

基于深度学习特征匹配的视频超分辨率方法

程松盛 潘金山

南京理工大学计算机科学与工程学院 南京 210094

(835357340@qq.com)

摘要 视频复原的目标是从给定的退化视频序列中把潜在的高质量视频复原出来。现有的视频复原方法主要集中在如何有效地找到相邻帧之间的运动信息,然后利用运动信息建立相邻帧之间的匹配。与这些方法不同,文中提出了基于深度学习特征匹配的方法来解决视频超分辨率问题。首先,通过深度卷积神经网络计算出相邻帧之间的运动信息;然后,采用一个浅层深度卷积神经网络从输入的视频帧中提取特征,基于估计到的运动信息,将浅层深度卷积神经网络提取到的特征匹配到中间视频帧对应的特征中,并将得到的特征进行有效融合;最后,采用一个深度卷积神经网络重建视频帧。大量的实验结果验证了基于深度学习特征匹配的方法能有效地解决视频超分辨率问题。与现有的基于视频帧匹配的方法相比,所提方法在现有的公开视频超分辨率数据集上取得了较好的效果。

关键词: 视频复原;视频超分辨率;深度卷积神经网络;特征匹配;运动估计

中图分类号 TP751

Video Super-resolution Method Based on Deep Learning Feature Warping

CHENG Song-sheng and PAN Jin-shan

School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

Abstract Video restoration aims to restore potential clear videos from a given degraded video sequence. Existing video restoration methods usually focus on modeling the motion information among adjacent frames and establishing the alignment among them. Different from these methods, this paper proposes a feature warping method based on deep learning for video super-resolution. Firstly, the proposed algorithm estimates the motion information based on deep convolutional neural networks. Then, it develops a shallow deep convolutional neural network to estimate the features from input frames. Based on the estimated motion information, this paper warps the deep features to those of the central frames. Next, the proposed method fuses the deep features effectively. Finally, this paper proposes a restoration network which is able to reconstruct clear frames. Experimental results demonstrate the effectiveness of the proposed algorithm. The proposed algorithm performs well on the benchmark datasets compared to existing methods.

Keywords Video restoration, Video super-resolution, Deep convolutional neural network, Feature warping, Motion estimation

1 引言

视频超分辨率是从给定的低质量视频中复原出高质量的视频。近年来,随着各种成像设备的普及,这些设备所导致的低质量视频给后续相关智能分析带来了极大的困难。如何从低质量视频中复原出高质量视频是一个非常重要且极具挑战性的问题,这一问题逐渐受到学者们的广泛关注。

由于相邻的视频帧之间通常包含丰富的图像信息,如何从相邻帧获取到更多有用的信息是视频超分辨率任务的一个极大的挑战。同时,相邻帧之间通常会有不同程度、不同种类的运动模糊,因此如何设计有效的运动信息来估计和匹配模

块是视频超分辨率的另一大难点。

视频超分辨率是一个病态问题,传统的方法大多通过设计各种各样的正则化方法来约束该问题,然后通过求解相应的变分模型^[1-6]来解决此问题。求解过程主要包括相邻帧之间的运动信息估计和匹配以及视频帧的复原过程。虽然基于正则化的方法在一定程度上取得了较好的效果,但是这些方法往往涉及到针对高度非凸能量函数求解问题。此外,由于正则化项的设计大多基于统计特征的方法,因此在一定程度上不能有效地刻画视频数据的分布特性,从而导致算法不能有效地恢复视频帧中的一些结构性细节。

深度学习为解决图像超分辨率问题提供了一种有效的途

到稿日期:2020-08-31 返修日期:2020-10-26

基金项目:国家自然科学基金(61872421)

This work was supported by the National Natural Science Foundation of China(61872421).

通信作者:潘金山(jspan@njust.edu.cn)

径。随着端到端的深度卷积神经网络在单幅图像超分辨率方法^[7]中获得成功,大量的基于深度卷积神经网络的单幅图像超分辨率方法^[8-13]相继被提出。虽然这些基于单幅图像的超分辨率方法取得了很大的成功,但它们没有利用时序信息,因此不能直接有效地应用到视频超分辨率问题中。为了解决这一问题,一些工作尝试结合运动信息估计和匹配方法来解决视频超分辨率的问题。常用的方法包括亚像素运动补偿^[14]、可形变的对齐网络^[15-16]、空间对齐网络^[17-19]、核预估网络^[20]等。为了更好地刻画视频的时序信息,循环神经网络与生成对抗网络被应用于视频超分辨率。本文发现,这些方法获得成功的最主要的因素就是解决了如何通过帧间的运动信息实现相邻帧匹配的问题。然而,这些方法的匹配主要是在原始的图像空间进行,很少有方法考虑在特征空间进行匹配。

本文的核心贡献在于挖掘特征空间匹配方法对视频超分辨率问题的影响,提出了基于深度学习特征匹配的视频超分辨率方法。其主要贡献如下:

- (1)提出了基于深度学习特征匹配的方法,用于解决视频超分辨率问题;
- (2)分析了特征匹配在视频超分辨率问题上的作用,说明在特征空间进行匹配能够有效地解决视频超分辨率问题;
- (3)在现有的视频超分辨率数据集上定性和定量地验证了所提算法的有效性。

2 相关工作

本节介绍了基于变分模型的视频超分辨率方法和基于深度学习的视频超分辨率方法。

(1)基于变分模型的方法。由于视频超分辨率是一个病态问题,因此早期的方法^[1-4]主要通过设计有效的先验知识来解决该问题。还有一些方法^[5-6]在最大后验估计框架下同时估计运动场、模糊核和潜在清晰图像。Liu等^[5]通过使用贝叶斯模型,同时估计运动场、模糊核、潜在清晰图像和噪声来解决视频超分辨率问题。Ma等^[6]提出了一种有效的最大期望模型,用于同时解决视频超分辨率和模糊估计问题。虽然这些方法已经取得了一些较好的成果,但这些算法通常都存在着复杂的优化问题,并且最终的结果受限于人为的先验。

(2)基于深度学习的方法。基于深度学习的单幅图像超分辨率的方法^[7-14]取得了很大的进展。为此,很多方法^[15-25]尝试通过捕获时序信息来实现视频超分辨率。例如,Huang等^[21]设计了一种有效的双向递归卷积网络来刻画长期的上下文信息;有些算法^[22-23]基于人为设计的先验来进行运动估计,然后使用深度神经网络模型来复原高质量图像;Caballero等^[18]提出了一种有效的运动补偿算法,并通过捕获的时序信息来完成视频超分辨率的任务;Liu等^[17]提出了一个自适应的时序神经网络和空间对齐网络来更好地收集时序信息;Tao等^[24]提出了一种基于运动场估计的亚像素运动补偿层;Xue等^[19]论证了光流对视频图像恢复的影响,并提出了一种有效的视频复原框架,解决了一般的视频复原问题。还有一些方法不同于上述工作中显式地使用光流来进行对准。例

如,Jo等^[20]通过动态估计上采样的滤波器来进行视频复原;Harris等^[25]基于递归网络扩展出一种深度-反投影方法^[3];Wang等^[16]引入可变形卷积^[15]并提出了一种有效的时间和空间注意力机制来解决视频复原问题。为了生成更加真实的图片,基于生成式对抗网络的方法被用来解决单幅图像^[13,26-27]和视频^[28-29]超分辨率问题。这些算法在视频超分辨率任务上都达到了较好的效果。然而,这些方法大多是在图像空间中完成运动信息估计和相邻帧的匹配,没有充分利用特征信息在相邻帧匹配中的作用。

3 本文方法

本文的重点不在于网络结构设计,而是挖掘深度学习特征对于视频超分辨率方法的影响,从而提出基于深度学习特征匹配的视频超分辨率算法。本文算法主要包括4个模块:运动信息估计模块 \mathcal{M} 、时空信息特征提取模块 \mathcal{T} 、特征融合模块 \mathcal{F} 和清晰帧重建模块 \mathcal{R} 。网络的整体框架结构如图1所示。下文分别针对这4个模块进行详细的阐述。

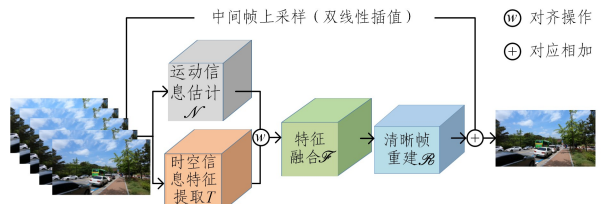


图1 基于深度学习特征匹配的视频超分辨率算法网络结构

Fig. 1 Network architecture of video super-resolution based on deep feature warping

3.1 运动信息估计

运动信息估计模块 \mathcal{M} 主要用来估计相邻帧之间的运动信息,从而借助运动信息构建形变矩阵,实现相邻帧之间的匹配。考虑到模型大小和运行效率的问题,本文采用PWC-Net^[30]作为运动信息估计模块 \mathcal{M} 的网络。具体地,给定视频图像序列中第 i 帧和第 j 帧低分辨率图像 $\{L_i, L_j\}$, L 的尺寸为 $c \times h \times w$,其中 c 代表图像的通道数, h 和 w 分别代表图像的高度和宽度。本文通过式(1)计算运动信息。

$$u_{i \rightarrow j} = \mathcal{M}(L_i, L_j) \quad (1)$$

其中, $u_{i \rightarrow j}$ 表示从 L_j 到 L_i 的运动信息, \mathcal{M} 表示PWC-Net。对于任意其他相邻帧的运动信息,网络 \mathcal{M} 均共享参数。

3.2 时空信息特征提取

对于给定的相邻帧 $\{L_i, L_j\}$,在估计出运动信息 u 之后,已有方法通常采用 u 构建形变矩阵 \mathbf{M} ,然后作用于视频帧,使得作用后的视频帧能够与中间的视频帧匹配,这一过程往往通过双线性插值的方法来实现。与此方法不同,本文采用特征空间匹配,在特征空间中对相邻帧进行对齐。具体地,给定低分辨率视频帧 $\{L_{i-2}, L_{i-1}, L_i, L_{i+1}, L_{i+2}\}$,通过一个浅层深度卷积神经网络 \mathcal{T} ,分别得到深度特征 $\{f_{i-2}, f_{i-1}, f_i, f_{i+1}, f_{i+2}\}$, f 的尺寸为 $c_f \times h \times w$,其中 c_f 为特征通道数。其具体计算方式如下:

$$\begin{aligned} f_{i-2} &= \mathcal{T}(L_{i-2}) & f_{i-1} &= \mathcal{T}(L_{i-1}) & f_i &= \mathcal{T}(L_i) \\ f_{i+1} &= \mathcal{T}(L_{i+1}) & f_{i+2} &= \mathcal{T}(L_{i+2}) \end{aligned} \quad (2)$$

其中, \mathcal{T} 表示时空信息特征提取网络。

该特征提取网络具体包括 1 个卷积层(Conv)和 5 个残差块(ResBlock)^[12], 其中卷积层和每个残差块中的滤波器个数均为 128。所有卷积层中卷积核的尺寸为 3×3 。本模块的具体网络结构如图 2 所示。

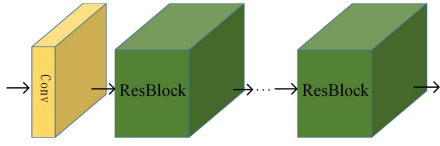


图 2 时空信息特征提取模块 \mathcal{T} 的网络结构图

Fig. 2 Architecture of spatial-temporal feature extraction module \mathcal{T}

针对这些特征, 本文通过以下方式得到匹配的特征信息:

$$\mathbf{d}_{i-2}(x) = \mathcal{W}(f_{i-2}(x), u_{i-2 \rightarrow i}(x)) = f_{i-2}(x) + u_{i-2 \rightarrow i}(x)$$

$$\mathbf{d}_{i-1}(x) = \mathcal{W}(f_{i-1}(x), u_{i-1 \rightarrow i}(x)) = f_{i-1}(x) + u_{i-1 \rightarrow i}(x)$$

$$\mathbf{d}_{i+1}(x) = \mathcal{W}(f_{i+1}(x), u_{i+1 \rightarrow i}(x)) = f_{i+1}(x) + u_{i+1 \rightarrow i}(x)$$

$$\mathbf{d}_{i+2}(x) = \mathcal{W}(f_{i+2}(x), u_{i+2 \rightarrow i}(x)) = f_{i+2}(x) + u_{i+2 \rightarrow i}(x)$$

(3)

其中, $\mathbf{d}(x)$ 表示在像素位置 x 处特征匹配的结果, \mathcal{W} 代表对齐操作, 与文献[30]中的方法类似, 本文方法采用双线性插值来实现。 \mathbf{d} 的尺寸为 $c_f \times h \times w$ 。

3.3 特征融合

在得到匹配的特征信息 \mathbf{d} 后, 本文通过以下公式得到融合的特征 \mathbf{k} 并将其用于清晰图像重建。

$$\mathbf{k} = \mathcal{F}(\mathcal{C}(\mathbf{d}_{i-2}, \mathbf{d}_{i-1}, \mathbf{f}_i, \mathbf{d}_{i+1}, \mathbf{d}_{i+2}))$$

(4)

其中, \mathcal{C} 表示按顺序拼接操作, \mathcal{F} 表示特征融合网络。

对于特征融合网络 \mathcal{F} , 本文采用 3 层卷积来实现, 3 层卷积中的滤波器数量分别为 512, 256, 128。前两层卷积后加有非线性激活函数 ReLU。每个卷积层中卷积核的尺寸为 3×3 。本模块的具体网络结构如图 3 所示。

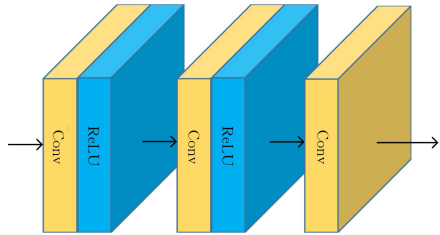


图 3 特征融合模块 \mathcal{F} 的网络结构图

Fig. 3 Architecture of feature fusion module \mathcal{F}

3.4 清晰帧重建

关于清晰帧重建网络 \mathcal{R} , 本文采用了一个比较常用的网络结构来预测中间帧超分辨率的残差图像。最终, 清晰视频帧通过以下公式计算得到:

$$H_i = \mathcal{R}(\mathbf{k}) + L_i^{\uparrow s}$$

(5)

其中, $L_i^{\uparrow s}$ 表示低分辨率图像序列中间帧上采样 s 倍的结果, 本文方法使用双线性插值来实现, H_i 表示中间帧重建结果。

对于清晰帧重建网络 \mathcal{R} , 本文使用了 40 个残差块, 2 个用于上采样(UpSample)的卷积层以及 2 个用于图像重建的卷积层。其中, 上采样使用 PixelShuffle^[14] 的方式。每个残差块

中的滤波器数量为 128, 用于上采样的 2 个卷积层中滤波器的数量均为 512, 用于图像重建的 2 个卷积层中滤波器数量分别为 128 和 3。每个卷积层中卷积核的尺寸为 3×3 。本模块的具体网络结构如图 4 所示。

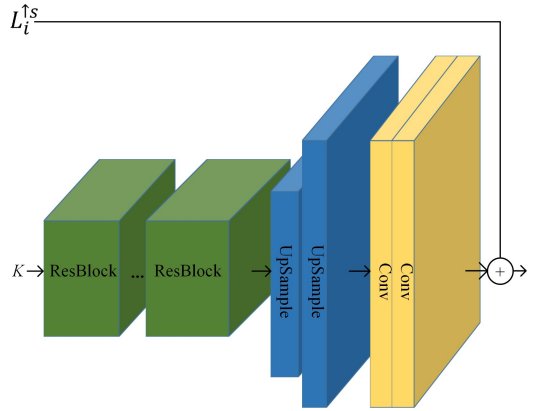


图 4 清晰帧重建模块 \mathcal{R} 的网络结构图

Fig. 4 Architecture of reconstruction module \mathcal{R}

3.5 损失函数

本文采用如下的损失函数来约束网络训练:

$$Loss = \|H_i - I_i\|_1$$

(6)

其中, $\|\cdot\|_1$ 表示 l_1 -范数, I_i 表示低分辨率图像 L_i 对应的真实高分辨率图像。

3.6 实现细节

3.6.1 数据集

本文采用 REDS 数据集^[31]训练出的网络。该数据集包含 240 个训练视频、30 个验证视频、30 个测试视频, 每个视频包含 100 帧图像, 每帧图像的分辨率为 720×1280 。同时, 本文使用与文献[16]类似的数据组织方式, 使用 236 个视频进行训练, 选择 4 个视频作为测试集(即 REDS4^[16])。

3.6.2 参数设置和训练细节

本文采用了与文献[16]相似的数据增强方法, 即随机对训练数据做水平、竖直翻转及旋转操作。网络训练过程中的批量大小设置为 8, 低分辨率图像块的大小为 64×64 。本文使用连续的 5 帧低分辨率图像作为网络的输入, 网络的输出为第 3 帧图像的超分辨率结果。

网络训练中采用文献[32]提出的 ADAM 优化器, 优化器中的参数取值为 $\beta_1 = 0.9$, $\beta_2 = 0.999$ 和 $\epsilon = 10^{-8}$ 。对于运动信息估计网络 \mathcal{N} , 本文采用预训练模型^[30]进行初始化, 初始学习率设置为 10^{-6} , 其他模块网络的初始学习率设置为 10^{-4} , 所有学习率每经过 100 步降低为前一学习率的 1/2 倍。整个网络联合训练, 共训练 500 步得到最后的超分辨率网络模型。

本文算法基于 PyTorch 框架实现, 使用 NVIDIA Tesla V100 GPU 进行训练。本文的训练代码和训练模型可从 OS-ID 码中获取。

4 本文方法分析与讨论

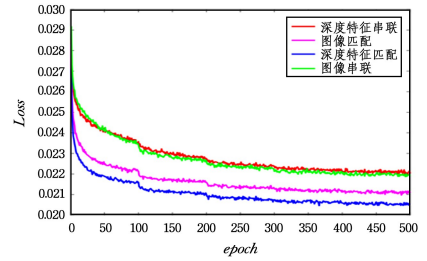
为了更好地理解本文方法的机理及每个模块对视频超分辨率的作用, 本节首先讨论深度特征匹配对于视频超分辨率

的作用,与图像匹配方法进行比较,验证本文提出的深度特征匹配方法的有效性;其次讨论运动估计模块 \mathcal{N} 相较于对视频超分辨率的影响;最后对深度特征提取网络做相关分析。本文采用了峰值信噪比(Peak Signal-to-Noise Ratio, PSNR)和结构相似性(Structural Similarity Image Measurement, SSIM)两个指标来衡量图像超分辨率结果的质量。表 1 列出了上述方法在 REDS4 测试集上的超分辨率结果的指标。图 5 给出了上述方法损失函数与 PSNR 的收敛情况。图 6 给出了上述方法的主观视觉效果。

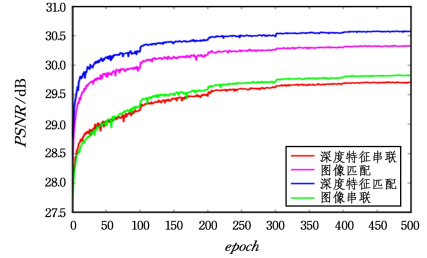
表 1 本文方法各模块在视频超分辨率中的作用

Table 1 Quantitative evaluations of different modules in the proposed method

方法	图像串联	深度特征串联	图像匹配	深度特征匹配
时空信息特征提取模块 \mathcal{T}	×	✓	×	✓
运动信息估计模块 \mathcal{N}	×	×	✓	✓
图像对齐操作 \mathcal{H}	×	×	✓	✓
清晰帧重建模块 \mathcal{R}	✓	✓	✓	✓
PSNR/SSIM	29.82/0.8520	29.71/0.8492	30.31/0.8652	30.55/0.8693



(a) Loss 对比图



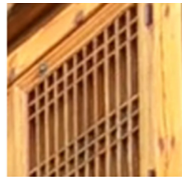
(b) PSNR 对比图

图 5 各方法的损失函数及 PSNR 的收敛曲线图

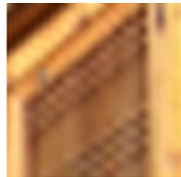
Fig. 5 Loss function and PSNR convergence curve of different methods



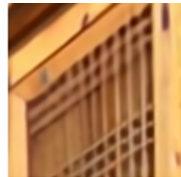
(a) 真实高分辨率图像



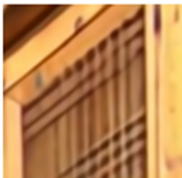
(b) 高分辨率图像块



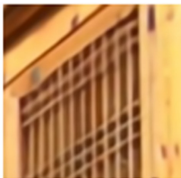
(c) bicubic



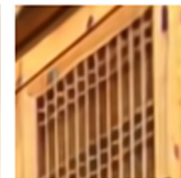
(d) 图像串联



(e) 深度特征串联



(f) 图像匹配



(g) 深度特征匹配

图 6 各方法在 REDS4 数据集上的视觉对比结果

Fig. 6 Visual comparisons of different methods on REDS4 dataset

4.1 深度特征匹配的作用

为了验证本文采用的深度特征匹配方法的有效性,本文将所提方法中的深度特征匹配换成图像空间的匹配(本文采用“图像匹配”表示这一基准方法)。从表 1 可以看出,采用深度特征匹配的方法(第 5 列)相比图像匹配的方法(第 4 列)在 PSNR 上提升了 0.24 dB。从图 6(f)和图 6(g)的比较结果可以看出,采用深度特征匹配的方法相比图像匹配的方法能更有效地恢复出视频帧的结构和细节。

4.2 运动信息估计的作用

为了验证运动估计对视频超分辨率问题的作用,本文与以下基准模型进行了比较,即将所提方法去掉运动信息估计模块 \mathcal{N} (本文采用“深度特征串联”表示这一基准方法)。由表 1 可以看出,采用运动估计的网络(第 5 列)比直接进行深度特征串联网络(第 3 列)的结果在 PSNR 上提高了 0.84 dB。由图 6(e)和图 6(g)的对比可以看出,本文方法的结果比深度特征串联网络具有更清晰的结构。

4.3 深度特征估计网络的验证

为了验证深度特征估计网络对视频超分辨率问题的作用,本文进行了以下实验,直接将低分辨率图像串联送入清晰帧重建网络 \mathcal{R} 中,恢复出高分辨率的中间帧(本文采用“图像串联”表示这一基准方法)。从表1可以看出,相比直接将低分辨率图像串联(第2列),采用运动估计并进行特征匹配(第5列)使得重建结果的PSNR提升了0.73 dB。从图6(d)和图6(g)的比较结果可以看出,采用运动估计并且进行特征匹配比单纯的串联输入视频帧更能有效地恢复出高质量的视频帧。

5 实验结果

为了说明本文模型的有效性,本节给出了与目前先进的超分辨率方法的比较结果。本文同大多数方法^[16,19-20,25]一样采用相同的标准,用4倍超分辨率的结果进行比较。通过采

用相同的测试集,并从客观指标PSNR,SSIM以及主观视觉效果上来分析本文模型的有效性。

5.1 定量比较结果

本文首先与目前常用的bicubic插值算法进行比较,然后与目前先进的视频超分辨率方法进行比较,其中包括DUF方法^[20]和TOFlow方法^[19]。除此之外,本文也与目前最先进的单幅图像超分辨率方法RCAN^[11]进行了比较。

表2列出了在REDS4数据集上关于4倍超分辨率方法的定量比较结果。从表中可以看出,本文方法恢复出了更高质量的视频帧。

图7给出了各种方法的可视化比较结果。由图7(c)可以看出,与RCAN相比,本文方法估计出了较好的结构细节;与图7(e)~图7(g)相比可知,本文提出的基于深度学习特征匹配的视频超分辨率方法估计出了较好的图像细节。

表2 本方法与现存方法在REDS4数据集上的客观指标

Table 2 Quantitative evaluations of the proposed method and existing methods on REDS4 dataset

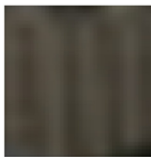
PSNR/SSIM	bicubic	RCAN	DUF	TOFlow	深度特征匹配
Clip000	24.07/0.6221	26.14/0.7346	27.29/0.7941	26.25/0.7479	27.65/0.8112
Clip011	25.49/0.7048	29.26/0.8244	28.34/0.8061	27.72/0.7846	31.49/0.8756
Clip015	27.98/0.7892	31.76/0.8867	31.48/0.8847	30.32/0.8543	33.47/0.9125
Clip020	24.80/0.7146	27.68/0.8279	27.28/0.8167	26.77/0.7929	29.60/0.8781
average	25.59/0.7077	28.71/0.8184	28.60/0.8254	27.77/0.7949	30.55/0.8693



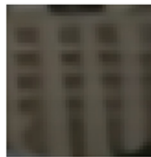
(a)真实高分辨率图像



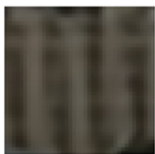
(b)高分辨率图像块



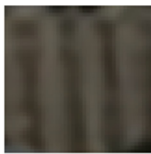
(c)bicubic



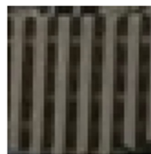
(d)RCAN



(e)DUF



(f)TOFlow



(g)深度特征匹配

图7 本文方法与现存方法在REDS4数据集上的视觉对比结果

Fig. 7 Visual comparisons of the proposed method and existing methods on REDS4 dataset

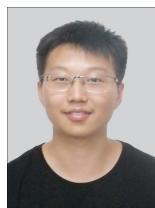
结束语 本文提出了一种基于深度学习特征匹配的视频超分辨率方法。该方法包括运动信息估计 \mathcal{M} 、时空信息特征提取 \mathcal{T} 、特征融合 \mathcal{F} 以及清晰帧重建 \mathcal{R} 模块。本文并没有设计非常复杂的网络结构与损失函数,而是挖掘了深度学习特征在视频超分辨率方法中的影响。不同于以往的大部分方法在图像空间中对视频相邻帧进行匹配,本文选择在特征空间中对相邻帧进行匹配,可以更充分地利用视频帧间的时序信

息,从而恢复出更好的清晰视频帧。本文展示了所提方法可以恢复出更清晰的图像以及更好的细节。最后,定性以及定量的实验结果证明,本文方法取得了较好的效果。

参考文献

- [1] FARSIU S, ROBINSON M D, ELAD M, et al. Fast and robust multiframe super resolution [J]. IEEE Transactions on Image Processing, 2004, 13(10): 1327-1344.
- [2] BASCLE B, BLAKE A, ZISSERMAN A. Motion deblurring and super-resolution from an image sequence [C] // Proceedings of the European Conference on Computer Vision. 1996: 571-582.
- [3] CAIN S C, HAYAT M M, ARMSTRONG E E. Projection-based image registration in the presence of fixed-pattern noise [J]. IEEE Transactions on Image Processing, 2001, 10(12): 1860-1872.
- [4] SHAHAR O, FAKTOR A, IRANI M. Space-time super-resolution from a single video [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2011: 3353-3360.
- [5] LIU C, SUN D. On Bayesian adaptive video super resolution [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(2): 346-360.
- [6] MA Z, LIAO R, TAO X, et al. Handling motion blur in multi-frame super-resolution [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 5224-5232.
- [7] DONG C, LOY C C, HE K, et al. Image super-resolution using deep convolutional networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(2): 295-307.
- [8] KIM J, KWON LEE J, MU LEE K. Accurate image super-reso-

- lution using very deep convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016;1646-1654.
- [9] DONG C, LOY C C, TANG X. Accelerating the super-resolution convolutional neural network[C]// Proceedings of the European Conference on Computer Vision. 2016;391-407.
- [10] HARIS M, SHAKHNAROVICH G, UKITA N. Deep back-projection networks for super-resolution[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018;1664-1673.
- [11] ZHANG Y, LI K P, LI K, et al. Image super-resolution using very deep residual channel attention networks[C]//Proceedings of the European Conference on Computer Vision. 2018;286-301.
- [12] LIM B, SON S, KIM H, et al. Enhanced deep residual networks for single image super-resolution[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2017;136-144.
- [13] LEDIG C, THEIS L, HUSZAR F, et al. Photo-realistic single image super-resolution using a generative adversarial network [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017;4681-4690.
- [14] SHI W, CABALLERO J, HUSZAR F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016;1874-1883.
- [15] TIAN Y, ZHANG Y, FU Y, et al. TDAN: Temporally-Deformable Alignment Network for Video Super-Resolution [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2020;3360-3369.
- [16] WANG X, CHAN K C K, YU K, et al. Edvr: Video restoration with enhanced deformable convolutional networks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2019;1954-1963.
- [17] LIU D, WANG Z, FAN Y, et al. Robust video super-resolution with learned temporal dynamics[C]// Proceedings of the IEEE International Conference on Computer Vision. 2017;2507-2515.
- [18] CABALLERO J, LEDIG C, AITKEN A, et al. Real-time video super-resolution with spatio-temporal networks and motion compensation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017;4778-4787.
- [19] XUE T, CHEN B, WU J, et al. Video enhancement with task-oriented flow [J]. International Journal of Computer Vision, 2019,127(8):1106-1125.
- [20] JO Y, OH S W, KANG J, et al. Deep video super-resolution network using dynamic upsampling filters without explicit motion compensation [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018;3224-3232.
- [21] HUANG Y, WANG W, WANG L. Bidirectional recurrent convolutional networks for multi-frame super-resolution[C]// Advances in Neural Information Processing Systems. 2015;235-243.
- [22] LIAO R, TAO X, LI R, et al. Video super-resolution via deep draft-ensemble learning [C]//Proceedings of the IEEE International Conference on Computer Vision. 2015;531-539.
- [23] KAPPELER A, YOO S, DAI Q, et al. Video super-resolution with convolutional neural networks[J]. IEEE Transactions on Computational Imaging, 2016,2(2):109-122.
- [24] TAO X, GAO H, LIAO R, et al. Detail-revealing deep video super-resolution[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017;4472-4480.
- [25] HARIS M, SHAKHNAROVICH G, UKITA N. Recurrent back-projection network for video super-resolution [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019;3897-3906.
- [26] WANG X, YU K, WU S, et al. Esrgan: Enhanced super-resolution generative adversarial networks [C]// Proceedings of the European Conference on Computer Vision Workshops. 2018;63-79.
- [27] BULAT A, YANG J, TZIMIROPOULOS G. To learn image super-resolution, use a gan to learn how to do image degradation first[C]//Proceedings of the European Conference on Computer Vision. 2018;185-200.
- [28] CHU M, XIE Y, LEAL-TAIXÉ L, et al. Temporally coherent gans for video super-resolution (tecogan) [J]. arXiv: 1811.09393, 2018,1(2):3.
- [29] LUCAS A, LOPEZ-TAPIA S, MOLINA R, et al. Generative adversarial networks and perceptual losses for video super-resolution[J]. IEEE Transactions on Image Processing, 2019,28(7):3312-3327.
- [30] SUN D, YANG X, LIU M Y, et al. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018;8934-8943.
- [31] NAH S, BAIK S, HONG S, et al. Ntire 2019 challenge on video deblurring and super-resolution: Dataset and study [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2019;1996-2005.
- [32] KINGMA D P, BA J. Adam: A method for stochastic optimization[J]. arXiv:1412.6980, 2014.



CHENG Song-sheng, postgraduate. His research interests include image super resolution, image/video enhancement and related vision problems.



PAN Jin-shan, professor. His main research interests include image deblurring, image/video analysis and enhancement, and related vision problems.