

# 基于增强特征对齐循环结构的视频超分辨

黄相丞, 刘恒

安徽工业大学计算机科学与技术学院, 安徽 马鞍山

收稿日期: 2023年6月10日; 录用日期: 2023年7月1日; 发布日期: 2023年7月12日

## 摘要

不同于单张图像超分辨, 视频超分辨需要时空联合考虑以处理连续的低分辨率视频从而获取清晰的高分辨率连续帧序列。当前, 随着深度学习在视频超分辨领域的广泛应用, 深度视频超分辨虽已取得显著效果, 但仍然存在视频特征信息挖掘不足的问题。特别是, 从时间维度看, 充分提取连续的视频时空特征并有效融合这些特征来实现帧内细腻、帧间稳定的视频超分辨, 仍然是当前视频超分辨率研究的主要问题。在这项工作中, 本文提出了一个基于增强特征对齐循环结构的视频超分辨网络。首先, 我们通过多分支的特征提取模块从不同的深度对输入的特征进行信息提取。其次, 在增强特征对齐循环结构中, 本文提出从当前帧的多个方向上同时融合相邻帧的信息, 并使用相应方向上的光流信息进行辅助对齐。最后, 本文提出在多个上采样尺度上对超分结果进行增强。实验结果表明, 所提出的方法能获得细节清晰、帧序列稳定的视频超分辨效果, 在定量的评估指标和定性的可视化结果等方面都超越了近些年来的其他先进方法。

## 关键词

视频超分辨, 增强特征对齐, 循环结构, 多尺度上采样

# Video Super-Resolution Based on Enhanced Feature Alignment with Recurrent Structure

Xiangcheng Huang, Heng Liu

School of Computer Science and Technology, Anhui University of Technology, Ma'anshan, Anhui

Received: Jun. 10<sup>th</sup>, 2023; accepted: Jul. 1<sup>st</sup>, 2023; published: Jul. 12<sup>th</sup>, 2023

## Abstract

Unlike single image super-resolution (SISR), video super-resolution (VSR) needs to be considered

spatially and temporally to process continuous low-resolution video to obtain clear high-resolution continuous frames. At present, with the extensive application of deep learning in the field of video super-resolution, deep video super-resolution has achieved remarkable results, but there is still a problem of insufficient video feature information mining. In particular, from the perspective of time dimension, it is still the main problem of current video super-resolution research to fully extract continuous video spatio-temporal features and effectively integrate these features to achieve video super-resolution that is exquisite within frames and stable between frames. In this work, an enhanced feature alignment recurrent structure for video super-resolution network is proposed. Firstly, the network extracts information from the input features from different depths through a multi-branch feature extraction module. Secondly, in the enhanced deformable feature alignment recurrent structure, we propose to fuse the information of adjacent frames from multiple directions of the current frame at the same time, and use the optical flow information in the corresponding directions to assist alignment. Finally, we propose to enhance the SR results on multiple upsampling scales. The experimental results show that the proposed method can achieve video super-resolution effect with clear details and stable frame sequence, which surpasses other advanced methods in recent years in terms of quantitative evaluation and qualitative visualization results.

## Keywords

Video Super-Resolution, Enhanced Feature Alignment, Recurrent Structure, Mutil-Scale Upsampling

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

视频超分辨是指从输入的低分辨视频中恢复出带有细节清晰和帧内连续的高分辨率图像帧序列, 是计算机视觉中的一项基本任务。视频超分技术在我们的日常生活中具有重要价值, 该技术被广泛应用于高清视频播放、视频监督和刑侦[1]等领域。由于连续的图像帧之间具有高度的相关性, 解决不同帧的时空特征的充分融合仍然是当前视频超分辨的主要挑战。

在早期的基于传统方法的视频超分辨算法中, Schultz 等人[2]通过仿射模型去简单地估计帧间的运动信息。Liu 等人[3]提出了一种贝叶斯方法, 用于同时估计潜在的运动、模糊核和噪声水平, 并重建高分辨率帧。Ma 等人[4]采用了期望最大化方法估计模糊核, 指导高分辨率帧的重建。然而, 这些不能够适应视频中的各种场景, 没有很好的泛化性。

近年来, 随着深度学习的快速发展, 基于卷积神经网络的视频超分辨算法能够在庞大的数据集上学习低分辨率视频和高分辨率视频之间复杂的映射关系。与传统方法相比, 这些方法显著提高了性能。当前, 这些基于深度学习去实现视频超分算法的框架大体可以分成两类: 一类是基于滑动窗口的方法, 另一类是基于循环神经网络的方法。基于滑动窗口的方法, 如基于时空相关性的视频超分辨率重建算法[5]、VESPCN [6]、TDAN [7]、EDVR [8]等方法, 通常先在滑动窗口中通过空间卷绕的方式将所有的输入帧显式或隐式地对齐到中间帧上, 然后进行信息融合, 最后输出中间帧的超分辨率结果。这种方法的优势是能够比较充分地利用所有的输入帧的信息, 并将其增强到中间帧上, 达到局部最好的效果。但是由于可以利用的视频帧的数量有限, 当在中间帧附近发生遮挡等特殊情况, 由于不能借鉴到有用信息, 则会导

致生成的效果较差。

与上述方法不同的是, 基于递归的方法[9]一般输入比较长, 并且能够一次输出所有的输入帧的超分结果。在处理时, 模型会递归地处理每一帧图像, 并将本次的输出应用到下一次的迭代过程中。为了使局部信息能够整个时间序列上进行传递, 后续的算法[10]中又提出了双向递归的架构——首先从后向前地进行一次循环, 再从前向后进行处理。该方法可以实现在输入数量较多具有更好的性能。但是由于在每次的迭代过程中只利用了来自单方向的上一帧的输出结果, 在输入帧数较少时效果反而不如基于滑动窗口的方法。

当然, 除了使用的框架会影响视频超分辨的生成结果外, 在框架之内, 如何更加有效的利用周边帧的信息也是决定视频超分结果好坏的一个重要因素。常见的一种做法是将相邻帧对齐到当前要处理的帧上。早期的算法中使用光流来实现帧之间的对齐。Kappeler 等人在 VSRNet [11]中第一次使用光流来实现运动补偿然后使用卷积神经网络来重建高分辨率帧。此后, Caballero 等人在 VESPCN [6]提出了一个端到端的方案来完成视频超分辨任务, 即将帧对齐和重建视为超分网络的一个组件。Xue 等人在 TOFlow [12]中指出, 在网络中同时学习光流模块的参数来适应一个特定的任务的做法比先学习光流网络, 然后将学习好的光流网络直接应用到超分辨网络中这种做法的性能要好。吴等人[13]提出对高分辨率的视频帧与预测的高分辨率光流进行运动补偿, 并将补偿结果输入到超分辨率融合网络来得到更好的效果。储等人[14]创新地将注意力机制和光流结合在一起, 提出双阶段特征对齐思路, 分别处理微小运动信息和幅度较大的运动信息。但是如果运动过程中发生了遮挡或者运动较快, 这时候就很难估计出准确的光流, 进而影响到超分辨输出结果。

另一种对齐的做法是通过可变形卷积去实现。Tian 等人在 TDAN [7]中第一次提出了在特征层面通过可变形卷积来实现帧对齐。之后, 在此基础上, Wang 等人在 EDVR [8]提出了多级的可变形卷积对齐并设计了时空注意力模块来融合这些对齐的特征同时通过注意力机制减少错误的对齐情况。

其他的做法是做隐式的对齐或者不进行对齐直接处理。Zhu 等人[15]设计了一个轻量的残差可逆块并提出了一个稀疏的特征融合机制去结合时空特征, 然后重建出最后的视频帧。Li 等人在 MuCAN [16]中提出了一种时间多重对应聚合策略去利用跨帧的相似补丁, 并设计了跨尺度非局部对应聚合方案, 以探索帧内不同尺度的自相似性。刘等人[17]提出了一个双分支网络, 通过引入梯度信息来避免超分辨率图像出现结构失真等问题。

Chan 等人在 BasicVSR [18]中将递归神经网络和光流对齐相结合, 并提出了一种轻量级的基准模型, 在输入帧数较多时实现了优异的性能。但是当输入帧数较少时, 比如说以 5 帧作为输入进行训练, 则其模型的性能将会下降很多。Cao 等人在 VSR-T [19]中提出了时空卷积自注意力层去探索局部信息并通过双向光流前馈层对齐图像帧, 实现了在滑动窗口方法中性能最优。但是其庞大的参数量, 也导致它难以更多的帧数进行训练。针对以上这些问题, 本文提出了一个基于增强特征对齐循环结构的视频超分辨网络。对于视频超分辨中空间信息丢失和不同帧的时空特征难以充分地利用的问题, 所提出的增强特征对齐循环结构可以有效地将相邻帧的信息对齐当前帧上, 其次, 该结构可以让帧间信息在时序上进行长距离传播。同时, 所提出的多尺度上采样模块, 分别从两个超分尺度上对融合特征进行了特征增强, 提升了重建视频的结果。

综上, 本文提出了一个可以应用于视频超分辨任务的新算法, 提出了一个增强特征对齐循环结构, 结合光流和可变形卷积, 可以更好地聚合相邻帧的信息; 其次提出了一个多尺度的上采样模块, 可以进一步提升输入的超分辨率帧的质量。实验结果证明了所提出的方法的可行性, 与现有的方法在两种不同的训练帧数下训练的模型的结果比较都证明了所提方法的有效性。

## 2. 网络模型架构

### 2.1. 模型架构

本节将描述所提出的基于增强特征对齐循环结构的视频超分辨率网络的结构细节。如图 1 所示, 网络的主要模块包括特征提取模块、增强特征对齐循环模块和多尺度上采样模块。将输入网络的低分辨率视频序列记为  $I_{LR}$ , 输出的高分辨率视频序列记为  $I_{SR}$ 。首先, 网络会在特征提取模块对输入的视频序列进行特征提取, 该过程可用(1)表示:

$$F = f_{extract}(I_{LR}) \tag{1}$$

其中,  $f_{extract}$  表示特征提取操作,  $F$  表示提取的中间特征, 被提取的连续序列特征  $F$  接下来将被送到增强的增强特征对齐循环模块中。该模块将递归的对每一帧特征进行对齐和融合, 并将融合后的结果应用到下一次的递归过程中。最后在多尺度上采样模块中, 之前输出的融合结果将会与输入图像进行不同尺度的特征融合, 最终通过一层卷积在 4 倍的尺度上输出每一帧的超分结果。

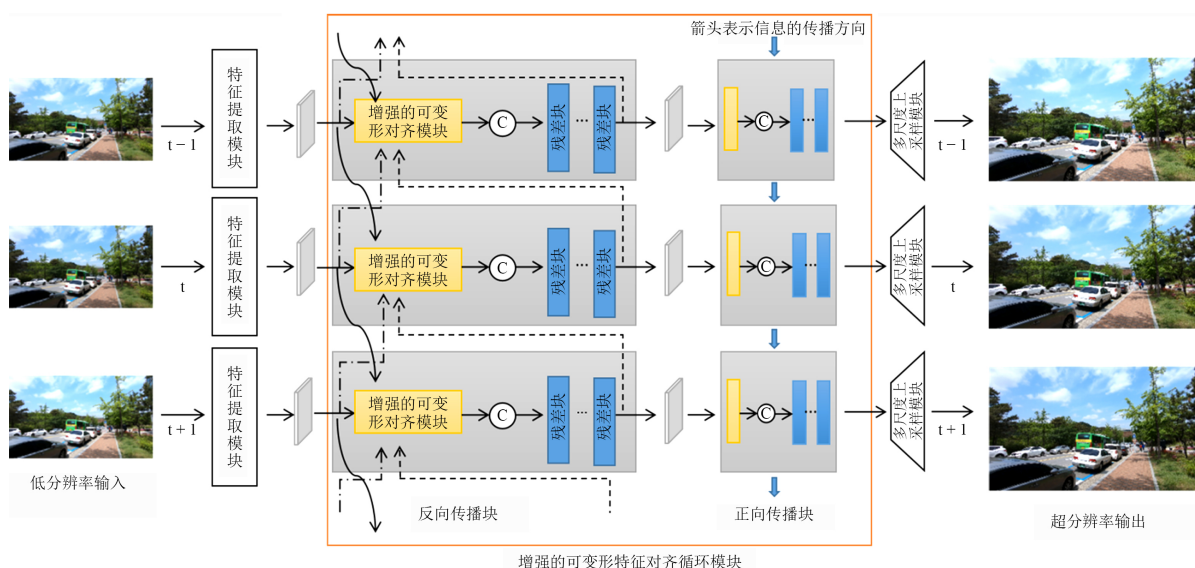


Figure 1. Overall architecture of the network  
图 1. 整体网络架构图

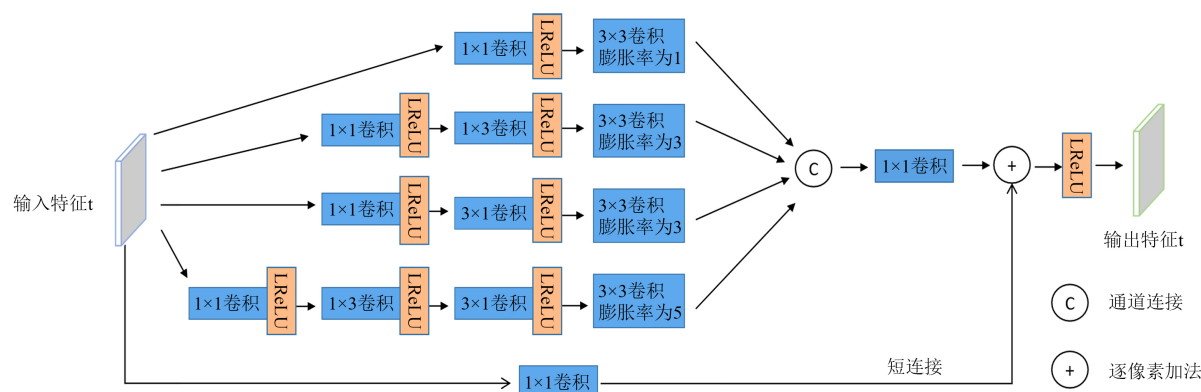


Figure 2. The structure of extractor module  
图 2. 特征提取模块结构图

## 2.2. 特征提取模块

所提出的特征提取模块如图 2 所示, 该模块通过 5 个分支从不同的深度对特征进行信息提取。输入的图像帧序列首先会通过一个  $3 \times 3$  的卷积层, 通过卷积操作从图像层面转换到特征层面, 通道数由 3 通道变成 64 通道。之后, 特征序列将会依次经过所提出的特征提取模块。以第  $t$  帧特征为例, 该输入特征将会从四个分支同时进行卷积, 这里使用了四个分支是考虑了通过不同尺度的感受野来提升特征的多样性。结构中最后会将不同膨胀率的卷积层输出进行通道连接, 以达到融合不同特征的目的。进行通道连接后的特征会经过  $1 \times 1$  卷积进行特征融合, 之后参考残差网络的结构, 与输入特征经过  $1 \times 1$  后的结果进行短连接, 相加之后经过 LReLU 激活层, 得到最后的输出特征。

## 2.3. 增强特征对齐循环模块

所提出的增强特征对齐循环模块可分为反向传播块和正向传播块。其操作一致, 只是在信息的传播方向上相反。这里以反向传播块为例, 如图 1 所示。假设当前处理的是第  $t$  帧特征, 其首先会与上一轮的递归输出特征  $t+1$  以及其相邻特征  $t-1$  做可变形卷积, 以实现输出的  $t+1$  帧特征和相邻的  $t-1$  帧特征对齐到当前特征  $t$  上。

对齐的过程如图 3 所示, 具体细节为: 记当前帧特征为  $F_{ref}$ , 要对齐的相邻特征为  $F_{nbr}$ ,  $F_{nbr}$  到  $F_{ref}$  的光流记为 flow, 该光流是通过 SpyNet [20] 网络进行估计的, 则首先模型会通过一个  $3 \times 3$  的卷积对特征  $F_{ref}$  进行卷积偏移, 得到输出记为 out, 然后在通道维度将 out 等分成 3 份, 分别记为  $O_1$ 、 $O_2$  和 mask。其中, 所提方法会对  $O_1$  和  $O_2$  进行通道连接, 并将连接结果与光流相加, 得到偏移量 offset。而 mask 则会先经过 Sigmoid 函数, 其结果会与  $F_{nbr}$  和 offset 一起送到基本的可变形卷积[21]中, 最后得到对齐的输出特征  $F_{aligned}$ 。

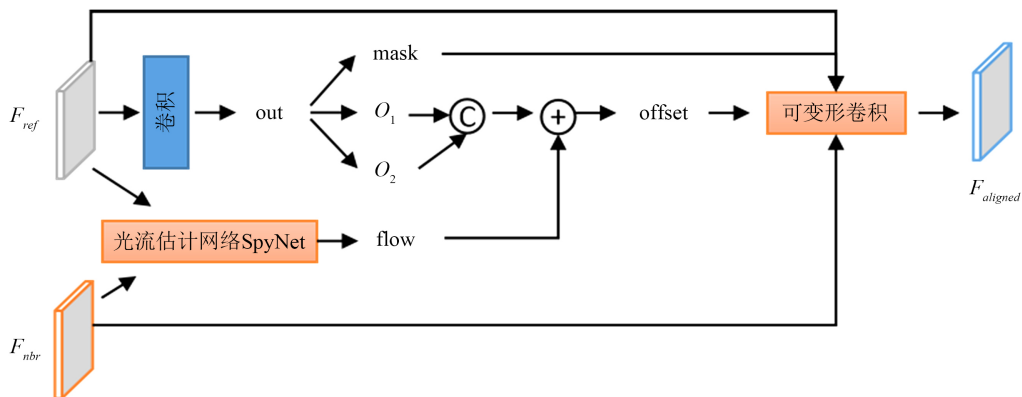


Figure 3. The enhanced feature alignment module

图 3. 增强特征对齐模块

在分别得到输出的  $t+1$  帧特征和相邻的  $t-1$  帧特征对齐到当前特征  $t$  上的对齐特征之后, 该方法会将图像  $I_t$ , 特征  $F_t$ , 以及得到的 2 个对齐特征进行通道连接, 再通过一个  $3 \times 3$  的卷积和 ReLU 非线性层进行特征融合后, 送到 30 层的残差块中, 得到当前第  $t$  帧的传播输出  $Prop_t$ , 并将该输出应用到下一次的递归过程中。

## 2.4. 多尺度上采样模块

所提出的多尺度上采样模块如图 4 所示, 可分为 2 倍和 4 倍两个上采样尺度。对当前的输入特征  $F_t$ ,

所提出的方法首先会对其进行 2 倍的 pixelshuffle [22] 上采样, 此时得到的中间特征  $M_t$  的大小将是输入特征  $F_t$  的 2 倍。同时, 网络会对输入的图像  $I_t$  进行 2 倍的线性上采样, 并将上采样的结果与中间特征  $M_t$  进行通道连接, 然后通过  $1 \times 1$  卷积进行融合。在第二层尺度上, 该方法会将融合后的结果进行 2 倍的 pixelshuffle [22] 上采样得到输出特征, 此时得到的特征图大小变成输入特征  $F_t$  的 4 倍, 然后通过卷积层将所得到的输出特征由 64 通道变成 3 通道, 转换到 RGB 图像层面, 得到输出结果  $O_t$ 。最后对图像  $I_t$  进行 4 倍的线性上采样, 并将上采样的结果与输出结果  $O_t$  相加, 得到最后的超分结果  $S_t$ 。

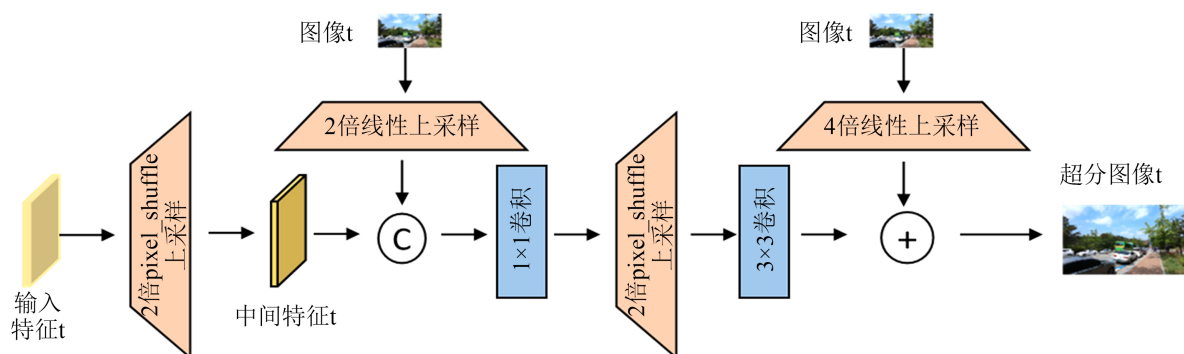


Figure 4. The multi-scale upsampling module  
图 4. 多尺度上采样模块

## 2.5. 损失函数

实验中所使用的重建损失为 Charbonnier [23] 损失, 定义如下:

$$l_{rec} = \sqrt{\|GT - SR\|^2 + \varepsilon^2} \quad (2)$$

其中,  $GT$  表示高分辨视频帧,  $SR$  表示网络模型输出的超分辨率视频帧, 设为  $1 \times 10^{-8}$ 。

## 3. 实验结果与分析

本节将首先介绍所使用的数据集。在 3.2 节, 将介绍本文所提出的视频超分网络模型的实验细节。在 3.3 节, 将展示所提出的方法与当前其他的先进方法的定性和定量对比。

### 3.1. 数据集

REDS [24] 是 2019 年 NTIRE 比赛中提出的并在此后被广泛应用的数据集。它包含了 240 个训练序列、30 个验证序列和 30 个测试序列, 其中每个序列中包含 100 个连续的图像帧。与 EDVR 相同, 在训练中本文使用 REDS4 [24] 数据集作为测试集, 即选择训练集中的第 000、011、015 和 020 四个序列作为测试集, 其余的序列都用来训练。

### 3.2. 实施细节

训练中, 本文分别使用了 15 帧和 5 帧连续的图像帧作为网络模型的输入来训练模型参数。输入的图像序列大小均为  $64 \times 64$ , 经过 4 倍的超分之后, 对应的输出序列的图像大小为  $256 \times 256$ 。图像的输入通道均为 3 通道, 经过特征提取模块之后, 特征通道将变成 64。所使用的 SpyNet [20] 的方式同 BasicVSR [18]。本文采用 Adam 优化器 [25] 和余弦退火机制 (Cosine Annealing scheme) [26] 来优化网络参数。网络的初始学习率设置为  $4 \times 10^{-4}$ 。训练中设置的批处理大小是根据所使用的 GPU 内存而设置的, 因为所使用的显卡为 2 张显存大小为 12G 的 RTX 1080Ti, 所以这里每张卡上的批处理大小分别设置为 4, 所使用的深度模

型框架为 Pytorch。至于损失函数, 和先前的模型一样, 因为 Charbonnier [23] 损失比损失和损失具有更好的效果, 因此本文选择其作为训练过程的损失函数。关于评价指标, 本文选择峰值信噪比(PSNR)和结构相似性指数 SSIM [27] 来衡量超分结果和真实的图像序列之间的差异。

### 3.3. 对比实验

在这一部分, 本文首先展示了在测试集 REDS4 [24] 上的定量比较结果。为了体现所提出的方法的优越性, 本文与以下的先进方法进行了比较: TOFlow [12]、DUF [28]、EDVR [8]、PFNL [29]、MuCAN [16]、BasicVSR [18]、IconVSR [18] 和 VSR-T [19]。表中所示的参数量的单位为百万(M), “-”表示该值无法获得; PSNR 和 SSIM 指标越高代表模型的性能越好, 其中最好的结果和第二好的分别用红色和蓝色表示。本文首先是用 15 帧连续图像进行训练并比较, 从表 1 中可以看出, 所提出的方法在 PSNR 和 SSIM [27] 指标上基本上都是最优的。为了进一步验证所提出的方法在低帧输入场景下的性能, 本文也用 5 帧图像作为输入进行了训练。定量对比如表 2 所示, 从表中可以看出, 即使以 5 帧作为输入进行训练, 本文的方法依旧拥有良好的性能。与此相比, BasicVSR [18] 和 IconVSR [18] 的性能就下降了很多, 甚至都没有 EDVR [8] 和 VSR-T [19] 高。

**Table 1.** Quantitative comparison (PSNR/SSIM) on REDS4 for  $4 \times$  VSR with a training input of 15 frames. The higher the PSNR and SSIM, the better the performance. The results are tested on RGB channels. Red and blue indicate the best and the second best performance with the same training frame number, respectively. “-” means unavailable

**表 1.** 训练输入为 15 帧的 4 倍视频超分辨的定量比较(PSNR/SSIM)。PSNR 和 SSIM 越高, 代表模型的性能越好。所有的测试均是在 RGB 通道上进行的。红色和蓝色的数字分别代表相同训练帧数下最好的和第二好的性能。“-”代表无法得到的数据

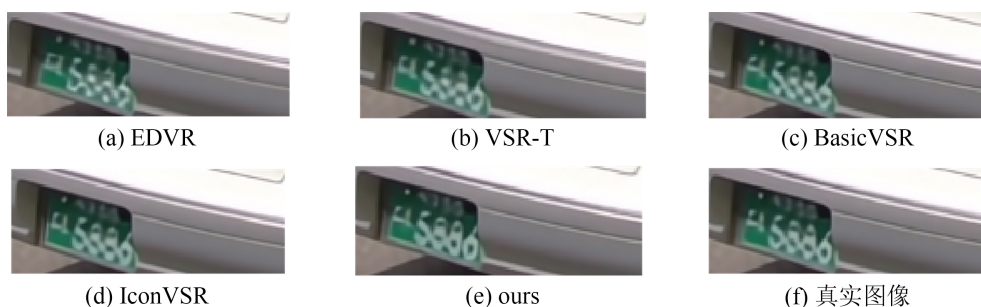
方法	参数量	序列 000	序列 011	序列 015	序列 020	平均值
BasicVSR	6.3	28.40/0.8419	32.44/0.8966	34.19/0.9235	30.54/0.8986	31.42/0.8909
IconVSR	8.7	28.55/0.8478	32.89/0.9024	34.54/0.9270	30.79/0.9033	31.69/0.8951
<b>ours</b>	6.8	28.58/0.8486	32.85/0.9026	34.56/0.9267	30.85/0.9039	31.71/0.8954

**Table 2.** Quantitative comparison (PSNR/SSIM) on REDS4 for  $4 \times$  VSR with a training input of 5 frames. The higher the PSNR and SSIM, the better the performance. The results are tested on RGB channels. Red and blue indicate the best and the second best performance with the same training frame number, respectively. “-” means unavailable

**表 2.** 训练输入为 5 帧的 4 倍视频超分辨的定量比较(PSNR/SSIM)。PSNR 和 SSIM 越高, 代表模型的性能越好。所有的测试均是在 RGB 通道上进行的。红色和蓝色的数字分别代表相同训练帧数下最好的和第二好的性能。“-”代表无法得到的数据

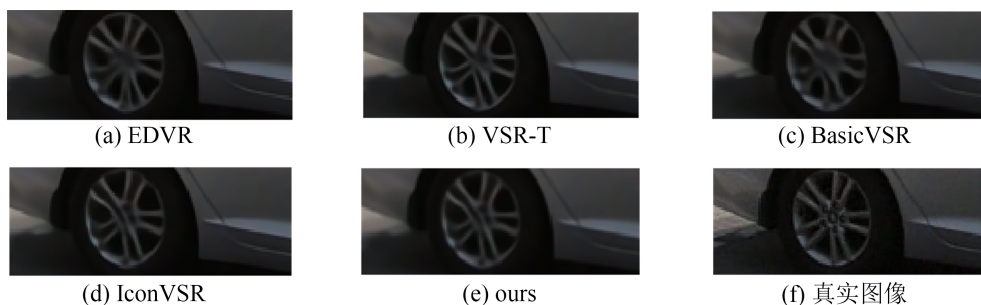
方法	参数量	序列 000	序列 011	序列 015	序列 020	平均值
Bicubic	-	24.55/0.6489	26.06/0.7261	28.52/0.8034	25.41/0.7386	26.41/0.7279
TOFlow	-	26.52/0.7540	27.80/0.7858	30.67/0.8609	26.92/0.7953	27.98/0.7990
DUF	5.8	27.30/0.7937	28.38/0.8056	31.55/0.8846	27.30/0.8164	28.63/0.8251
EDVR-L	20.6	28.01/0.8250	32.17/0.8864	34.06/0.9206	30.09/0.8881	31.09/0.8800
BasicVSR	6.3	27.67/0.8114	31.27/0.8740	33.58/0.9135	29.71/0.883	30.56/0.8698
IconVSR	8.7	27.83/0.8182	31.69/0.8798	33.81/0.9164	29.90/0.8841	30.81/0.8746
VSR-T	32.6	28.06/0.8267	32.28/0.8883	34.15/0.9199	30.26/0.8912	31.19/0.8815
<b>ours</b>	6.8	28.27/0.8388	32.10/0.8925	34.07/0.9201	30.47/0.8969	31.22/0.8871

此外, 定性结果比较可以在图 5 和图 6 中看到, 图中显示了其他的可获得训练模型的先进方法和本文所提出的方法与真实图像之间的在 REDS4 [24] 部分测试集上细节的比较, 所比较的序列分别为序列 000 和序列 015。可以明显的看出, 本文所提出的方法与其他方法相比恢复出了更多的纹理细节, 在数字“5886”以及车轮等处的细节恢复中都更加地接近真实图像。



**Figure 5.** Comparison of qualitative effects with other SOTA methods on REDS4 sequence 000

**图 5.** 与其他 SOTA 方法在 REDS4 序列 000 上定性效果的对比



**Figure 6.** Comparison of qualitative effects with other SOTA methods on REDS4 sequence 015s

**图 6.** 与其他 SOTA 方法在 REDS4 序列 015 上定性效果的对比

## 4. 结论与展望

本文提出了一个基于增强特征对齐循环结构的视频超分辨率网络, 通过利用增强特征对齐循环结构和多尺度上采样模块, 完成了视频超分辨率的任务。实验的定量结果和定性比较, 都证明了所提出的网络的可行性及其性能。在未来, 我们的研究还将集中在使模型更加轻量化上, 以便模型可以部署在移动端, 使研究内容更加实用化。

## 基金项目

安徽省高校协同创新项目(GXXT-2022-044)。

## 参考文献

- [1] 陈艳, 金伟其, 王岭雪, 等. 基于视频序列的超分辨率视频图像处理软件在刑侦上的应用[J]. 警察技术, 2011(1): 24-26.
- [2] Schultz, R.R. and Stevenson, R.L. (1996) Extraction of High-Resolution Frames from Video Sequences. *IEEE Transactions on Image Processing*, **5**, 996-1011. <https://doi.org/10.1109/83.503915>
- [3] Liu, C. and Sun, D. (2013) On Bayesian Adaptive Video Super Resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **36**, 346-360. <https://doi.org/10.1109/TPAMI.2013.127>
- [4] Ma, Z., Liao, R., Tao, X., et al. (2015) Handling Motion Blur in Multi-Frame Super-Resolution. *Proceedings of the*



- IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 7-12 June 2015, 5224-5232.
- [5] 李金航, 肖亮, 韦志辉. 基于时空相关性的视频超分辨率重建算法[J]. 计算机与数字工程, 2020, 48(4): 946-951.
- [6] Caballero, J., Ledig, C., Aitken, A., *et al.* (2017) Real-Time Video Super-Resolution with Spatio-Temporal Networks and Motion Compensation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 2848-2857. <https://doi.org/10.1109/CVPR.2017.304>
- [7] Tian, Y.P., Zhang, Y.L., Fu, Y. and Xu, C.L. (2020) TDAN: Temporally-Deformable Alignment Network for Video Super-Resolution. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 13-19 June 2020, 3357-3366. <https://doi.org/10.1109/CVPR42600.2020.00342>
- [8] Wang, X.T., Chan, K.C.K., Yu, K., Dong, C. and Loy, C.C. (2019) EDVR: Video Restoration with Enhanced Deformable Convolutional Networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Long Beach, 16-17 June 2019, 1954-1963. <https://doi.org/10.1109/CVPRW.2019.00247>
- [9] 刘佳, 安鹤男, 李蔚, 等. 基于帧循环网络的视频超分辨率技术[J]. 电子技术应用, 2020, 46(9): 43-46.
- [10] Huang, Y., Wang, W. and Wang, L. (2017) Video Super-Resolution via Bidirectional Recurrent Convolutional Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **40**, 1015-1028. <https://doi.org/10.1109/TPAMI.2017.2701380>
- [11] Kappeler, A., Yoo, S., Dai, Q.Q. and Katsaggelos, A.K. (2016) Video Super-Resolution with Convolutional Neural Networks. *IEEE Transactions on Computational Imaging*, **2**, 109-122. <https://doi.org/10.1109/TCI.2016.2532323>
- [12] Xue, T., Chen, B., Wu, J., *et al.* (2019) Video Enhancement with Task-Oriented Flow. *International Journal of Computer Vision*, **127**, 1106-1125. <https://doi.org/10.1007/s11263-018-01144-2>
- [13] 吴昊, 赖惠成, 钱绪泽, 等. 基于光流残差的视频超分辨率重建算法[J]. 计算机工程与应用, 2022, 58(15): 220-228.
- [14] 储岳中, 乔雨楠. 多注意力结合光流的视频超分辨方法[J]. 重庆工商大学学报(自然科学版), 2022, 39(4): 1-8.
- [15] Zhu, X., Li, Z., Zhang, X.Y., *et al.* (2019) Residual Invertible Spatio-Temporal Network for Video super-Resolution. *Proceedings of the AAAI Conference on Artificial Intelligence*, **33**, 5981-5988. <https://doi.org/10.1609/aaai.v33i01.33015981>
- [16] Li, W., Tao, X., Guo, T., *et al.* (2020) MuCAN: Multi-Correspondence Aggregation Network for Video Super-Resolution. In: Vedaldi, A., Bischof, H., Brox, T. and Frahm, J.M., Eds., *ECCV 2020: Computer Vision—ECCV 2020*, Springer, Cham, 335-351. [https://doi.org/10.1007/978-3-030-58607-2\\_20](https://doi.org/10.1007/978-3-030-58607-2_20)
- [17] 刘翠微, 丁彩胤, 李侃, 等. 基于时空注意力的双分支视频超分辨率网络[J]. 沈阳航空航天大学学报, 2022, 39(1): 45-53.
- [18] Chan, K.C.K., Wang, X.T., Yu, K., Dong, C. and Loy, C.C. (2021) BasicVSR: The Search for Essential Components in Video Super-Resolution and Beyond. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, 20-25 June 2021, 4945-4954. <https://doi.org/10.1109/CVPR46437.2021.00491>
- [19] Cao, J.Z., Li, Y.W., Zhang, K., *et al.* (2021) Video Super-Resolution Transformer. arXiv: 2106.06847.
- [20] Ranjan, A. and Black, M.J. (2017) Optical Flow Estimation Using a Spatial Pyramid Network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 2720-2729. <https://doi.org/10.1109/CVPR.2017.291>
- [21] Dai, J.F., Qi, H.Z., Xiong, Y.W., *et al.* (2017) Deformable Convolutional Networks. *Proceedings of the IEEE International Conference on Computer Vision*, Venice, 22-29 October 2017, 764-773. <https://doi.org/10.1109/ICCV.2017.89>
- [22] Shi, W., Caballero, J., Huszár, F., *et al.* (2016) Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 1874-1883. <https://doi.org/10.1109/CVPR.2016.207>
- [23] Charbonnier, P., Blanc-Feraud, L., Aubert, G. and Barlaud, M. (2019) Two Deterministic Half-Quadratic Regularization Algorithms for Computed Imaging. *Proceedings of 1st International Conference on Image Processing*, Austin, 13-16 November 1994, 168-172.
- [24] Nah, S., Baik, S., Hong, S., *et al.* (2019) Ntire 2019 Challenge on Video Deblurring and Super-Resolution: Dataset and Study. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Long Beach, 16-17 June 2019, 1996-2005. <https://doi.org/10.1109/CVPRW.2019.00251>
- [25] Kingma, D.P. and Ba, J. (2014) Adam: A Method for Stochastic Optimization. arXiv: 1412.6980.
- [26] Loshchilov, I. and Hutter, F. (2016) SGDR: Stochastic Gradient Descent with Warm Restarts. arXiv: 1608.03983.
- [27] Wang, Z., Bovik, A.C., Sheikh, H.R. and Simoncelli, E.P. (2004) Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, **13**, 600-612. <https://doi.org/10.1109/TIP.2003.819861>

- 
- [28] Jo, Y., Oh, S.W., Kang, J. and Kim, S.J. (2018) Deep Video Super-Resolution Network Using Dynamic Upsampling Filters without Explicit Motion Compensation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 3224-3232. <https://doi.org/10.1109/CVPR.2018.00340>
- [29] Yi, P., Wang, Z., Jiang, K., Jiang, J.J. and Ma, J.Y. (2019) Progressive Fusion Video Super-Resolution Network via Exploiting Non-Local Spatio-Temporal Correlations. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Seoul, 27 October-2 November 2019, 3106-3115. <https://doi.org/10.1109/ICCV.2019.00320>