doi:10. 16055/j. issn. 1672-058X. 2022. 0004. 001

多注意力结合光流的视频超分辨方法

储岳中, 乔雨楠
(安徽工业大学 计算机科学与技术学院,安徽 马鞍山 243000)

摘 要:针对现如今监控摄像、卫星遥感以及视频娱乐等领域对视频图像的清晰度要求越来越高,而目前大部分视频超分辨方法存在参数量大、恢复的视频存在抖动等问题,提出了一种多注意力结合光流的视频超分辨方法,通过引入多个注意力包括空间注意力、通道注意力以及自注意力来提升超分辨性能。具体而言,作为一种特征加权的增强方法,这些注意力可以捕获视频帧的时空特征并增强自适应性和通道间的依赖性,实现全局学习的功能;同时,提出双阶段特征对齐思路,首先利用光流对视频进行估计,进行第一阶段的特征对齐,然后引入长短是记忆网络结构增强位置和通道的特征融合,进行第二阶段的特征对齐,以防止恢复的视频帧出现抖动。实验结果表明:该方法在评估标准和可视化效果方面都取得了令人满意的效果。

关键词:深度学习;注意力机制;特征对齐;光流估计;视频超分辨
中图分类号:TP391 **文献标志码**:A **文章编号**:1672-058X(2022)04-0001-08

0 引 言

视频超分辨率已经成为非常流行和具有挑战性的计算机视觉任务,越来越多的基于深度学习的方法被用来解决超分辨率问题。一类直接的视频超分辨率方法是使用 3D 卷积提取空间信息以保留视频的空间特征。但是,一旦引入了 3D 卷积,则意味着新引入了一个维度,不仅会带来更多的参数,增加计算成本,而且还会限制网络的深度并影响超分辨(Super-Resolution)SR 性能。

此外,更多方法选择逐帧处理视频[1-2],然后根据图像 SR 方法进行超分辨任务。不过,这种方法很难保证视频的连贯性,尤其是对于运动幅度大的视频,局部特征和全局依赖无法很好地集成。可以选择使用递归神经网络用于维持视频的连贯性,但是在保留空间信息方面效果却并不好。

众所周知,典型的深度学习方法总是选择残差连接来传达特征。当特征沿着网络的深度方向前馈时,残差连接可以减少特征的退化,从而可以将特征表达到网络的任何位置。尽管残差连接在特征传递方面很方便,但它并不能完全挖掘不同层之间的特征信息。因此,代替简单的残差跳跃连接,出现了一些复杂的残差变体网络,例如 DRRN[3]、RDN[4](残差密集网络)等。这里,RDN(残差密集网络)是这种变体网络的代表,它不仅使用局部密集残差学习,而且还使用全局残差学习来提取和自适应融合来自所有观察层的局部特征和全局特征。由于 RDN 充分利用了 LR 图像中的多个层次结构特征,因此可以提高图像 SR 的性能。然而,使用残差模块会增

收稿日期:2021-03-30;**修回日期**:2021-05-14.
基金项目:国家自然科学基金资助项目(61971004).
作者简介:储岳中(1971—),男,安徽安庆人,副教授,博士,从事模式识别与数据挖掘研究.
通讯作者:乔雨楠(1993—),女,安徽马鞍山人,硕士,从事模式识别、深度学习研究. Email:qynJoe@ 163. com.

加计算复杂度,并且也会阻碍特征融合和上采样。与此同时,卷积运算可能会给全局学习带来一些缺陷。此外,许多现有方法还会选择使用光流和运动补偿[5-6]来增加帧之间的一致性,这无疑将给整个模型的计算带来负担。随着生成对抗网络(GAN)[7]的出现,用于超分辨率任务的基于GAN的神经网络越来越多。例如,Ledig等[8]提出用于图像超分辨率的对抗网络SRGAN。对于视频超分辨任务,出现了许多基于卷积神经网络(CNN)的视频任务模型。最近,Li等[9]提出了一种快速时空域残差网络(FSTRN),该网络将传统的3D卷积和残差块组合在一起,它不仅可以提取时空域的特征,而且可以减轻计算负担。Wang等[10]提出的EDVR[10]使用可变形卷积将帧从粗到细对齐以便在帧之间进行特征提取;Xiang等[11]提出了一种基于可变形采样的网络,设计了一种新的可变形卷积加ConvLSTM模型来增强时序对齐能力,并利用全局时序上下文信息来处理视频中的大运动。

目前,注意力在许多模型中被广泛使用。例如,在超分辨率下,Zhang等[12]提出将通道注意力与残差相结合以提高网络性能;Wang等[10]在EDVR中提出了时空注意力(TSA),目的在于帮助融合多个对齐的特征信息并且引导图像重建。

注意力机制的优势在于可以快速提取数据的重要特征,注意力机制的改进版本即自我注意力机制可以减少对外部信息的依赖性,并且更擅长捕获远程依赖性以及数据或要素的内部相关性。Wang等[13]提出了一种非局部操作神经网络,该网络可以计算空间中任意位置之间的关系,并且可以作为一个组件插入任何现有结构中。受此启发,Zhang等[14]提出了一种自我注意力生成对抗网络,更好地学习全局特征;Fu等[15]在场景分割任务中引入了双重自注意力,目的是自适应地整合局部特征和全局依赖性;Wang等[16]在立体图像超分辨率的视差注意力中添加了残差块,以处理视差变化较大的不同立体图像,同时提高SR性能。

光流指视频图像当前帧中某一物体或对象像素点所在位置与下一帧中该物体或对象像素点所在位置的位移量。目前常被引用的光流方法包括FlowNet[17]、FlowNet2.0[18]。Alexey等提出FlowNet方法,一方面,将两帧输入图像叠加在一起送到简单光流网中,让网络自动提取运动信息;另一方面,将这两帧输入图像分别送入相同但是独立的处理流网络,方便网络找到对应运动信息。之后利用扩大部分同时保留较粗的高级信息和精细的局部特征,这样提升了光流估计的准确度和速率。在FlowNet2.0方法中,对FlowNet进行了一些改进,增加了训练数据,改进了训练策略;利用堆叠的结构提升效果;引入特定的子网解决空间位移量小的情况。Pan等[19]利用光流估计结合时间清晰度先验进行视频去模糊取得了不错的效果。

若将视频视为一个序列,则循环结构可以起到帧间融合的作用。在现有的许多工作中,都有对循环网络或是变形的应用。Sajjadi等[1]提出一种帧迭代方法,具体做法是评估当前帧LR和前一帧LR之间的光流,然后使用双线性插值方法获得HR光流图,之后进行仿射变换和深度空间操作获得SR;Haris等[2]提出一种使用编码器-解码器方法(Encoder-Decoder)的RBPN,通过反投影合并在单个图像超分辨(SISR)和多个图像超分辨(MISR)中提取细节,扩大RNN中的时间间隔,这样网络对具有更大时间跨度的帧也可以更好地利用。

使用3D卷积的方法一般不使用特征对齐,这样做除了会引入更多的参数量,恢复的视频也难以保持连贯一致性。使用图像超分辨方法,恢复的视频容易产生抖动,清晰度也不高。针对这些问题提出了一种含有多个注意力结合光流的视频超分辨率网络(Multi-attention combined with optical flow video super- resolution network,MAFnet)。一方面,对于视频超分辨中空间信息容易丢失的问题,引入了通道注意力、空间注意力以及自注意力来保留空间信息实现全局学习;另一方面,对于恢复的视频容易出现抖动,无法保持时序上连续性的问题,提出双阶段特征对齐思路,分别对微小运动对象和幅度较大的运动对象进行特征对齐。具体而言,在给定视频序列的情况下,使用残差密集块进行特征提取,然后利用通道和空间注意力将权重分配给不同通道的每个空间位置,有效地使用通道和空间信息,并利用一个自注意力结构捕获空间中长距离依赖关系实现全局学

习。同时,将给定的视频序列分别经过通道注意力和空间注意力,得到的特征输出一起送入一个注意力光流估计分支,进行第一阶段的特征对齐;之后,得到特征先进行上采样然后再送入可变形卷积 LSTM[11] 中进行第二阶段的特征对齐;最后,进行重建得到恢复的视频帧。

本文提出了一个可以应用于视频超分辨率任务的新框架,该模型简单明了,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,第一阶段处理微小运动信息,第二阶段处理幅度较大的运动信息。实验证明所提出方法的可行性,并与现有方法的比较证明了所提出方法在视频 SR 中的有效性。

1　模型介绍

1.1　模型结构

本节将描述所提出的多注意力光流网络(MAFnet)的结构细节。如图 1 所示,网络的主要模块包括特征提取、多注意力分支、注意力光流估计分支、可变形卷积 LSTM、重建部分。将多注意力光流网络的输入大小表示为 $M_L \times N_L$,输入的 LR 帧是 $(2n+1)$ 个 LR 帧序列 $\{I_{t-n}^{\mathrm{LR}},\cdots,I_{t+n-1}^{\mathrm{LR}},I_{t+n}^{\mathrm{LR}}\}$。其中 I_{t+n}^{LR} 是输出的 HR 帧,将其表示为 I_{SR},大小为 $M_H \times N_H$,并且 $M_H > M_L, N_H > N_L$。首先,将输入的视频序列送入第一部分进行特征提取,可以表示为

$$F = H_{\mathrm{rdb}}(H_{c1}(H_{c0}(I_{\mathrm{LR}})))$$

其中,$H_{\mathrm{rdb}}(\cdot)$ 表示残差密集块操作,$H_c(\cdot)$ 表示卷积操作。

随后,将提取的特征分别送入多注意力分支和注意力光流估计分支,可以得到两个分支的输出:

$$\mathcal{F}_a = H_{se}(H_{sa}(H_{ca}(\mathcal{F})))$$

$$\mathcal{F}_f = H_f(H_{sa}(\mathcal{F}), H_{ca}(\mathcal{F}))$$

其中,$H_{se}(\cdot)$,$H_{sa}(\cdot)$ 和 $H_{ca}(\cdot)$ 以及 $H_f(\cdot)$ 分别表示自注意力模块、空间注意力和通道注意力以及光流模块的函数。

之后,将 $\mathcal{F}_a$ 和 $\mathcal{F}_f$ 同时进行上采样,可以表示为

$$y_1 = H_{c4}(\uparrow(H_{c2}(\mathcal{F}_a)))$$

$$y_2 = H_{c5}(\uparrow(H_{c3}(\mathcal{F}_f)))$$

其中,$\uparrow$ 表示上采样。最后,将 y_1, y_2 送入 DLSTM,再经过一层卷积得到最后的输出:

$$\mathcal{F}_d = H_{c6}(\mathrm{DLSTM}(y_1, y_2))$$

其中,$\mathcal{F}_d$ 表示经过 DLSTM 和最后重建卷积得到的特征。整个网络最后可以表示为

$$I_{\mathrm{SR}} = H_{\mathrm{MAFnet}}(I_{\mathrm{LR}})$$

接下来,选择 L_1 损失函数进行优化。给定一个训练集 $\{I_{\mathrm{LR}}^i, I_{\mathrm{HR}}^i\}_{i=1}^N$,其中包含 N 个 LR 输入帧及其对应的 HR。训练 MAFnet 的目标是使损失函数最小化:

$$\mathcal{L}(\theta) = \frac{1}{N}\sum_{i=1}^{N} \| H_{\mathrm{MAFnet}}(I_{\mathrm{LR}}^i) - I_{\mathrm{HR}}^i \|$$

其中,θ 表示网络的参数集。

同时,本文还选择了 Charbonnier Loss[21] 来帮助模型更好地恢复边缘信息,提升性能。计算公式为

$$\mathcal{F}(\theta) = \frac{1}{N}\sum_{i=1}^{N} \rho(I_{\mathrm{HR}}^i - I_{\mathrm{LR}}^i)$$

其中,$\rho(x) = \sqrt{x^2+\varepsilon^2}$ 是 Charbonnier 惩罚函数,ε 在这里设置为 1E-6。训练的更多细节将在实验部分详细表述。

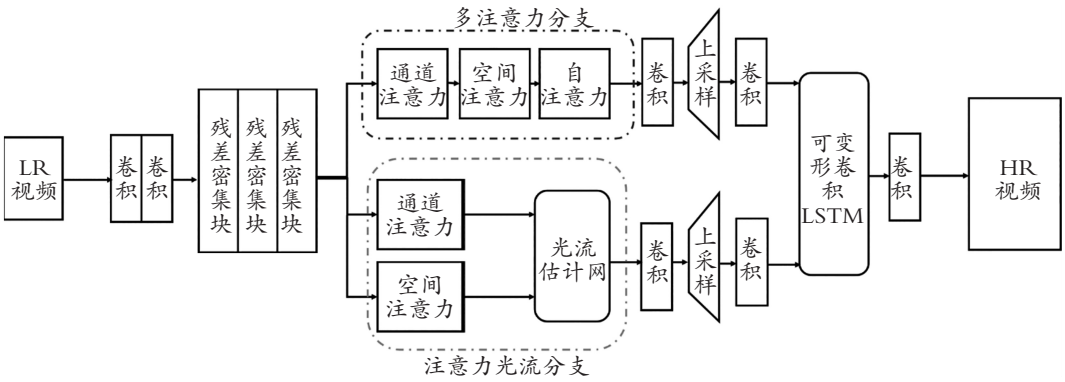


图 1　多注意力光流网络的结构

Fig. 1　The structure of multi-attention optical flow network

1.2 多注意力分支

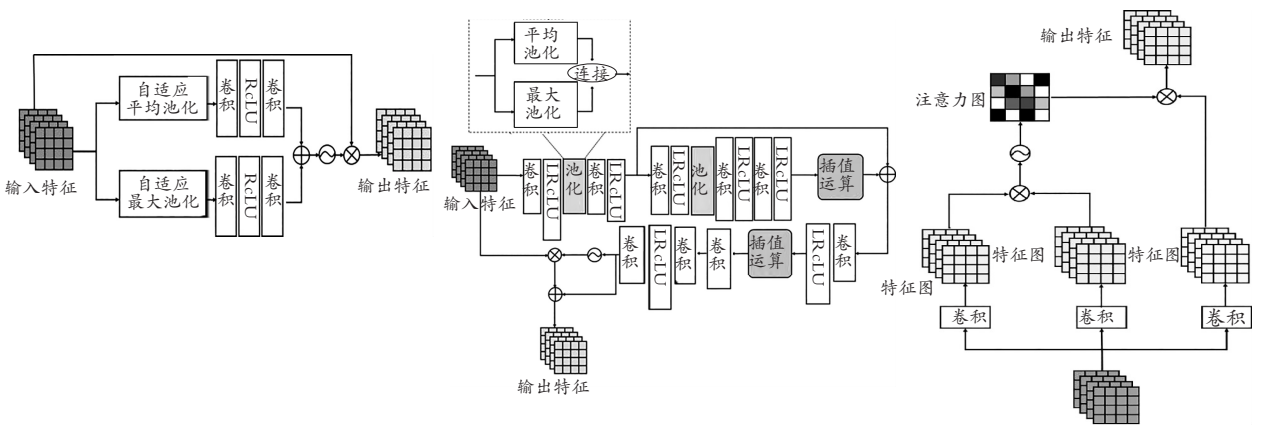
受非局部操作网络[13]的启发,本文提出了多注意力分支(MAB)。针对视频超分辨中空间信息不易保留的问题,在增强通道依赖性的情况下,保留空间信息,并且自适应地实现全局学习功能。

该分支的工作原理:首先,将特征$\mathcal{F}$送入到通道注意力中得到$\mathcal{F}_{ca}$,通道注意力自适应调整通道特征;随后,再将$\mathcal{F}_{ca}$送入空间注意力得到$\mathcal{F}_{sa}$,空间注意力采取了金字塔结构,目的是为通道的每个位置分配权重,有效利用和保留空间信息;最后,利用自注意力(self-attention)结构可以计算任意两个位置之间的依赖关系,从而实现全局学习。

通道注意力的结构如图 2(a)所示。将经过第一个特征提取部分得到的特征作为输入特征,此时特征尺寸大小为 $H\times W\times C$,分别经过自适应平均池化和自适应最大池化后,特征尺寸为 $1\times1\times C$,再分别经过一个卷积核大小为 3 的卷积,将特征尺寸变为 $1\times1\times\frac{C}{r}$,$r$ 是通道缩小比,在这里设置为 16,之后经过激活函数 ReLU;随后,都经过一个 3×3 大小的卷积恢复通道,尺寸为 $1\times1\times C$;将得到的特征相加经过一个 Sigmoid 函数得到注意力特征图,再将注意力特征图与输入特征作矩阵乘法得到输出特征,尺寸为 $H\times W\times C$。

空间注意力结构如图 2(b)所示。在多注意力分支中,空间注意力将通道注意力的输出特征作为输入特征先经过 1×1 大小的卷积和激活函数 LReLU,在空间注意力中之所以选择 LReLU 而非 ReLU,是考虑 ReLU 在训练过程中可能会导致神经元死亡,无法进一步更新参数梯度,使用 LReLU 能够缓和该问题,更好地保留空间信息。经过池化层,池化层是由平均池化和最大池化以及连接操作构成,经过池化层后接着经过 1×1 的卷积和 LReLU 得到的特征记为特征 1;之后,经过重复的 1×1 卷积、LReLU、池化层结构,接着经过 3×3 的卷积和 LReLU 并重复一次该结构,进行插值运算得到的特征记为特征 2;将特征 1 和特征 2 相加后经过 1×1 卷积、LReLU,并再进行一次插值运算,将特征依次送入 3×3 卷积、1×1 卷积、LReLU,得到特征记为特征 3;利用 Sigmoid 函数得到注意力特征图,将注意力特征图和输入特征作矩阵乘法,将结果与特征 3 相加得到输出特征。

自注意力结构如图 2(c)所示。在多注意力分支中,自注意力结构的输入特征是空间注意力的输出,该结构中卷积核大小都为 1×1,得到的特征图作矩阵乘法并经过 softmax 函数得到注意力图与另一个特征图再作矩阵乘法得到输出。



(a) 通道注意力 (b) 空间注意力 (c) 自注意力

图 2 不同注意力的结构图

Fig. 2 Structure charts of different attentions

1.3　注意力光流分支

传统运动补偿方法存在计算复杂、准确度不高的问题。本文采用将注意力与光流相结合的方式处理小运动对象的信息,同时保留对象相关信息,达到第一阶段的特征对齐。将第一部分特征提取得到的特征分别经过通道注意力和空间注意力,将两者的输出送入光流估计网络得到该分支的输出。

给定任何两个相邻帧 I_i, I_{i+1},则光流计算公式可以表示为

$$f_{i\to i+1} = N_f(I_i, I_{i+1})$$

其中,N_f 表示光流估计网络。

1.4　可变形卷积 LSTM 的应用

将可变形卷积[22]加入到传统 LSTM 中。可变形卷积相较于传统卷积可以对空间位置信息的位移进行调整,而相较于空洞卷积[23],不易引入网格伪影。它不仅保留了 LSTM 原本的优点,而且增强了视频帧在时序上对齐的能力,有效地利用上下文信息处理视频中的大运动信息,保证了视频的连续性。

2　实　验

2.1　数据集

Vimeo-90k[24]是一个被广泛应用的数据集,选择 settuplet 子集作为本实验的训练和测试数据集,选择 PSNR 和 SSIM 作为评估标准。在训练过程中,将每个视频剪辑 5 个连续帧输入模型,学习率设置为 1E-4;同时,像大多数图像超分辨和视频超分辨方法一样,将超分辨上采样系数设置为 4;批处理大小是根据 GPU 内存设置的,通常将其设置为 64;然后使用 PyTorch 框架在一张 RTX 2080Ti 显卡上进行实验(表 1)。

表 1　在 Vimeo-90k 测试集上的 PSNR 和 SSIM 的比较

Table 1　Comparison of PSNR and SSIM on the Vimeo-90k test set

方法/对象	字母		花朵		相机		平均值	
	PSNR	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**
VSRnet[25]	31.771	0.793 8	32.363	0.824 8	32.412	0.800 4	32.192	0.755 1
RCAN[12]	32.363	0.811 0	32.882	0.844 7	33.140	0.828 7	33.694	0.828 1
TOFlow[26]	32.465	0.804 1	32.762	0.837 4	32.842	0.814 4	33.081	0.818 6
DUF[27]	32.622	0.808 8	32.931	0.846 6	33.064	0.825 6	33.872	0.827 0
MAFnet	33.494	0.851 6	33.802	0.879 4	34.575	0.896 7	33.908	0.831 0

2.2　对比实验

本文在定量和定性两个方面将提出的方法与不同的超分辨方法进行了比较,包括经典和最新的图像和视频超分辨率方法。所有定量结果都可以在表 1 中找到,选择 PSNR 和 SSIM 作为评估指标。与现有的超分辨方法相比,所提出的多注意力结合光流方法有一定提升。此外,定性结果可以在图 3 中看到,它们显示了 Ground Truth(GT)和超分辨放大倍率 4 倍的结果,可以通过细节图观察到所提出方法可以轻松恢复纹理细节。

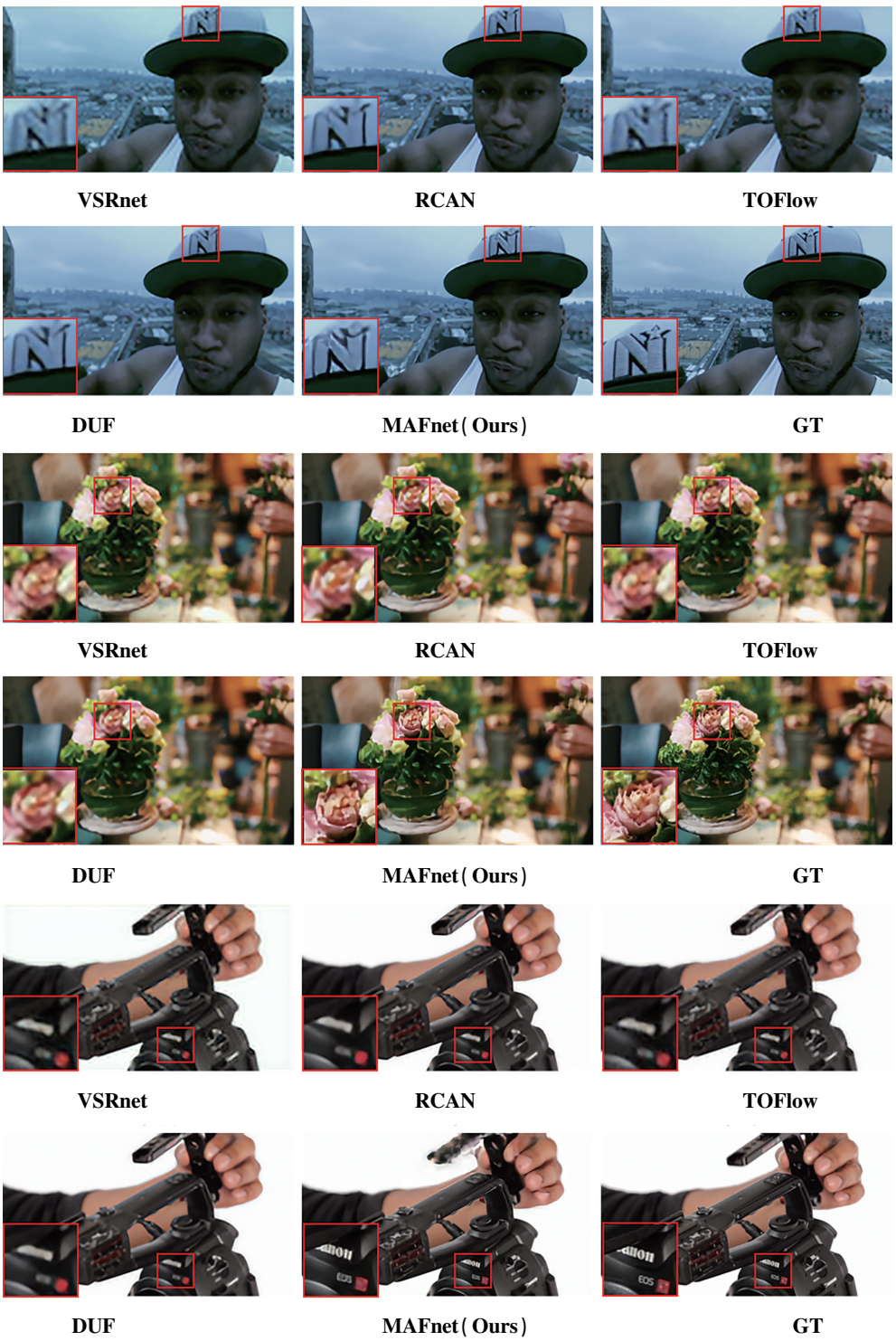


图 3 Vimeo-90k 测试集上的可视化结果

Fig. 3 Visualization results on the Vimeo-90k test set

3 结 语

本文提出利用多个注意力结合光流的网络结构完成视频超分辨任务,并且利用使用了可变形卷积的 LSTM 网络,配合光流估计网络实现双阶段特征对齐的思路,实验结果证明了网络的可靠性和可行性。虽然所提出的模型在可视化效果上取得了令人满意的效果,但是模型不够轻量化,如何设计轻量模型,降低计算复杂度同时保证超分辨性能是接下来要研究和解决的方向。

参考文献(References):

[1] SAJJADI M S, VEMULAPALLI R, BROWN M. Frame-recurrent video super-resolution[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018.

[2] 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. Piscataway: IEEE, 2019.

[3] TAI Y, YANG J, LIU X. Image super-resolution via deep recursive residual network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017.

[4] ZHANG Y, TIAN Y, KONG Y, et al. Residual dense network for image super-resolution[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018.

[5] WANG T C, LIU M Y, ZHU J Y, et al. Video-to-video synthesis[C]//Proceedings of the IEEE International Conference on Neural Information Processing Systems (NeurIPS). Piscataway:IEEE,2018.

[6] 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. Piscataway: IEEE, 2017.

[7] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Advances in Neural Information Processing Systems. Piscataway: IEEE, 2014.

[8] LEDIG C, THEIS L, HUSZÁR 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. Piscataway: IEEE, 2017.

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

[10] WANG X, CHAN K C, 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. Piscataway: IEEE, 2019.

[11] XIANG X, TIAN Y, ZHANG Y, et al. Zooming slow-mo: fast and accurate one-stage space-time video super-resolution[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020.

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

[13] WANG X, GIRSHICK R, GUPTA A, et al. Non-local neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018.

[14] ZHANG H, GOODFELLOW I, METAXAS D, et al. Self-attention generative adversarial networks[C]//Proceeding of the 36th International Conference on Machine Learning. Piscataway: IEEE, 2018.

[15] FU J, LIU J, TIAN H, et al. Dual attention network for scene segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019.

[16] WANG L, WANG Y, LIANG Z, et al. Learning parallax attention for stereo image super-resolution[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019.

[17] FISCHER P, DOSOVITSKIY A, ILG E,et al. Flow net: learning optical flow with convolutional networks[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE, 2016.

[18] ILG E, MAYER N, SAIKIA T, et al. Flow net 2.0: evolution of optical flow estimation with deep networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017.

[19] PAN J, BAI H, TANG J. Cascaded deep video deblurring using temporal sharpness prior[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020.

[20] FENG Y, MA L, LIU W, et al. Spatio-temporal video re-localization by warp LSTM[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019.

[21] LAI W S, HUANG J B, AHUJA N, et al. Fast and accurate image super-resolution with deep laplacian pyramid networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2018, 41(11):2599—2613.

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

[23] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[C]//International Conference on Learning Representations (ICLR), 2016.

[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. Piscataway: IEEE, 2017.
[25] 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.
[26] 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.
[27] JO Y, WUG OH S, 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. Piscataway: IEEE, 2018.
[28] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016.
[29] 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.
[30] ZHANG K, MU G, YUAN Y, et al. Video super-resolution with 3D adaptive normalized convolution[J]. Neurocomputing, 2012, 94(1):140—151.
[31] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8):1735—1780.
[32] TRAN D, BOURDEV L, FERGUS R, et al. Learning spatiotemporal features with 3D convolutional networks [C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE, 2015.

Video Super-resolution Method of Multi-attention Combined with Optical Flow

CHU Yue-zhong, QIAO Yu-nan

(School of Computer Science and Technology, Anhui University of Technology, Anhui Ma'anshan 243000, China)

Abstract: In view of the increasingly higher requirements for the definition of video images in the fields of surveillance cameras, satellite remote sensing, and video entertainment, most of the current video super-resolution methods have problems such as large parameters and jitter in the restored video. A video super-resolution method combining multiple attention and optical flow is proposed. By introducing multiple attentions including spatial attention, channel attention and self-attention, the super-resolution performance is improved. Specifically, as a feature-weighted enhancement method, these attentions can capture the spatio-temporal features of video frames and enhance the adaptability and inter-channel dependence to achieve the function of global learning. At the same time, the idea of two-stage feature alignment is proposed. First, the optical flow is used to estimate the video, the first stage of feature alignment is performed, the introduction of length is a feature fusion of memory network structures that enhance locations and channels, and the second stage of feature alignment is performed to prevent the recovered video frame from shaking. The experimental results show that the method has satisfactory results in both the evaluation standard and the visualization effect.

Key words: deep learning; attention mechanism; feature alignment; optical flow estimation; video super-resolution

责任编辑:李翠薇

引用本文/Cite this paper:

储岳中, 乔雨楠. 多注意力结合光流的视频超分辨[J]. 重庆工商大学学报(自然科学版),2022,39(4):1—8.

CHU Yue-zhong, QIAO Yu-nan. Video super-resolution method of multi-attention combined with optical flow[J]. Journal of Chongqing Technology and Business University (Natural Science Edition),2022,39(4):1—8.