyxuan@mail.scut.edu.cn

杨春玲(通信作者) 女,1970 年生于河南新乡,华南理工大学电子与信息学院博士生导师. 研究方向:图像/视频压缩编码,图像质量评价.
E-mail: eeclyang@scut.edu.cn