基于层次特征复用的视频超分辨率重建

周 圆 [1] 王明非 [1] 杜晓婷 [1] 陈艳芳 [1]

摘 要 当前的深度卷积神经网络方法, 在视频超分辨率任务上实现的性能提升相对于图像超分辨率任务略低, 部分原因是它们对层次结构特征中的某些关键帧间信息的利用不够充分. 为此, 提出一个称作层次特征复用网络 (Hierarchical feature reuse network, HFRNet) 的结构, 用以解决上述问题. 该网络保留运动补偿帧的低频内容, 并采用密集层次特征块 (Dense hierarchical feature block, DHFB) 自适应地融合其内部每个残差块的特征, 之后用长距离特征复用融合多个 DHFB 间的特征, 从而促进高频细节信息的恢复. 实验结果表明, 提出的方法在定量和定性指标上均优于当前的方法.

关键词 层次特征复用, 卷积神经网络, 特征融合, 视频超分辨率重建

引用格式 周圆, 王明非, 杜晓婷, 陈艳芳. 基于层次特征复用的视频超分辨率重建. 自动化学报, 2024, **50**(9): 1736−1746

DOI 10.16383/j.aas.c210095

Video Super-resolution via Hierarchical Feature Reuse

ZHOU Yuan[1] WANG Ming-Fei[1] DU Xiao-Ting[1] CHEN Yan-Fang[1]

Abstract The performance improvement of current deep convolution neural network methods in video super-resolution task is slightly lower than that in image super-resolution task, partly because they do not make full use of some key inter-frame information in hierarchical structure features. In this paper, we propose hierarchical feature reuse network (HFRNet) to solve the problem mentioned above. The network retains the low-frequency content of the motion compensation frame, and use dense hierarchical feature block (DHFB) to adaptively fuse the features of each residual block within it, then long-term feature reuse is proposed to fuse the features between multiple dense hierarchical feature block, so as to promote the recovery of high-frequency detail information. Experimental results show that the proposed method is superior to the current method in both quantitative and qualitative metrics.

Key words Hierarchical feature reuse, convolutional neural network (CNN), feature fusion, video super-resolution

Citation Zhou Yuan, Wang Ming-Fei, Du Xiao-Ting, Chen Yan-Fang. Video super-resolution via hierarchical feature reuse. *Acta Automatica Sinica*, 2024, **50**(9): 1736−1746

视频超分辨率重建 (Super-resolution, SR) 旨在从低分辨率视频帧 (Low resolution frames, LR) 中恢复出对应的高分辨率视频帧 (High resolution frames, HR), 并广泛用于各种任务. 然而, 视频帧间的时空信息复杂性是视频超分辨率重建任务所面临的一个严峻挑战.

视频超分辨率重建任务有以下几大类方法: 传统的基于重建的方法、基于示例的方法和基于深度学习的方法.

传统的基于重建的视频超分辨率方法使用贝叶斯模型对病态的超分辨率问题进行正则化[1]. 然而, 这些基于重建的方法虽然可以在一定程度上保证性能, 但在较大倍数的超分辨率重建和较大尺度运动的场景下表现欠佳. 基于示例的视频超分辨率重建方法可以利用视频帧之间的自相似性完成重建[2−3]. 然而, 基于示例的方法使用许多重叠的图像小块, 会占用相当多的存储与计算资源. 基于深度学习的方法通过卷积神经网络 (Convalutional neural network, CNN) 学习视频帧细节特征来重建高分辨率的视频, 取得了比传统方法更好的结果. 在这些深度学习方法中, 其中一些方法直接以低分辨率视频帧作为网络输入[4−5]; 而另一些深度学习方法是先对视频帧进行运动补偿, 然后再将运动补偿的结果作为网络输入[6−7]. 随着网络深度的增加, 其中的每个卷积层能提取具有不同层级的特征. 然而, 这些深度学习方法的卷积神经网络多是简单堆叠而成, 忽略了在重建过程中对来自不同层次的深层网络特征进行充分的整合利用.

为解决视频超分辨率重建过程中层次特征利用不充分的问题, 本文提出用于视频超分辨率重建的

收稿日期 2021-01-28 录用日期 2021-05-12
Manuscript received January 28, 2021; accepted May 12, 2021
国家自然科学基金联合基金项目 (U2006211), 国家重点研发计划 (2020YFC1523204) 资助
Supported by National Natural Science Foundation of China (U2006211) and National Key Research and Development Program of China (2020YFC1523204)
本文责任编委 黄华
Recommended by Associate Editor HUANG Hua
1. 天津大学电气自动化与信息工程学院 天津 300072
1. School of Electrical and Information Engineering, Tianjin University, Tianjin 300072

层次特征复用网络 (Hierarchical feature reuse network, HFRNet). 它通过运动补偿的预处理方法对齐多个视频帧的光流, 并充分利用深度神经网络中来自不同层的特征. 本文采用密集层次特征块 (Dense hierarchical feature block, DHFB) 作为网络的基本构建模块, 其内部包含几个残差块和短距离特征复用. DHFB 通过自适应地保留内部残差块的信息来提取、融合短距离密集特征. 在使用多个 DHFB 提取多个短期密集特征后, 本文使用长距离特征复用自适应融合所有 DHFB 的特征. 同时, 为探究 DHFB 间不同特征复用方式的影响, 本文提出两种网络结构. 广泛的实验证明 HFRNet 优于许多视频超分辨率重建算法.

1 相关工作

1.1 视频超分辨率重建

视频超分辨率任务以一个低分辨率的视频序列作为输入, 重建对应的高分辨率序列. 作为图像超分辨率重建任务[8−13] 的推广, 视频超分辨率重建方法可分为以下几大类: 传统的基于重建的方法、基于示例的方法和基于深度学习的方法.

基于重建的视频超分辨率方法利用视频中的运动信息, 在较小的全局运动下, 生成图像的保真度较高. 许多基于重建的方法都集中在使用贝叶斯框架重建高分辨率图像, 例如 Liu 等[1] 的贝叶斯方法可同时估计超分辨率重建图像、物体运动的场和对应的模型参数. 但是, 这类基于核回归方法的运算成本很高.

基于示例的视频超分辨率方法主要利用耦合字典或视频帧内的自相似性来学习从低分辨率到高分辨率视频帧的非线性映射. Shahar 等[2] 假设视频或图像块遵循多尺度关系, 并从较粗尺度的块中恢复给定尺度的高分辨率图像块. 然而, 这类方法使用许多重叠的图像小块, 会占用相当大的存储与计算资源.

上述方法提取的特征依赖于先验知识, 但依赖先验知识设计的特征并不丰富, 难以适应各种场景的视频. 近年来, 基于深度学习的方法在各领域取得了重大进展, 如语义分割[14]、显著性检测[15−16]. 深度学习方法使用到的神经网络可以针对各种场景的视频自适应地提取更丰富的特征, 从而推动视频超分辨率重建的发展, 改善重建效果[17−18]. Kappeler 等[6] 提出用于视频超分辨率重建的视频超分辨率重建网络 (Video super resolution network, VSRNet), 后由 Li 等[7] 扩展为添加运动补偿的残差视频超分辨率重建网络 (Motion compensation residual network, MCResNet), 它在 VSRNet 输入与输出框架之间添加了跳跃连接. 同时, 上述方法在输入 CNN 前, 使用局部−全局结合的全变分算法 (Combined local and global total variation, CLG-TV)[19] 对视频帧进行运动补偿, 可进一步提高视频超分辨率的精度. Caballero 等[4] 提出一种时空子像素卷积网络 (Video efficient sub pixel convolutional neural network, VESPCN), 可联合训练运动补偿和视频超分辨率重建. 在此基础上, Tao 等[5] 提出亚像素运动补偿 (Sub pixel motion compensation, SPMC), 并利用编解码器网络进行特征融合. 此外, 生成对抗网络也得以用于视频超分辨率重建中[20]. 这类网络通过使用对抗损失和感知内容这两个目标函数来改善重建视频质量, 提升重建效果.

1.2 卷积神经网络的设计

近年来, 卷积神经网络已经迅速解决了许多具有挑战性的视觉问题, 例如人群计数[21]、图像显著性检测[22−24]. Szegedy 等[25] 的研究表明, 更深的网络结构可学习到更复杂的映射, 提取到更多层次的特征, 从而提升性能. 但是, 训练也会变得更加困难. 为了有效地训练更深层的网络架构, He 等[26] 的残差神经网络 (Residual network, ResNet) 在各层之间使用跳跃连接来加速训练过程. 同时, 为了更好地利用深层网络中丰富的特征图, Huang 等[27] 提出的密集连接网络 (Densely connected convolutional neural network, DenseNet) 使用密集连接, 将当前层前面所有层的输出拼接起来作为当前层的输入来丰富不同层次的特征. 所有这些网络都具有一个关键思想, 即必须在各层之间建立许多跳跃连接以有效地训练非常深的网络. 跳跃连接与密集连接的思想在超分辨率领域也有重要应用. Zhang 等[28] 将跳跃连接与密集连接相结合, 提出残差密集网络 (Residual dense network, RDN) 实现图像深度特征的多级融合, 提升静态图像的高质量超分辨率重建. Zhou 等[29] 设计的网络也利用了不同尺度的深浅层特征完成特征融合.

1.3 视频超分辨率中的时域信息融合模式

在视频超分辨率任务中, 时域信息的融合模式对超分辨率性能起着重要作用. Yi 等[30] 提出密集记忆模块 (Ultra-dense memory block, UDMB), 将卷积−长短时记忆嵌入密集残差模块中建模视频时空特征, 从而建立高性能视频超分辨率模型. Yi 等[31] 提出基于逐步融合的视频超分辨率重建网络 (Progressive fusion network, PFN), 通过非局部运算有效地融合来自不同时间的视频帧的特征. Yi 等[32] 提出基于 PFN 框架的生成对抗模型, 综合利用 PFN

的时域特征融合能力和生成对抗的视觉感知评价能力, 从而使得生成器输出更加真实的高分辨率图像.

2 层次特征复用网络 HFRNet

本节首先描述层次特征复用网络 (HFRNet) 的整体结构和其内部每一个组件的结构, 其次, 在描述 HFRNet 的长距离特征复用的基础上, 提出两种可能的网络结构, 以探讨 DHFB 间不同的特征复用方式对 HFRNet 的影响.

HFRNet 是以连续的低分辨率视频帧 $\{I^{\mathrm{L}}_{-T}, I^{\mathrm{L}}_{-T+1}, \cdots, I^{\mathrm{L}}_{0}, \cdots, I^{\mathrm{L}}_{T-1}, I^{\mathrm{L}}_{T}\}$ 作为输入, 以中间位置帧的超分辨率重建结果 I^{SR}_{0} 作为输出, 来估计真实高分辨率帧 I^{HR}_{0}. 其中, $I^{\mathrm{L}}_{t} \in [0, 1]^{H\times W\times C}$ 是指第 t 个编号范围从 $-T$ 到 T 的低分辨率视频帧, 是对应的真实高分辨率帧 $I^{\mathrm{HR}}_{t} \in [0, 1]^{sH\times sW\times C}$ 以倍率 s 下采样而得.

2.1 网络结构

HFRNet 由运动补偿模块、浅层特征模块、特征融合模块和上采样重建模块顺序组成, 如图 1 所示. 这四个组成部分分别负责对齐多个视频帧、提取帧的浅层特征、充分融合网络中不同层中丰富的特征图、增加图像尺寸和细节. 下面对这四个部分进行介绍.

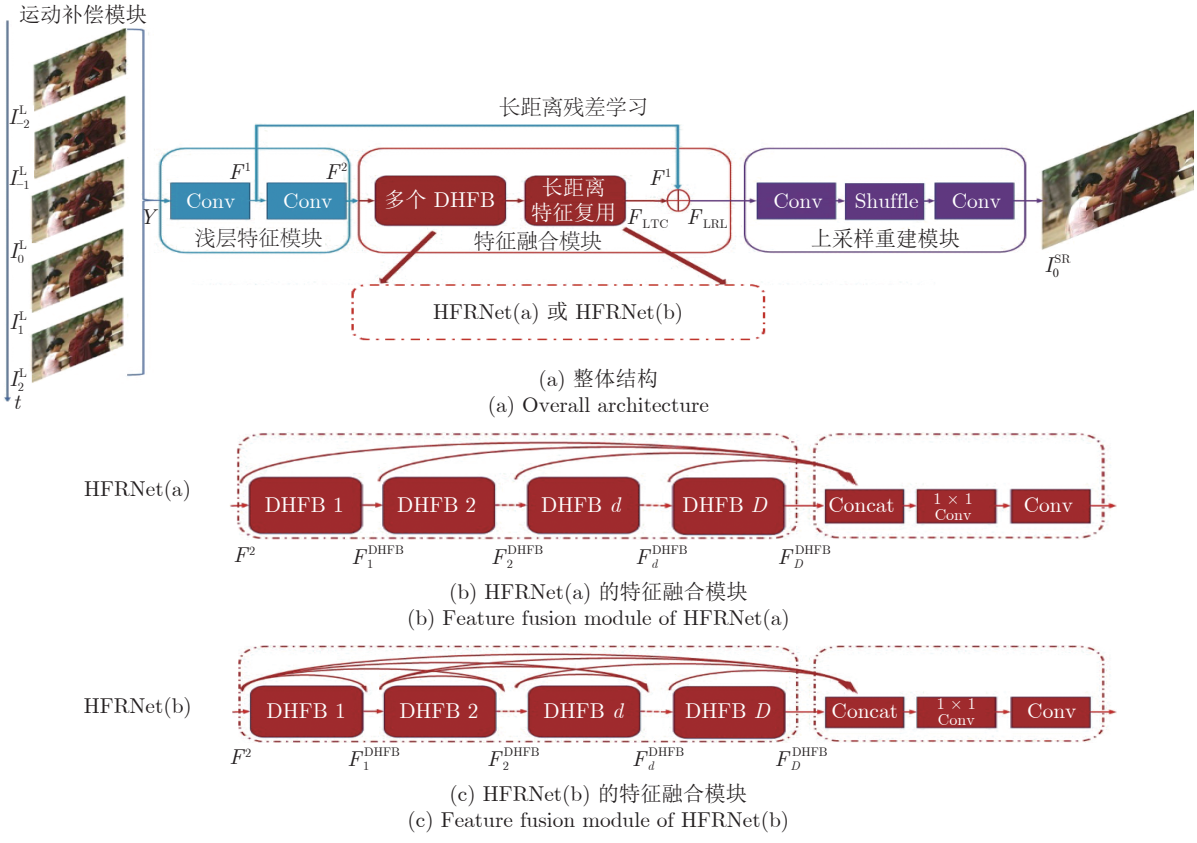
2.1.1 运动补偿模块

运动补偿模块使用适当的光流估计算法对齐视频帧, 以减少视频帧间的大位移或者运动模糊对视频超分辨率的影响. 本文首先利用 Drulea 等[19] 提出的 CLG-TV 算法作为处理运动的运动估计方法, 随后应用普通运动补偿 (Motion compensation, MC) 方法或自适应运动补偿 (Adaptive motion compensation, AMC) 方法对运动估计的结果进行运动补偿, 与 Kappeler 等[6] 采用的方法一样. 如果运动补偿误差非常大, 则 AMC 的效果将比 MC 好, 这是因为 AMC 会减小不可靠的相邻帧影响. AMC 计算式为

$$I^{\mathrm{amc}}_{t}(i, j) = (1-r(i, j))I^{\mathrm{L}}_{0}(i, j)+r(i, j)I^{\mathrm{mc}}_{t}(i, j) \quad (1)$$

$$r(i, j) = \exp(-ke(i, j)) \quad (2)$$

其中, k 是一个常数 (实验中取 0.125), $e(i, j)$ 是参考帧 (序号为 0 的视频帧) 与相邻其他视频帧之间的运动补偿误差. I^{mc}_{t} 是第 t 个经过普通运动补偿方



(a) 整体结构
(a) Overall architecture

(b) HFRNet(a) 的特征融合模块
(b) Feature fusion module of HFRNet(a)

(c) HFRNet(b) 的特征融合模块
(c) Feature fusion module of HFRNet(b)

图 1 层次特征复用网络 (HFRNet) 的结构

Fig. 1 Architecture of hierarchical feature reuse network (HFRNet)

法预处理的相邻帧, I_t^{amc} 是第 t 个经过自适应运动补偿方法预处理的相邻帧.

2.1.2　浅层特征模块

浅层特征模块利用两个卷积层提取视频序列的浅层特征. 本文将使用 MC 或者 AMC 方法进行运动补偿后的视频帧序列 $Y=\{I_{-T}^{\text{mc}}, I_{-T+1}^{\text{mc}}, \cdots, I_0^{\text{mc}}, \cdots, I_{T-1}^{\text{mc}}, I_T^{\text{mc}}\}$ 或 $Y=\{I_{-T}^{\text{amc}}, I_{-T+1}^{\text{amc}}, \cdots, I_0^{\text{amc}}, \cdots, I_{T-1}^{\text{amc}}, I_T^{\text{amc}}\}$ 作为此模块的输入, 其计算式为

$$F^1=f_{\text{SF1}}(Y) \tag{3}$$

$$F^2=f_{\text{SF2}}(Y) \tag{4}$$

其中, f_{SF1} 和 f_{SF2} 是这个模块的第 1 个和第 2 个卷积层. F^1 是第 1 个卷积层的输出, 用来辅助长距离残差学习; F^2 是第 2 个卷积层的输出, 作为特征融合模块的输入.

2.1.3　特征融合模块

特征融合模块通过长距离特征存储机制, 对此模块内的每个密集层次特征块 (DHFB) 的分层特征进行特征复用. 假设在特征融合模块中有 D 个 DHFB, 则其输出为

$$F_{\text{LRL}}=f_{\text{FF}}(F_D^{\text{DHFB}}, \cdots, F_d^{\text{DHFB}}, \cdots, F_1^{\text{DHFB}}, F^2, F^1) \tag{5}$$

其中, f_{FF} 表示特征融合模块. $F_D^{\text{DHFB}}, \cdots, F_d^{\text{DHFB}}, \cdots, F_1^{\text{DHFB}}, F^2, F^1$ 和 F_{LRL} 是特征融合模块的输入和输出. 下面描述特征融合模块的细节. 特征融合模块由多个 DHFB、长距离特征复用和长距离残差学习组成. 图 1 展示了两种网络结构, HFRNet(a) 和 HFRNet(b), 以探究 DHFB 间不同的连接与特征复用方式对性能的影响. 图 2 展示了 DHFB 内部结构. 如图 2 所示, DHFB 由几个残差块和短距离特征复用组成. 第 d 个 DHFB 内的第 c 个残差块的输出为

$$F_{d,c}^{\text{Res}}=f_{\text{RB}}(F_{d,c-1}^{\text{Res}})=\gamma f_{\text{Res}}(F_{d,c-1}^{\text{Res}})+F_{d,c-1}^{\text{Res}} \tag{6}$$

其中, f_{RB} 表示第 d 个 DHFB 中第 c 个残差块的复合运算. γ 是残差比例因子, f_{Res} 表示由两个卷积层组成的残差函数. $F_{d,c-1}^{\text{Res}}$, $F_{d,c}^{\text{Res}}$ 分别是第 d 个 DHFB 的第 c 个残差块的输入和输出. 当 $c=0$ 时, $F_{d,0}^{\text{Res}}=F_{d,\text{pre}}^{\text{DHFB}}$, 对应 DHFB 的输入.

在 HFRNet(a) 中, $F_{d,\text{pre}}^{\text{DHFB}}$ 是 F_{d-1}^{DHFB}, 而在 HFRNet(b) 中, $F_{d,\text{pre}}^{\text{DHFB}}$ 是拼接起来的特征 $[F_{d-1}^{\text{DHFB}}, \cdots, F_1^{\text{DHFB}}, F^2]$.

短距离特征复用用于自适应地融合当前 DHFB 中输入的特征图和其内部每个残差块输出的特征图. 为了减少特征维度, 本文引入了 1×1 卷积层, 以自适应地控制输出特征图信息. 该操作可以表示为

$$F_d^{\text{DHFB}}=f_{\text{STC}}\left([F_{d,\text{pre}}^{\text{DHFB}}, F_{d,1}^{\text{Res}}, \cdots, F_{d,C}^{\text{Res}}]\right) \tag{7}$$

其中, $[:,:,:]$ 是级联运算符, f_{STC} 表示第 d 个 DHFB 中 1×1 卷积的运算. $F_{d,\text{pre}}^{\text{DHFB}}$ 和 F_d^{DHFB} 是第 d 个 DHFB 的输入和输出.

2.1.4　上采样重建模块

上采样重建模块学习从低质量图像到高质量图像的直接映射关系, 最终生成高分辨率的重建图像. 本文在此模块中使用了 Caballero 等[4] 提出的亚像素卷积层–卷积层的两层网络结构. 其输出为

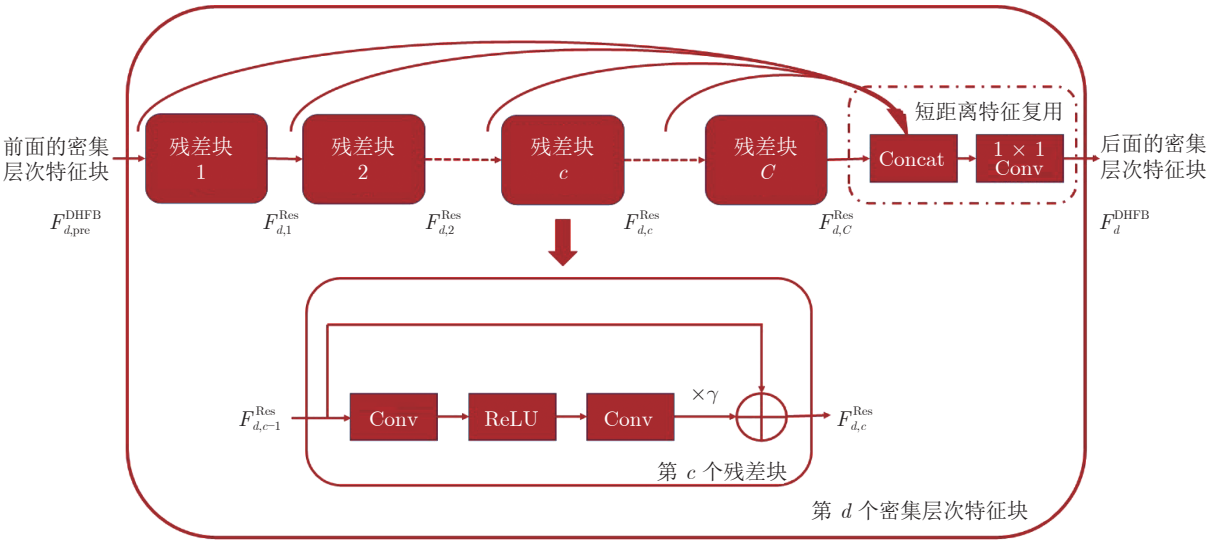


图 2　DHFB 的详细结构

Fig. 2　Detailed architecture of dense hierarchical feature block (DHFB)

$$I_0^{\mathrm{SR}} = f_{\mathrm{Rec}}(F_{\mathrm{LRL}}) \tag{8}$$

其中, f_{Rec} 表示由上述两层网络组成的复合函数. F_{LRL} 是特征融合模块的输出, I_0^{SR} 是中间视频帧的超分辨率重建结果.

2.2 DHFB 间的特征复用

在使用 DHFB 提取多级短距离密集特征后, 本节介绍 DHFB 间的特征复用, 包括长距离特征学习和 DHFB 之间的连接.

2.2.1 长距离特征学习

长距离特征学习包含长距离特征复用和长距离残差学习, 用以自适应融合所有 DHFB 的输出特征, 其表达式为

$$F_{\mathrm{LTC}} = f_{\mathrm{LTC}}\left(\left[F_D^{\mathrm{DHFB}}, \cdots, F_d^{\mathrm{DHFB}}, \cdots, F_0^{\mathrm{DHFB}}\right]\right) \tag{9}$$

其中, f_{LTC} 由两个卷积核大小分别为 1×1 和 3×3 的卷积层组成. 1×1 卷积用于减少所有 DHFB 的融合特征图数目, 而 3×3 卷积则用于提取深层特征. F_d^{DHFB} 是第 d 个 DHFB 的输出.

2.2.2 DHFB 间的连接

在长距离特征复用方法的基础上, 为了探究 DHFB 间不同的特征连接复用方式对 HFRNet 的影响, 本文提出了两种网络结构, 对应图 1 中的 HFRNet(a) 和 HFRNet(b). 下面对这两种网络结构分别进行介绍.

对于 HFRNet(a), 每个 DHFB 的输入都是前一个 DHFB 的直接输出. 在最后一个 DHFB 输出后, 长距离特征复用方法将前面所有 D 个 DHFB 的输出拼接在一起. 第 d 个 DHFB 输出为

$$F_d^{\mathrm{DHFB}} = f^a_{\mathrm{DHFB},\,d}(F^{\mathrm{DHFB}}_{d,\,\mathrm{pre}}) = f^a_{\mathrm{DHFB},\,d}(F_{d-1}^{\mathrm{DHFB}}) = f^a_{\mathrm{DHFB},\,d}(f^a_{\mathrm{DHFB},\,d-1}(\cdots(f^a_{\mathrm{DHFB},\,1}(f_0^{\mathrm{DHFB}}))\cdots) \tag{10}$$

其中, $f^a_{\mathrm{DHFB},\,d}$ 表示 HFRNet(a) 中第 d 个 DHFB. F_{d-1}^{DHFB} 和 F_d^{DHFB} 分别是 HFRNet(a) 中第 d 个 DHFB 的输入和输出, 且 $F_0^{\mathrm{DHFB}} = F^2$.

对于 HFRNet(b), 每个 DHFB 的输入是将前面所有 DHFB 输出拼接在一起, 每个 DHFB 的输出将传递给后面所有 DHFB 的输入. 在最后一个 DHFB 输出后, 长距离特征复用方法将前面所有 D 个 DHFB 的输出拼接在一起. 第 d 个 DHFB 的输出为

$$F_d^{\mathrm{DHFB}} = f^b_{\mathrm{DHFB},\,d}(F^{\mathrm{DHFB}}_{d,\,\mathrm{pre}}) = f^b_{\mathrm{DHFB},\,d}\left(\left[F_{d-1}^{\mathrm{DHFB}}, \cdots, F_1^{\mathrm{DHFB}}, F_0^{\mathrm{DHFB}}\right]\right) \tag{11}$$

其中, $f^b_{\mathrm{DHFB},\,d}$ 表示 1×1 卷积层与 HFRNet(b) 中第 d 个 DHFB 的复合运算. 由于在 HFRNet(b) 中, F_d^{DHFB} 的输入是由其前面 d 个 DHFB 的输出拼接而成, 具有较高的特征维度, 而每个 DHFB 的输入维度和输出维度是相同的. 为了保证输入 DHFB 间特征维度的一致性, 本文使用了 1×1 的卷积运算来减少输入特征图维度, 并达到自适应地融合来自不同 DHFB 的特征的最终目的.

长距离残差学习可以进一步改善将要输入到重建模块的信息流. 对应表达式为

$$F_{\mathrm{LRL}} = F_{\mathrm{LTC}} + F^1 \tag{12}$$

其中, F^1 是浅层特征网络的特征图. F_{LTC} 是长距离特征复用的输出. 经过长距离残差学习之后, 本文获得了具有更多的高频信息的特征图 F_{LRL}, 并以此作为上采样重建网络的输入.

3 实验与结果分析

本节首先描述实验设置, 包括训练和测试数据集、网络结构和训练细节; 其次, 分析网络中 DHFB 的数目 D 和每一个 DHFB 所具有的残差块数目 R 对结果的影响. 另外, 对不同的 DHFB 间特征复用方式, 即 HFRNet(a) 和 HFRNet (b) 的性能进行比较. 最后, 本文将提出的方法与其他一些视频超分辨率方法进行了定性和定量的对比.

3.1 实验设置

本文使用公开的 4K 分辨率视频数据集中的 Myanmar 数据集[33] 进行训练, 在 Myanmar 数据集和 VIDEO4 数据集[1] 进行测试. 为了方便与其他算法进行对比, 本文将 Myanmar 数据集下采样为 960×540 像素的分辨率. Myanmar 数据集中包含 59 个序列, 本文选择其中的 53 个用以训练、6 个用以测试. 对于每一个测试序列, 本文选择了与文献 [7] 相同的 4 帧图像进行测试. VIDEO4 数据集[1] 包含 WALK、FOLIAGE、CITY、CALENDAR 四个序列.

本文从训练序列中共提取 518 760 组 $5 \times 48 \times 48$ 的高分辨率彩色小块, 这些高分辨率彩色小块和对应的低分辨率彩色小块共同构成训练集. 每一组高分辨率彩色小块由 5 个时间间隔为 10 帧彩色图像小块组成, 尺寸为 48×48. 所有彩色小块先转换到 YCbCr 彩色空间中, 并仅取 Y 通道作为训练输入. 在测试阶段, 本文采用峰值信噪比 (Peak signal noise ratio, PSNR) 和结构相似度 (Structure similarity, SSIM) 这两个客观指标评价重建的超分辨率视频帧, 同时, 本文也展示在 YCbCr 色彩空间中的主观视觉重建结果, 其中 Cb 和 Cr 两个通道也是高分辨率 YCbCr 图像先下采样、再上采样回相同尺寸后获得的.

在本文的网络结构中, 除了短距离特征复用与长距离特征复用中的 1×1 卷积层外, 所有其他卷积层的卷积核大小均为 3×3. 除了重建模块中的 3×3 卷积外, 剩下的 3×3 卷积都用零填充来保障卷积前后特征图大小不变, 并且所有卷积的输出特征维度均为 64. 本文在重建模块中采用文献 [4] 的方法, 使用亚像素卷积层而不是反卷积层来生成高质量的 Y 通道重建结果.

本文使用 I_0^{HR} 和 I_0^{SR} 之间的 L2 损失函数训练网络, 使用随机梯度下降算法来优化网络参数, 总共训练 100 个 epoch, 设置批次大小为 64, 初试学习率为 10^{-4}, 并在第 75 个 epoch 时将学习率下降为初始值的 1/10, 使用 TensorFlow 框架在一块 Titan XP GPU 上训练需约 2 天时间.

3.2　实验结果与分析

3.2.1　密集层次特征块数 (D) 和残差块数 (R) 对结果的影响

在本实验中, 本文研究网络中每种基本组件的数目, 即 DHFB 的数目 D 和每个 DHFB 内残差块数目 R 对测试结果的影响. 本文以 HFRNet(a) 的结构作为基础来构建网络. 表 1 展示了 R4D6、R6D4、R6D6、R6D8 和 R8D6 在 CITY, WALK, FOLIAGE 和 CALENDAR 四个序列进行 2 倍视频超分辨率重建任务下得到的 PSNR. 从表 1 可以看出, 随着 DHFB 数目和每个 DHFB 中的残差块数目增加, 每一个序列中的 PSNR 指标都有所提升. 但是, 当模型规模大于 R6D6 时, 虽然仍然可以通过增加参数量提升性能, 但是训练更大模型需要数据量扩充和训练时间的延长, 否则原 R6D8 和 R8D6 容易在第 3.1 节中的实验配置下陷入轻微过拟合状态, 性能有所下降. 在综合考虑了算法性能、网络复杂度与训练难度等因素后, 本文在实验中采用 R6D6 这种组合方式进行网络搭建.

3.2.2　DHFB 之间的连接方式对结果的影响

在本实验中, 本文分析图 1 中 HFRNet(a) 和 HFRNet(b) 这两种不同的结构对于测试的影响. 两种结构的相同之处是均以 R6D6 作为主干网络结构, 并在最后的一个 DHFB 之后, 将前面所有 D 个 DHFB 的输出进行长距离特征复用, 并进行长距离残差学习. 两种结构的不同之处是 DHFB 间特征复用方式, 即当前 DHFB 的输入. HFRNet(a) 的当前 DHFB 输入是上一个 DHFB 的输出; HFRNet(b) 采用密集连接方式, 当前 DHFB 的输入则是将先前所有 DHFB 输出拼接在一起.

表 2 展示了这两种结构在 2 倍和 3 倍视频超分辨率重建任务中的结果及所需参数量. 实验结果表明, 本文提出的层次特征复用机制能够显著提升超分辨率性能. 当去除层次特征复用机制后, 模型在 2 倍、3 倍超分辨率任务上的 PSNR 值均有接近 1 dB 的下降. 在本文提出的两种包含层次特征复用机制的结构中, HFRNet(b) 的综合性能优于 HFRNet(a),

表 1　不同 DHFB 数目 (D) 和每个 DHFB 残差块数目 (R) 对 2 倍率超分辨率重建性能的影响 (PSNR (dB))
Table 1　The impact (PSNR (dB)) of different numbers of DHFBs (D) and residual blocks (R) on the performance of 2× super-resolution reconstruction task

模块组合方式	CITY 序列	WALK 序列	FOLIAGE 序列	CALENDAR 序列	平均 PSNR
R4D6	34.342	36.846	32.045	27.071	32.576
R6D4	34.339	37.101	32.117	27.067	32.656
R6D6	34.896	37.210	32.224	27.137	32.866
R6D8	34.901 (±0.035)	37.102 (±0.054)	32.187 (±0.069)	27.140 (±0.007)	32.833 (±0.041)
R8D6	34.633 (±0.039)	36.873 (±0.025)	32.144 (±0.050)	27.109 (±0.019)	32.690 (±0.034)

表 2　不同网络结构实验结果的平均 PSNR 及所需参数量
Table 2　The average PSNR and number of parameters for different network architectures

尺度	网络结构	参数量	CITY 序列 (dB)	WALK 序列 (dB)	FOLIAGE 序列 (dB)	CALENDAR 序列 (dB)	平均 PSNR (dB)
×2	无层次特征复用	2.85 M	33.793	35.919	31.884	26.291	31.972
	HFRNet(a)	3.01 M	34.896	37.210	32.224	27.137	32.866
	HFRNet(b)	3.10 M	35.104	37.218	32.230	27.158	32.927
×3	无层次特征复用	2.85 M	27.220	30.113	27.019	23.344	26.924
	HFRNet(a)	3.01 M	28.235	31.513	27.539	24.190	27.869
	HFRNet(b)	3.10 M	28.240	31.613	27.587	24.217	27.914

在 2 倍、3 倍视频超分辨率重建任务中分别提升了 0.061 dB、0.045 dB. 主要原因是 HFRNet(b) 结构中的层次特征复用连接比 HFRNet(a) 更密集且层次融合形式更丰富. 这表明本文提出的层次特征复用机制在视频超分辨率任务上十分有效, 引入该结构之后能使超分辨率性能稳定提升.

3.2.3 不同光流估计法的影响

本实验研究不同的光流估计法对超分辨率效果的影响, 主要对比本文采用的 CLG-TV 算法与近期提出的一种基于 CNN 方法的光流重建网络 (Optical flow reconstruction network, OFRNet)[34]. 引入基于 CNN 的光流估计法之后, 网络的复杂度与训练难度也会随之提高, 但是实验结果表明 (如表 3 所示), 利用 CNN 估计光流并进行端对端训练, 可以使得超分辨率的性能获得进一步提升.

3.2.4 不同运动补偿方法的影响

本节实验讨论在使用不同的运动补偿方法 (MC 和 AMC) 或参数 (k 值) 时超分辨率结果的对比, 实验结果如表 4 所示. 实验结果表明, 在使用普通运动补偿 (MC) 时, k 值会对超分辨率结果产生一定影响, 但影响并不十分明显, 且当 $k = 0.125$ 时, 超分辨率结果的 PSNR 取得最大值. 而当使用自适应运动补偿 (AMC) 时, 超分辨率结果的 PSNR 值比使用 MC 时的最大值高出 0.213 dB (2×) 和 0.092 dB (3×).

3.3 与其他方法的比较

在本节中, 本文将 HFRNet 与几种基于单图像和基于多帧的超分辨率算法进行比较, 以证明本文提出的结构的优越性. 基于单图像的方法包括双三次插值 (Bicubic)、超分辨率重建卷积神经网络 (Super-resolution convolutional neural network, SRCNN)[35], 基于多帧的方法包括 VSRNet[6]、VESPCN[4]、MCResNet[7]、显示细节的视频超分辨率网络 (Detail-revealed video super-resolution network, DRVSR)[5]、残差循环网络 (Residual recurrent convolutional neural network, RRCN)[36]、三维超分辨率重建网络 (3D super-resolution network, 3DSRNet)[37]、残差特征生成对抗的视频超分辨率网络 (Residual feature general adversarial network for video super-resolution, VSRResFeatGAN)[20] 和多记忆视频超分辨率重建网络 (Multi-memory convolution neural network, MMCNN)[38]. 本文提出的 HFRNet 使用 R6D6 的结构.

图 3 展示了本文方法与其他超分辨率方法在 VIDEO4 测试数据集和 Myanmar 测试数据集上的 PSNR 和 SSIM 对比结果. 在 VIDEO4 测试数据集上, 本文方法比很多方法效果更好. 在 Myanmar 测试数据集上, 本文方法在 2 倍和 3 倍超分辨率任务上的性能明显优于现有方法.

图 4 和图 5 展示了本文方法与其他几个方法在超分辨率重建任务的主观对比. 我们添加了深层超分辨率卷积神经网络 (Very deep super-resolution convolutional neural network, VDSR)[39]、基于拉普拉斯金字塔的深层超分辨率网络 (Deep Laplacian pyramid network, LapSRN)[40]、基于光流的视频超分辨率网络 (Super-resolve optical flow video super-resolution network, SOF-VSR)[41] 用于对比. 图 4 是在 VIDEO4 数据集下进行 3 倍超分辨率重建, 与其他方法相比, 本文方法都可以更清晰地重建汽车的细节. 图 5 是对 Myanmar 测试数据集进行 4 倍超分辨率重建的结果. 图 6 对各种方法重建结果的细节放大, 以突出各种算法的细节处理能力. 从图中可以看出, 本文提出的算法重建出的高分辨

表 3 不同光流估计方法对超分辨率重建性能的影响 (PSNR (dB))

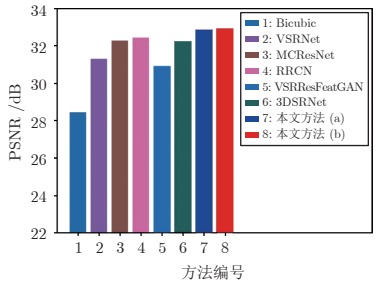
Table 3 The impact (PSNR (dB)) of different optical flow estimation methods on super-resolution reconstruction performance

尺度	光流估计算法	CITY 序列	WALK 序列	FOLIAGE 序列	CALENDAR 序列	平均 PSNR
×2	CNN-based	35.226	37.106	32.244	27.817	33.098
	CLG-TV	35.104	37.218	32.230	27.158	32.927
×3	CNN-based	28.255	32.103	27.590	24.766	28.179
	CLG-TV	28.240	31.613	27.587	24.217	27.914

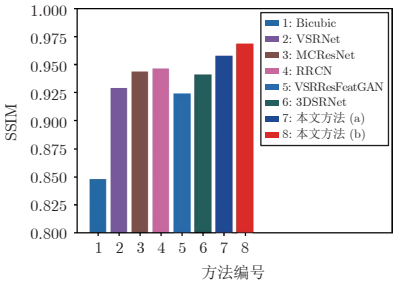
表 4 不同运动补偿算法对超分辨率重建性能的影响 (平均 PSNR (dB))

Table 4 Average PSNR (dB) in video super-resolution task, with different motion compensation algorithm

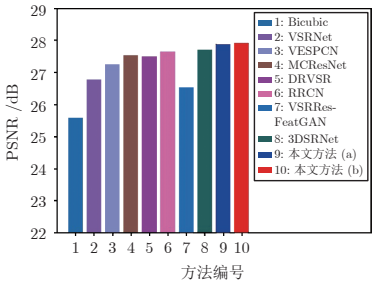
运动补偿算法与参数	尺度	MC ($k = 0.050$)	MC ($k = 0.100$)	MC ($k = 0.125$)	MC ($k = 0.175$)	AMC
平均 PSNR (dB)	×2	32.493	32.510	32.714	32.615	32.927
	×3	27.505	27.684	27.822	27.694	27.914



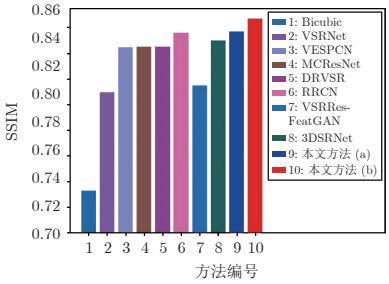
(a) VIDEO4 下 2 倍视频超分辨率的平均 PSNR
(a) Average PSNR for 2× video super-resolution task on VIDEO4 dataset



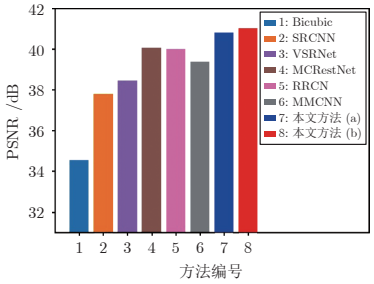
(b) VIDEO4 下 2 倍视频超分辨率的平均 SSIM
(b) Average SSIM for 2× video super-resolution task on VIDEO4 dataset



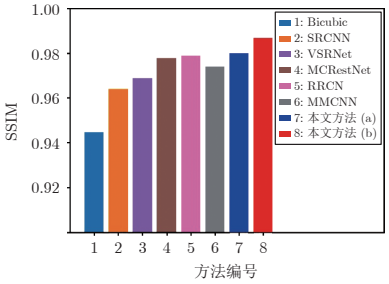
(c) VIDEO4 下 3 倍视频超分辨率的平均 PSNR
(c) Average PSNR for 3× video super-resolution task on VIDEO4 dataset



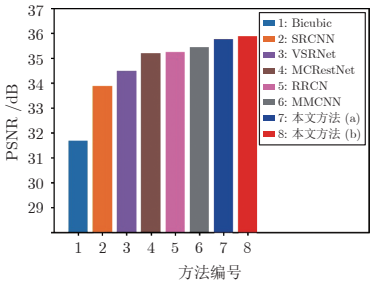
(d) VIDEO4 下 3 倍视频超分辨率的平均 SSIM
(d) Average SSIM for 3× video super-resolution task on VIDEO4 dataset



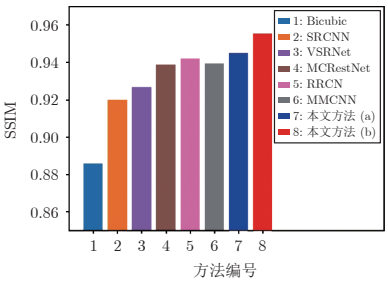
(e) Myanmar 下 2 倍视频超分辨率的平均 PSNR
(e) Average PSNR for 2× video super-resolution task on Myanmer dataset



(f) Myanmar 下 2 倍视频超分辨率的平均 SSIM
(f) Average SSIM for 2× video super-resolution task on Myanmar dataset



(g) Myanmar 下 3 倍视频超分辨率的平均 PSNR
(g) Average PSNR for 3× video super-resolution task on Myanmar dataset



(h) Myanmar 下 3 倍视频超分辨率的平均 SSIM
(h) Average SSIM for 3× video super-resolution task on Myanmar dataset

图 3　本文方法和其他方法在 VIDEO4 和 Myanmar 数据集下得到的平均 PSNR 和平均 SSIM

Fig. 3　Average PSNRs and SSIMs obtained by our method and other methods on VIDEO4 and Myanmar datasets

率图像从清晰度和真实性方面均优于现有算法. 这些比较结果表明, 本文方法在定量指标和主观对比上均优于其他很多方法. 这说明了学习层次结构特征的有效性.

3.4　算法的复杂度与时效性

本节讨论所提出算法的运算复杂度与时效性. 复杂度方面, 由于本文提出的 HFRNet(a) 和 HFRNet(b) 结构采取了全卷积结构, 所以运算复杂度与输入视频的尺寸 (像素数) 和帧数几乎成正比. 对于训练过程中采用的视频尺寸 (帧数为 5, 分辨率为 48×48), 其每秒浮点运算次数 (Floating point operations per second, FLOPS) 分别为 35.6 G 和 36.7 G. 时效性方面, 在本文实验所采用的硬件设备 Titan XP GPU 上, $48 \times 48 \times 5$ 的视频序列在两种网络中的处理时间分别为 0.126 s 和 0.143 s; 光流估计部分采用了 CLG-TV 算法, 由于该部分的实现主要在 CPU 上, 所以其耗时可能为超分辨率

图 4 HFRNet 与其他模型在 VIDEO4 数据集图像上超分辨率的定性对比

Fig. 4 Qualitative super-resolution comparison of HFRNet with other models on an image from VIDEO4 dataset

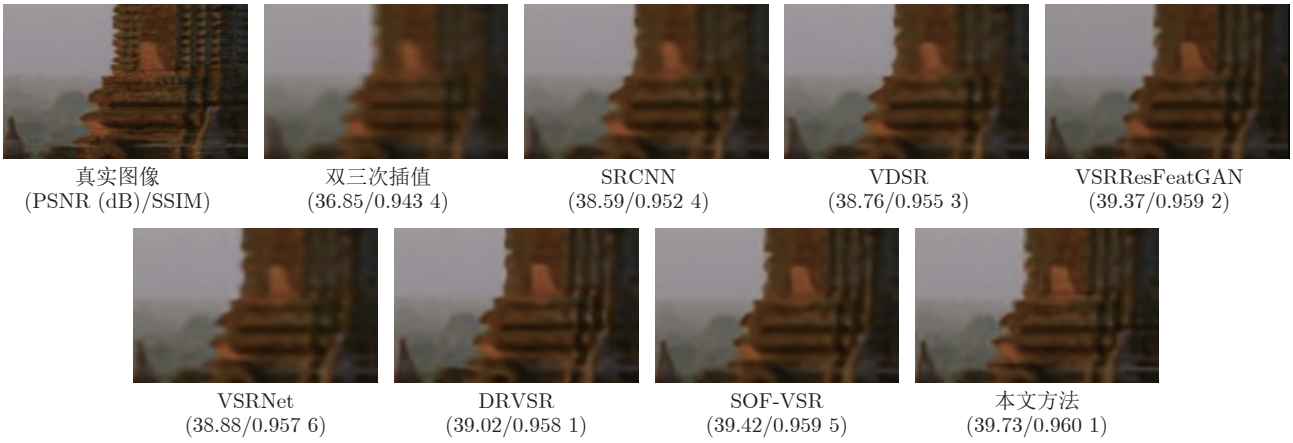


图 5 HFRNet 与其他模型在 Myanmar 数据集图像上超分辨率的定性对比

Fig. 5 Qualitative super-resolution comparison of HFRNet with other models on an image from Myanmar dataset

图 6 HFRNet 重建细节与其他模型超分辨率的定性对比

Fig. 6 Qualitative super-resolution comparison of the reconstruction details by HFRNet and other models

网络的 6 ~ 10 倍; 当光流估计部分替换为 CNN-based 模型后, 光流估计时间可缩短至约 0.2 s.

4 结束语

本文提出一种用于视频超分辨率重建的层次特征复用网络 (HFRNet). 在利用来自原始低分辨率帧的帧间信息的基础上, 本文使用由密集层次特征块 (DHFB) 构成的神经网络来提取并融合层次化的特征. 在一个密集层次特征块内, 短距离特征复用可以自适应地学习融合每个残差块提取的层次特征. 在多个密集层次特征块之间, 长距离特征复用能够自适应地控制所有密集层次特征块中保留的信息, 并使深层网络的训练过程更加稳定. 此外, 本文提出两种网络结构, 以探讨密集层次特征块间的不同特征复用方式对结果的影响. 通过短距离和长距离特征复用, 层次特征复用网络具有很强的特征提取和表达能力. 实验结果表明, 本文的方法在客观指标和主观结果方面, 相比之前其他视频超分辨率重建方法有所提升.

References

1 Liu C, Sun D. On Bayesian adaptive video super resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, **36**(2): 346−360

2 Shahar O, Faktor A, Irani M. Space-time super-resolution from a single video. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Colorado Springs, USA: IEEE, 2011. 3353−3360

3 Zhou Y, Wang Y, Zhang Y, Du X, Liu H, Li C. Manifold learning based super resolution for mixed-resolution multi-view video in visual internet of things. In: Proceedings of the International Conference on Artificial Intelligence for Communications and Networks. Harbin, China: Springer, 2019. 486−495

4 Caballero J, Ledig C, Aitken A, Acosta A, Totz J, Wang Z, et al. Real-time video super-resolution with spatio-temporal networks and motion compensation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017. 2848−2857

5 Tao X, Gao H, Liao R, Wang J, Jia J. Detail-revealing deep video super-resolution. In: Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017. 4482−4490

6 Kappeler A, Yoo S, Dai Q, Katsaggelos A K. Video super-resolution with convolutional neural networks. *IEEE Transactions on Computational Imaging*, 2016, **2**(2): 109−122

7 Li D, Wang Z. Video super resolution via motion compensation and deep residual learning. *IEEE Transactions on Computational Imaging*, 2017, **3**(4): 749−762

8 Zhou Y, Zhang Y, Xie X, Kung S-Y. Image super-resolution based on dense convolutional auto-encoder blocks. *Neurocomputing*, 2021, **423**(1): 98−109

9 Li Jin-Xin, Huang Zhi-Yong, Li Wen-Bin, Zhou Deng-Wen. Image super-resolution based on multi hierarchical features fusion network. *Acta Automatica Sinica*, 2023, **49**(1): 161−171
(李金新, 黄志勇, 李文斌, 周登文. 基于多层次特征融合的图像超分辨率重建. 自动化学报, 2023, **49**(1): 161−171)

10 Zhang Yi-Feng, Liu Yuan, Jiang Cheng, Cheng Xu. A curriculum learning approach for single image super resolution. *Acta Automatica Sinica*, 2020, **46**(2): 274−282
(张毅锋, 刘袁, 蒋程, 程旭. 用于超分辨率重建的深度网络递进学习方法. 自动化学报, 2020, **46**(2): 274−282)

11 Zhou Y, Feng L, Hou C, Kung S-Y. Hyperspectral and multispectral image fusion based on local low rank and coupled spectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, **55**(10): 5997−6009

12 Zhou Deng-Wen, Zhao Li-Juan, Duan Ran, Chai Xiao-Liang. Image super-resolution based on recursive residual networks. *Acta Automatica Sinica*, 2019, **45**(6): 1157−1165
(周登文, 赵丽娟, 段然, 柴晓亮. 基于递归残差网络的图像超分辨率重建. 自动化学报, 2019, **45**(6): 1157−1165)

13 Sun Xu, Li Xiao-Guang, Li Jia-Feng, Zhuo Li. Review on deep learning based image super-resolution restoration algorithms. *Acta Automatica Sinica*, 2017, **43**(5): 697−709
(孙旭, 李晓光, 李嘉锋, 卓力. 基于深度学习的图像超分辨率复原研究进展. 自动化学报, 2017, **43**(5): 697−709)

14 Xie X K, Zhou Y, Kung S-Y. Exploiting operation importance for differentiable neural architecture earch. arXiv preprint arXiv: 1911.10511, 2019.

15 Huo S, Zhou Y, Xiang W, Kung S-Y. Semi-supervised learning based on a novel iterative optimization model for saliency detection. *IEEE Transactions on Neural Network and Learning System*, 2019, **30**(1): 225−241

16 Zhou Y, Mao A, Huo S, Lei J, Kung S-Y. Salient object detection via fuzzy theory and object-level enhancement. *IEEE Transactions on Multimedia*, 2019, **1**(1): 74−85

17 Jo Y, Oh S W, Kang J, Kim S J. Deep video super-resolution network using dynamic upsampling filters without explicit motion compensation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 3224−3232

18 Pan Zhi-Yong, Yu Mei, Xie Deng-Mei, Song Yang, Jiang Gang-Yi. Fast video super-resolution reconstruction using a succinct convolutional neural network. *Journal of Optoelectronics · Laser*, 2018, **29**(12): 1332−1341
(潘志勇, 郁梅, 谢登梅, 宋洋, 蒋刚毅. 采用精简卷积神经网络的快速视频超分辨率重建. 光电子 · 激光, 2018, **29**(12): 1332−1341)

19 Drulea M, Nedevschi S. Total variation regularization of local-global optical flow. In: Proceedings of the IEEE Conference on Intelligent Transportation Systems. Washington D C, USA: IEEE, 2011. 318−323

20 Lucas A, López-Tapia S, Molina R, Katsaggelos A K. Generative adversarial networks and perceptual losses for video super-resolution. *IEEE Transactions on Image Processing*, 2019, **28**(7): 3312−3327

21 Zhou Y, Yang J X, Li H R, Cao T, Kung S-Y. Adversarial learning for multiscale crowd counting under complex scenes. *IEEE Transactions on Cybernetics*, 2021, **51**(11): 5423−5432

22 Zhou Y, Huo S, Xiang W, Hou C, Kung S-Y. Semi-supervised salient object detection using a linear feedback control system model. *IEEE Transactions on Cybernetics*, 2019, **49**(4): 1173−1185

23 Huo S, Zhou Y, Lei J, Ling N, Hou C. Iterative feedback control-based salient object segmentation. *IEEE Transactions on Multimedia*, 2018, **20**(6): 1350−1364

24 Zhou Y, Zhang T, Huo S, Hou C, Kung S-Y. Adaptive irregular graph construction based salient object detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 2020, **30**(6): 1569−1582

25 Szegedy C, Liu W, Jia Y Q, Sermanet P, Reed S E, Anguelov D,

et al. Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE, 2015. 1−9

26 He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016. 770−778

27 Huang G, Liu Z, Van Der Maaten L, Weinberger K Q. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017. 2261−2269

28 Zhang Y, Tian Y, Kong Y, Zhong B, Fu Y. Residual dense network for image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 2472−2481

29 Zhou Y, Du X T, Wang M F, Huo S W, Zhang Y D, Kung S-Y. Cross-scale residual network: A general framework for image super-resolution, denoising, and deblocking. *IEEE Transactions on Cybernetics*, 2022, **52**(7): 5855−5867

30 Yi P, Wang Z, Jiang K, Shao Z, Ma J. Multi-temporal ultra dense memory network for video super-resolution. *IEEE Transactions on Circuits and Systems for Video Technology*, 2020, **30**(8): 2503−2516

31 Yi P, Wang Z, Jiang K, Jiang J, Ma J. Progressive fusion video super-resolution network via exploiting non-local spatio-temporal correlations. In: Proceedings of the IEEE International Conference on Computer Vision. Seoul, South Korea: IEEE, 2019. 3106−3115

32 Yi P, Wang Z Y, Jiang K, Jiang J J, Lu T, Ma J. A progressive fusion generative adversarial network for realistic and consistent video super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, **44**(5): 2264−2280

33 H Inc. Myanmar 60p [Online], available: http://www.harmonicinc.com/resources/videos/4k-video-clip-center, May 20, 2021

34 Wang L, Guo Y, Liu L, Lin Z, Deng X, An W. Deep video super-resolution using HR optical flow estimation. *IEEE Transactions on Image Processing*, 2020, **29**(1): 4323−4336

35 Dong C, Loy C C, He K, Tang X. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, **38**(2): 295−307

36 Li D, Liu Y, Wang Z. Video super-resolution using motion compensation and residual bidirectional recurrent convolutional network. In: Proceedings of the IEEE International Conference on Image Processing. Beijing, China: IEEE, 2017. 1642−1646

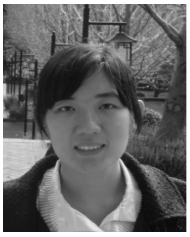
37 Kim S Y, Lim J, Na T, Kim M. Video super-resolution based on 3D-CNNs with consideration of scene change. In: Proceedings of the IEEE International Conference on Image Processing. Taipei, China: IEEE, 2019. 2831−2835

38 Wang Z, Yi P, Jiang K, Jiang J, Han Z, Lu T, et al. Multi-memory convolutional neural network for video super-resolution. *IEEE Transactions on Image Processing*, 2019, **28**(5): 2530−2544

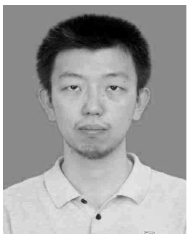
39 Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016. 1646−1654

40 Lai W S, Huang J B, Ahuja N, Yang M H. Deep Laplacian pyramid networks for fast and accurate super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017. 5835−5843

41 Wang L, Guo Y, Lin Z, Deng X, An W. Learning for video super-resolution through HR optical flow estimation. In: Proceedings of the Asian Conference on Computer Vision. Perth, Australia: Springer, 2018. 514−529



周 圆 天津大学电气自动化与信息工程学院副教授. 主要研究方向为计算机视觉与图像/视频通信. 本文通信作者.
E-mail: zhouyuan@tju.edu.cn
(**ZHOU Yuan** Associate professor at the School of Electrical and Information Engineering, Tianjin University. Her research interest covers computer vision and image/video communication. Corresponding author of the paper.)



王明非 天津大学电气自动化与信息工程学院硕士研究生. 主要研究方向为计算机视觉与机器学习.
E-mail: wmf997@126.com
(**WANG Ming-Fei** Master student at the School of Electrical and Information Engineering, Tianjin University. His research interest covers computer vision and machine learning.)

杜晓婷 天津大学电气自动化与信息工程学院硕士研究生. 主要研究方向为计算机视觉与机器学习.
E-mail: 18225511181@163.com
(**DU Xiao-Ting** Master student at the School of Electrical and Information Engineering, Tianjin University. Her research interest covers computer vision and machine learning.)

陈艳芳 天津大学电气自动化与信息工程学院博士研究生. 主要研究方向为计算机视觉与机器学习.
E-mail: chyf@tju.edu.cn
(**CHEN Yan-Fang** Ph.D. candidate at the School of Electrical and Information Engineering, Tianjin University. Her research interest covers computer vision and machine learning.)