基于三维时空注意的密集连接视频超分算法*

何啸林，吴丽君
（福州大学 物理与信息工程学院，福建 福州 350116）

摘 要：针对视频超分对时间帧间信息以及分层信息的利用不充分，设计了一种具有空间时序注意力机制的密集可变形视频超分辨率重建网络。利用三维卷积来提取经可变形卷积模块对齐后的相邻帧之间的时间序列信息，同时设计具有步幅卷积层的轻量级模块来提取空间注意力信息。在特征重构阶段引入密集连接，充分利用分层特征信息以实现更好的特征重建。选取公共数据集进行实验验证，结果表明，提出的算法在客观评价指标与视觉对比效果上都有提升。。

关键词：视频超分辨重建；三维时空注意力机制；可变形卷积；密集连接
中图分类号：TP391　　**文献标识码**：A　　**DOI**：10.19358/j.issn.2097-1788.2023.02.011
引用格式：何啸林，吴丽君. 基于三维时空注意的密集连接视频超分算法[J].网络安全与数据治理，2023，42(2)：70-75.

Densely connected video super-resolution based on three-dimensional spatial-sequential attention

He Xiaolin，Wu Lijun
（College of Physics and Information Engineering，Fuzhou University，Fuzhou 350116，China）

Abstract：Aiming at the insufficient utilization of temporal inter-frame information and hierarchical information in video super-resolution，a dense deformable video super-resolution reconstruction network with spatial-sequential attention mechanism is designed. Three-dimensional convolution is used to extract sequence information between adjacent frames aligned by deformable convolution module，and a lightweight module with strided convolution layer is designed to extract spatial attention information. Dense connections are introduced in the feature reconstruction stage to make full use of hierarchical feature information to achieve better feature reconstruction. The public datasets are selected for experimental verification. The results show that the proposed algorithm has improved both objective evaluation indicators and visual contrast effects.
Key words：video super-resolution；three-dimensional spatial-sequential attention；deformable convolution；dense connection

0 引言

视频超分辨(Video Super-Resolution，VSR)算法是一项具有挑战性的课题，倍受人们的关注。相较于单图像的超分辨率重建，视频超分辨率重建可以利用帧之间的相关性和连续帧间的时间信息。视频超分的目标是在相邻的低分辨率帧(Low Resolution，LR)的帮助下，重建出高分辨率帧(High Resolution，HR)。早期的研究[1-3]将视频超分视为图像重建的简单扩展，并没有考虑到物体运动，性能较差。对此，人们开始研究一些显式运动补偿的方法，最为广泛的是使用光流来估计帧之间的运动并执行变形。然而，对光流进行准确的预测是比较困难的，尤其是在存在遮挡或大运动时，当对光流量的不准确预测时可能会引入伪影[4]。为了解决这个问题，研究人员开始研究隐式运动补偿方法。在隐式补偿方法中，可变形卷积较为常用[5]。时序可变形对齐视频超分网络(Temporally Deformable Alignment Network，TDAN)[4]首次将可变形卷积引入视频超分任务中；增强型可变形卷积视频超分网络(Video Restoration with Enhanced Deformable Convolutional Networks，EDVR)[6]将跨帧信

* 基金项目：福建省自然科学基金(2022H0008，2021J01580)；福州市科技计划(2021-P-030，2021-P-059)

息与可变形网络和注意力机制融合在一起。相比光流法,可变形卷积的方法解决了伪影问题,但注意力机制的设计仍有改进空间。对于连续帧的视频任务,视频的序列信息是至关重要的。由于在时间注意力模块中仅仅采用二维卷积,无法提取时间序列维度的信息,以往方法中的时空注意力模块仅仅只是在两帧之间进行自注意力加权。

本文设计了一种具有三维空间顺序注意机制的密集可变形视频超分辨率重建网络。在视频帧对齐模块之后引入空间时序注意力模块,利用三维卷积操作来捕获帧间序列信息。在超分任务中,引入空间注意力中金字塔结构使得网络能够获得更大的感受野,但也带来了冗余参数。本文通过几个卷积层和池化层的组合来重新设计空间注意模块,利用更少的参数保持一个大的感受野。此外,为了在特征重建阶段充分利用分层特征,设计了一个由密集连接和残差组成的密集连接重建模块。

综上所述,本文设计了一种三维空间时序注意力机制。应用三维卷积来获取时间注意模块中的帧间序列信息。在空间注意力模块中,修改卷积的步长,使用卷积组结合池化来实现轻量化。同时设计密集连接重建模块,通过密集连接充分利用分层特征信息,更好地完成特征重建。

1 基于深度学习的视频超分辨率重建

自 Dong 等人[7]建立超分卷积神经网络(Super Resolution Convolutional Neural Network,SRCNN)工作以来,深度学习的方法显著提升了超分任务的性能。对于视频超分,时间帧的对齐至关重要,并得到了广泛研究。早期人们研究显示运动补偿的光流法来估计图像之间的运动并执行翘曲操作。基于循环神经网络设计的 BasicVSR[8](Basic Video Super-Resolution Network)使用光流和递归的结合,研究视频中的递归结构并取得了良好的性能。然而,目标图片中存在物体间的遮挡一些大型的运动变化,很难获得精确的光流。Kim 等人[9]提出了一种时空网络来减少光流估计中的遮挡问题,但效果提升较小;Xue 等人[10]也在基于光流法的视频超分网络(Video Enhancement with Task-Oriented Flow,TOFlow)中表明标准光流并不是视频恢复的最佳运动表示形式。对此,人们开始研究隐式运动补偿的方法。

隐式运动补偿的方法较为常用的是基于可变形卷积实现的。可变形卷积由 Dai 等人[11]首先提出,其原理是为标准卷积核的每一个位置额外学习一个偏移量,可变形卷积核的大小和位置可以根据图像内容动态调整。直观的效果是卷积核会根据图像内容自适应变化,从而适应不同物体的形状、大小等进行几何变形。TDAN 首先在视频超分任务中引入可变形卷积,在特征级别自适应地对齐输入帧,避免显式运动补偿所产生的问题;EDVR 设计具有金字塔和级联结构的可变形卷积对齐模块,它显示出比以前基于光流的超分网络更优越的性能;视频超分增强网络 VESR[12](Video Enhancement and Super-Resolution,VESR)在 EDVR 基础上进行改进,引入全局注意力机制并在残差块中引入通道注意力机制,进一步提高性能。

注意力机制被引入超分任务后对网络性能的提升效果十分显著。基于通道注意力的单图超分网络 RCAN[13](Residual Channel Attention Networks)首次将注意力机制引入到超分任务中,随着单图像超分到视频超分的发展,人们也进行了许多研究。Liu 等人[14]训练学习一组权重图来权衡来自不同时间分支的特征权重;Li 等人[15]提出时空注意力模块来同时利用视频帧的时间和空间信息。

然而,现有大多数网络并没有利用注意力模块中的帧间序列信息,并且在特征重建阶段只是简单地堆叠残差块,不能充分利用不同通道之间特征的相关性。对此,本文旨在利用三维卷积来捕获帧间信息,并采用密集连接结合残差块的方式设计特征重建模块来收集分层特征信息以完成特征重建,最终实现视频帧的超分重建。

2 网络模型架构

2.1 整体方案

本文的目标在于研究一种三维空间时序注意力机制,使超分网络能够充分获取帧间的信息。整体网络结构如图 1 所示,输入图像首先经过残差块堆叠的特征编码模块完成浅层特征提取,经过时间帧对齐模块完成运动补偿,对齐模块采用 EDVR 中基于可变形卷积的金字塔集联帧对齐模块(Pyramid,Cascading and Deformable Convolution,PCD)[6]。将注意力模块部署在时间帧对齐模块之后,以充分利用连续帧之间的信息,其中在时序注意力模块之后,进行特征的全连接融合操作,融合后的特征输入到空间注意力模块进行空间加权。加权后的特征图输入特征重建模块,此模块基于密集连接和残差块设计,

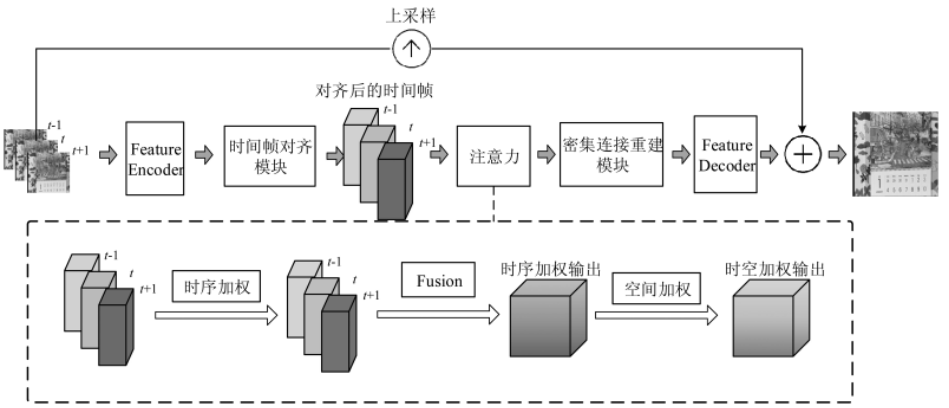


图 1 基于三维时空注意的密集连接视频超分网络

以充分利用分层特征信息完成特征重建。最后将重建后的特征图与原始中间帧上采样图进行元素相加,并通过简单卷积操作完成最终的解码,得到超分辨重建输出。

2.2 时序注意力模块

2.2.1 时序注意力

在时间帧完成对齐之后,可能出现部分帧的错位或者未对齐现象,这会对后续重建性能产生不利影响,因此需要进行时序加权操作。在原 EDVR 框架中,设计了一种基于自注意力机制的时序模块,通过设计嵌入空间计算帧间相似度,相似度越高的,权重则越大。两帧之间的相似度权重可以通过以下公式计算:

$$H(F^a_{t+1},F^a_t)=\mathrm{sigmoid}(\phi(F^a_{t+i})^{\mathrm{T}}\varepsilon(F^a_t)) \quad (1)$$

其中 ϕ 和 ε 是两个嵌入式空间,它们由简单的卷积操作实现,激活函数将输出约束在 0~1 之间,从而约束梯度的反向传播。然而,基于自注意力模块只计算了帧之间的特征相似度,并没有考虑帧间信息。由于视频超分任务对于帧间信息的收集是至关重要的,而现有时间注意力机制中使用的二维卷积无法获取时间维度的序列信息,因此引入三维卷积。

引入三维卷积的时序注意力模块如图 2 所示。本文将三维卷积获取的时序信息添加到原始时间注意力加权帧中,以获得最终的时间注意力加权帧。

2.2.2 空间注意力

空间注意在时间注意模块之后进行,由于超分任务需要一个较大的感受野,模块的设计在考虑轻量化的同时,尽可能地增大了感受野,空间注意力模块如图 3 所示。

首先,融合后的特征输入到空间注意力模块,执

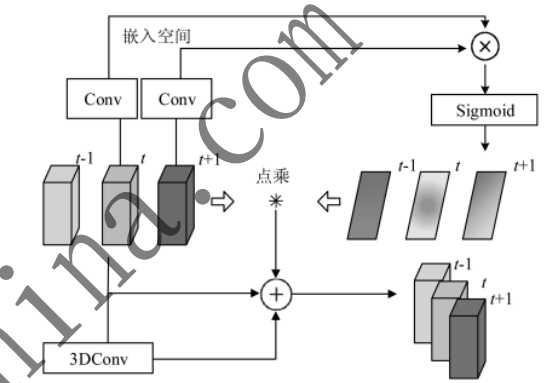


图 2 结合三维卷积的时序注意力模块

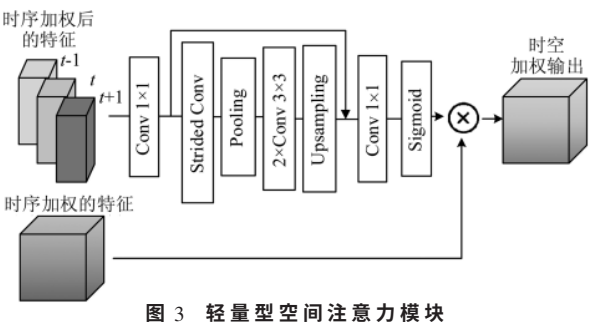


图 3 轻量型空间注意力模块

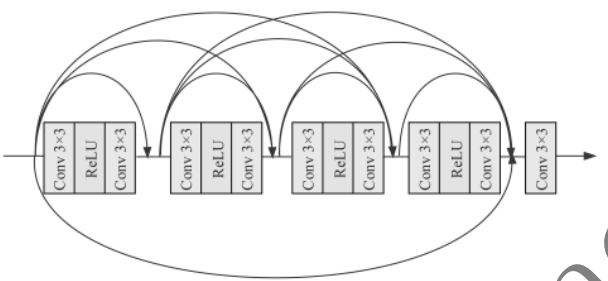
行 1×1 卷积来减少通道数以保持模块轻量化。为了增大感受野,后接一个步幅为 2 的卷积同时进行池化,其中池化由平均池化连接最大池化连接并后接一个一维卷积组成,一维卷积保证通道数的一致。然而由池化带来的感受野的扩大还是有限的,为了继续增大感受野,后接两个 3×3 卷积构成的卷积组。接着,连接一个上采样层来恢复空间维度大小,同时再次利用一个 1×1 卷积来恢复通道数的大小。最后,通过一个 Sigmoid 激活层来获取空间的权重值,并与融合后的特征值相乘得到空间加权后的输出。

最终,经过时序注意力模块和空间注意力模块

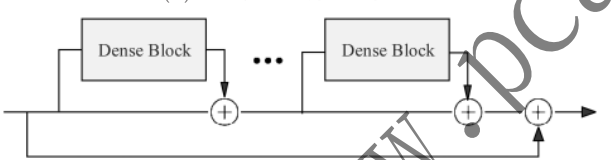
后，输出的时空加权融合特征输入后续的特征重建模块完成特征重建及后续操作。

2.3 密集连接重建模块

为了充分利用分层特征信息，本文采用残差和密集连接结合的方式设计特征重建模块。残差块的设计，使用两个 3×3 卷积层和一个 ReLU 层的组合。残差块侧重于捕捉帧间信息的相关信息，保证了 LR 帧和重建帧的结构相似性。结合残差块，同时引入密集连接来设计密集连接块。受密集残差模块(Residual-in-Residual Dense Block，RRDB)[16]的启发，将四个残差块密集连接，并后接一个 3×3 卷积来调整通道数。密集连接块的内部结构图如图 4(a)所示。最后将多个密集连接块级联，通过全局跳转连接将浅层信息发送到末端，构建成密集连接重建模块，如图 4(b)所示。



(a)密集连接块内部结构图



(b)密集连接块级联结构图

图 4 密集连接重建模块

3 实验验证

3.1 数据集与实验配置

在本次研究中，采用了在 NTIRE19[6](New Trends in Image Restoration and Enhancement 2019)比赛中构建的高质量(720P)视频帧数据集 REDS[17]。REDS 数据集包括 240 个训练片段、30 个验证片段和 30 个测试片段(每个片段包含连续 100 帧)。同时在数据集的高质量图片上做了 4 种不同的退化处理，行成4 个子数据集：双三次下采样子集、带运动模糊的双三次下采样子集、运动模糊子集、带压缩伪影的运动模糊子集。在训练过程中，通常取验证片段中的 4 个片段用作验证，剩下 26 个与训练集一起用于训练使用。取出的 4 个片段取名为 REDS4，作为测试集来使用。另外，Vid4[18]数据集作为常用的验证数据集，包含 4 个测试片段，也作为本实验的测试集使用。

由于可变形卷积在训练过程中随着网络的加深预测的偏移量可能发生跳变，对此本文设计分两个阶段训练网络。第一阶段先训练一个包含 5 个特征提取层和 10 个重建层的小规模网络；第二阶段训练一个包含 10 个特征提取层和 20 个重建层的大规模网络。为减少跳变现象，先训练小规模模型，然后加深网络，使用小规模网络模型作为预训练模型来训练大规模网络。本文所有实验的上采样因子设为 4，采用 5 帧 64×64 大小的连续帧作为输入，输出重建帧的大小为 256×256。残差块的通道数设置为 64，密集块中的增长通道数设置为 32，训练的批次大小 batch size 设置为 8。在训练过程中，通过随机翻转来扩充训练数据集。同时使用 Adam[19]优化器，设置 $\beta1=0.9$ 和 $\beta2=0.99$，学习率初始设置为 4×10^{-4}，损失函数采用 Charbonnier 惩罚函数[20]，如式(2)所示，其中 ε 设置为 10^{-3}。

$$L=\sqrt{\left|\left|\hat{O}_t-O_t\right|\right|^2+\varepsilon^2} \tag{2}$$

3.2 对比实验

本文选用的定量性能指标为峰值信噪比(Peak Signal to Noise Ratio，PSNR)及结构相似性(Structure Similarity Index Measure，SSIM)。对比实验主要依据 PSNR 指数表现以及视觉对比效果进行批判，消融实验依据 PSNR 与 SSIM 两项指标的表现进行批判。进行对比实验时，采用大规模模型训练方式，分别与单图超分网络 RCAN、光流法网络 TOFlow 和可变形卷积网络 EDVR 等几种经典超分网络进行了比较，基于 Vid4 和 REDS4 数据集，实验结果如表 1 和表 2 所示。从实验结果可以看出，本文设计的网络在所有的测试集片段上的表现都要优于其他网络。相比 EDVR，本网络在测试集上的 PSNR 指标有一定提高，平均在 Vid4 上提高了0.08 dB，在 REDS4 上提高了0.18 dB。

视觉对比效果图如图 5 所示。对比不同超分算法对 Vid4 数据集 City 片段中 15 号图片的恢复，可以发现本文网络对于大楼细节纹理的恢复要优于其他网络，且未产生伪影。另外，对于 REDS4 数据集 020 片段中 88 号图片的重建效果，本网络对于货车后备箱纹理线条的恢复要明显优于其他网络，模糊性大大减小，更接近于真实图片。综上所述，本

表 1 不同超分网络在 Vid4 上的 PSNR 指标对比

(dB)

网络结构	Calendar	City	Foliage	Walk	平均值
Bicubic	20.39	25.16	23.47	26.10	23.78
RCAN	22.33	26.10	24.74	28.65	25.46
TOFlow	22.47	26.78	25.27	29.05	25.89
EDVR	23.36	27.53	25.87	30.03	26.69
本文的方法	23.41	27.59	25.94	30.14	26.77

表 2 不同超分网络在 REDS4 上的 PSNR 指标对比

(dB)

网络结构	Clip000	Clip011	Clip015	Clip020	平均值
Bicubic	24.55	26.06	28.52	25.41	26.14
RCAN	26.17	29.34	31.85	27.74	28.78
TOFlow	26.52	27.80	30.67	26.92	27.98
EDVR	27.58	30.95	33.20	29.35	30.27
本文的方法	27.69	31.19	33.43	29.48	30.45

文提出算法不仅在客观评价指标上表现最佳,在视觉对比效果上也要优于其他网络。

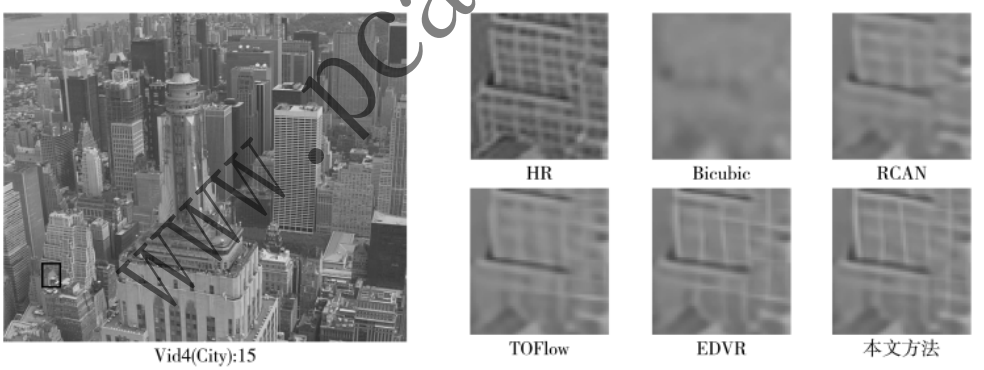
本文对所设计三维空间时序注意力模块和密集连接重建模块进行了消融实验,实验结果如表 3 所示,其中 A 表示三维时空注意力模块,D 表示密集连接重建模块。从表 3 可以看出,空间时序注意力模块的引入在基本不改变参数量的情况下,在 Vid4 和 REDS4 数据集上的性能指标表现均有提升,但提升的性能有限。对此,进一步引入密集连接重建模块,在引入额外参数的情况下,性能表现提升明显。最终,改进网络在 Vid4 数据集和 REDS4 数据集上的 PSNR 指数提升分别为0.08 dB 和 0.18 dB,SSIM 指数提升分别为 0.003 0 和 0.003 6。

4 结论

在视频超分任务中,时序信息是至关重要的,为了充分利用时间序列信息并提高特征重建能力,本文提出了一种三维空间时序注意力机制和用于视频超分的密集连接重建模块。基于三维卷积的注意力模块可以提取帧之间的时序信息用于特征融合;

表 3 在 Vid4 和 REDS4 数据集下的消融实验 PSNR/SSIM 指标对比

网络结构	Vid4	REDS4	参数量
EDVR	26.69/0.804 4	30.27/0.864 3	4.41 M
+A	26.70/0.804 4	30.31/0.864 5	4.47 M
+D	26.72/0.807 4	30.41/0.867 7	6.62 M
+A、D	26.77/0.807 4	30.45/0.867 9	6.68 M



(a) Vid4 数据集 City 片断 15 号图片

(b) REDS4 数据集 020 片段 88 号图片

图 5 不同超分网络重建效果的视觉对比图

密集连接可以充分利用分层特征进行高分辨率图像重建。实验结果表明，在同一实验配置下，本文改进方法对低分辨率图像具有更好的重建恢复效果。

参考文献

[1] SHAHAR O, FAKTOR A, IRANI M. Space-time super-resolution from a single video[C]//IEEE Conference on Computer Vision & Pattern Recognition, 2011.

[2] TAKEDA H, MILANFAR P, PROTTER M, et al. Super-resolution without explicit subpixel motion estimation[J]. IEEE Transactions on Image Processing, 2009, 18(9): 1958-1975.

[3] 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.

[4] TIAN Y, ZHANG Y, FU Y, et al. TDAN: temporally-deformable alignment network for video super-resolution[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 3360-3369.

[5] LIU H, RUAN Z, ZHAO P, et al. Video super-resolution based on deep learning: a comprehensive survey[J]. Artificial Intelligence Review, 2022: 1-55.

[6] WANG X, CHAN K C K, YU K, et al. EDVR: video restoration with enhanced deformable convolutional networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019.

[7] DONG C, LOY C C, HE K, et al. Learning a deep convolutional network for image super-resolution[C]//European Conference on Computer Vision. Springer, Cham, 2014: 184-199.

[8] CHAN K C K, WANG X, YU K, et al. BasicVSR: the search for essential components in video super-resolution and beyond[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 4947-4956.

[9] KIM S Y, LIM J, NA T, et al. 3DSRnet: video super-resolution using 3d convolutional neural networks[J]. arXiv preprint arXiv: 1812.09079, 2018.

[10] 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.

[11] DAI J, QI H, XIONG Y, et al. Deformable convolutional networks[C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 764-773.

[12] CHEN J, TAN X, SHAN C, et al. VESR-Net: the winning solution to youku video enhancement and super-resolution challenge[J]. arXiv preprint arXiv: 2003.02115, 2020.

[13] ZHANG Y, LI K, LI K, et al. Image super-resolution using very deep residual channel attention networks[C]//Proceedings of the European Conference on Computer Vision (ECCV), 2018: 286-301.

[14] 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.

[15] LI S, HE F, DU B, et al. Fast spatio-temporal residual network for video super-resolution[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 10522-10531.

[16] WANG X, YU K, WU S, et al. ESRGAN: enhanced super-resolution generative adversarial networks[C]//Proceedings of the European Conference on Computer Vision (ECCV) Workshops, 2018.

[17] 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/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019.

[18] LIU C, SUN D. On Bayesian adaptive video super resolution[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 36(2): 346-360.

[19] KINGMA D P, BA J. Adam: a method for stochastic optimization[J]. arXiv preprint arXiv: 1412.6980, 2014.

[20] LAI W S, HUANG J B, AHUJA N, et al. Deep laplacian pyramid networks for fast and accurate super-resolution[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 624-632.

（收稿日期：2022-12-13）

作者简介：

何啸林（1998-），男，硕士，主要研究方向：视频超分辨率重建。

吴丽君（1984-），通信作者，女，博士，副教授，硕士生导师，主要研究方向：计算机视觉。E-mail: lijun.wu@fzu.edu.cn。

版权声明

凡《网络安全与数据治理》录用的文章，如作者没有关于汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权等版权的特殊声明，即视作该文章署名作者同意将该文章的汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权授予本刊，本刊有权授权本刊合作数据库、合作媒体等合作伙伴使用。同时，本刊支付的稿酬已包含上述使用的费用，特此声明。

《网络安全与数据治理》编辑部