

引用格式:张培健,滕奇志,何小海. 基于级联卷积神经网络的轻量级视频插帧算法 [J]. 微电子学与计算机, 2021, 38(3): 39-45. [ZHANG P J, TENG Q Z, HE X H. Lightweight cascade network for video frame interpolation [J]. Microelectronics & Computer, 2021, 38(3): 39-45.]

基于级联卷积神经网络的轻量级视频插帧算法

张培健,滕奇志,何小海

(四川大学 电子信息学院,四川 成都 610065)

摘 要: 针对基于卷积神经网络的视频插帧算法模型参数过大、实时性差、内存占用高、难以广泛应用的问题,提出了一种基于双向光流和多尺度特征融合的轻量级级联推理网络模型. 将视频插帧任务分解为帧间运动合成和纹理重建两个步骤,设计了一种轻量级的级联双向光流预测网络,并提出了一种多尺度空间与纹理特征融合网络模型,实现了对视频帧多尺度纹理特征和复杂运动特征的充分提取和利用. 该模型使用相邻的两个视频帧和所需中间帧的位置作为网络输入,首先计算两个输入帧的空间金字塔特征与纹理金字塔特征;然后使用空间金字塔特征计算帧间的多尺度双向光流;随后结合双向光流,计算出中间帧的空间特征和纹理特征;最后融合中间帧的多尺度空间、纹理特征得到最终所需的视频帧. 在 Vimeo90K 和 UCF101 数据集上的实验表明,在保证精度的前提下,本文算法在计算速度和模型参数数量上具有更好的表现.

关键词: 视频插帧;深度学习;级联网络;轻量级网络;光流;多尺度特征融合

中图分类号: TP391. 41

文献标识码: A

文章编号: 1000-7180(2021)03-0039-07

Lightweight cascade network for video frame interpolation

ZHANG Pei-jian, TENG Qi-zhi, HE Xiao-hai

(College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China)

Abstract: Aiming at the problems that the model parameters of video interpolation algorithm based on convolutional neural network are too large, poor real-time, high memory occupation and difficult to be widely used, a lightweight cascading inference based on bidirectional optical flow and multi-scale feature fusion is proposed. The network model decomposes the task of video frame insertion into two steps of inter-frame motion synthesis and texture reconstruction, designs a lightweight two-way optical flow prediction network, and proposes a multi-scale spatial and texture feature fusion network model. The multi-scale texture features and complex motion features of video frames are fully extracted and utilized. The model uses the position of two adjacent video frames and the required intermediate frames as network inputs. First, the spatial pyramid features and texture pyramid features of the two input frames are calculated. Then the spatial pyramid features are used to calculate the multi-scale bidirectional optical flow between frames, meanwhile, calculating the spatial and texture features of the intermediate frame. Finally, a fusing network is introduced to generate the multi-scale space and texture features of the intermediate frame to generate the final required video frame. Experiments on the Vimeo90K and UCF101 datasets show that, under the premise of guaranteeing accuracy, the algorithm in this paper has better performance in terms of calculation speed and model parameters.

Key words: video frame interpolation; deep learning; cascade network; lightweight network; optical flow; multi-scale feature fusion

1 引言

视频帧内插技术广泛应用于慢动作生成、视频后处理和动画制作. 通过提高帧率来减少视频的闪烁、运动模糊和干扰. 但插帧过程仅依赖两个输入帧来合成中间帧,使得生成高质量的中间帧变得十分困难.

早期的视频插帧技术主要使用的运动估计和运动补偿^[1-2],但此过程计算量大、生成帧质量低. 近年来,基于深度学习的方法在图像和视频处理取得了突出的成绩. Meyer 等提出一种基于图像相位的网络来预测中间帧的相位值^[3]. 文献^[4-5]将帧内插过程视为相邻帧局部区域的选择过程,为感知运动和合成中间帧设计了自适应卷积核,这种基于自适应核的方法具有较好的生成质量,但这种方法的运动感受范围受到卷积核尺寸的限制,实际使用时需要极大尺寸的卷积核,导致内存占用高、参数量大. 另外,一些基于光流的方法在大尺度运动中表现出优势. Liu 等提出根据时空域上的体素流重建中间帧^[6]. Jiang 等^[7]采用类似的方案,并使用两个 U-Net^[8]结构进行光流估计和遮挡处理. Niklaus 等^[9]提出增加上下文特征来降低生成结果中的伪影. 目前,基于深度学习视频插帧算法大部分可分为两个部分:帧间光流估计、合成中间帧. 一方面,精确且高效地预测帧间光流十分困难,错误的光流估计会导致重建结果不可避免地出现伪影、模糊和失真,这在帧间存在大范围运动物体的情况下尤为严重,需要在后续合成过程加入大量网络层修复缺陷. 另一方面,网络过深会增大训练难度,且在许多实际应用中,插帧任务需要在计算受限的平台上执行,如移动平台和嵌入式设备,神经网络模型要求更低的参数量和计算复杂度.

为了解决上述问题,本文提出了一种级联推理网络,以实现高质量、轻量级和高速的视频插帧. 网络的关键思想是将视频插帧任务分解成时空域重建和纹理重建,并重利用空间特征以减少网络的参数量. 在 UCF101^[10]和 Vimeo90K^[11]数据集上的实验结果表明,本文算法具有更快的运算速度和更低的参数量,并兼顾了较好的生成质量. 与最新的方法相比,本文算法在仅损失 0.16 dB 精度的情况下减少了 60%的参数量.

2 网络结构

网络包括 4 个部分:特征提取、双向光流预测、

特征合成和目标帧重建. 算法的主要流程为:首先,使用两个编码器网络分别提取输入帧的空间金字塔特征和纹理金字塔特征,将多级空间金字塔特征输入至级联光流网络计算多级双向光流. 随后,使用文献^[12]提出的形变算法,结合光流信息,将两输入帧的空间特征与纹理特征合成为中间帧的特征. 最后,将形变后的中间特征重建为所需的视频帧.

网络整体结构如图 1 所示,输入为相邻两帧 I_0 、 I_1 和中间帧的位置 t .

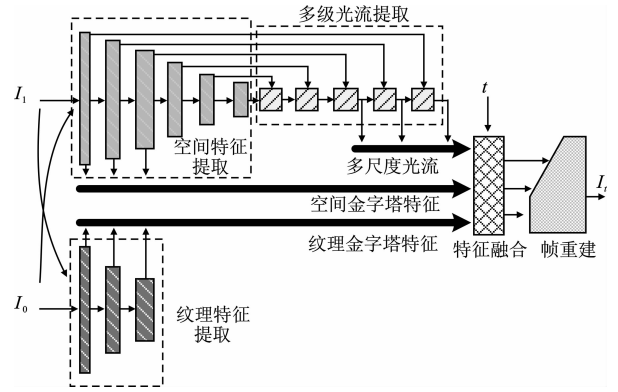


图 1 网络整体结构图

2.1 特征提取网络

本文使用两个不同的孪生网络提取两个输入视频帧的金字塔空间和纹理特征,分别使用 S_l^i 和 C_l^i 表示输入帧 i 的第 l 级空间特征和纹理特征. 本文采用 6 级空间金字塔(l 从 1 到 6),3 级纹理金字塔(l 从 1 到 3). 两个金字塔均采用步长为 2 的卷积核 2 倍降采样,每个卷积层后使用 Leaky Relu 作为激活函数($\alpha = 0.1$). 另外,在纹理金字塔中使用残差网络^[13]的 ResBlock 层. 具体参数设置如表 1 所示,格式为 {卷积层 - 尺寸 - 输出通道 - 步长}.

表 1 特征提取网络卷积层参数设置

尺度等级	空间特征	纹理特征
level-1	Conv-7-32-1	Conv-7-32-1
	Conv-3-32-2	
level-2	Conv-3-32-1	Conv-3-32-2
	Conv-3-32-1	ResBlock-32-1
level-3	Conv-3-32-1	Conv-3-64-2
	Conv-3-64-2	ResBlock-64-1
	Conv-3-64-1	ResBlock-64-1
level-4	Conv-3-96-2	
	Conv-3-96-1	
level-5	Conv-3-128-2	
	Conv-3-128-1	
level-6	Conv-3-192-2	
	Conv-3-192-1	

2.2 级联光流网络

在视频插帧算法中,精确的光流预测是十分重要的. 本文网络采用两种级联推断:(1)使用多尺度金字塔结构,特征随着尺度的增加扩大了运动感知范围,降低了噪声对运动检测的不利影响,并使用高尺度光流监督低尺度的光流预测,降低了网络预测

难度;(2)采用级联网络预测每级尺度的光流,通过对多级特征形变(warping layer)并匹配(cost volume)以实现遮挡的处理,并使用残差结构降低预测的难度. 第1级尺度的光流预测网络结构如图2所示,包括两个级联部分:光流校正网络(图2(a))和光流预测网络(图2(b)).

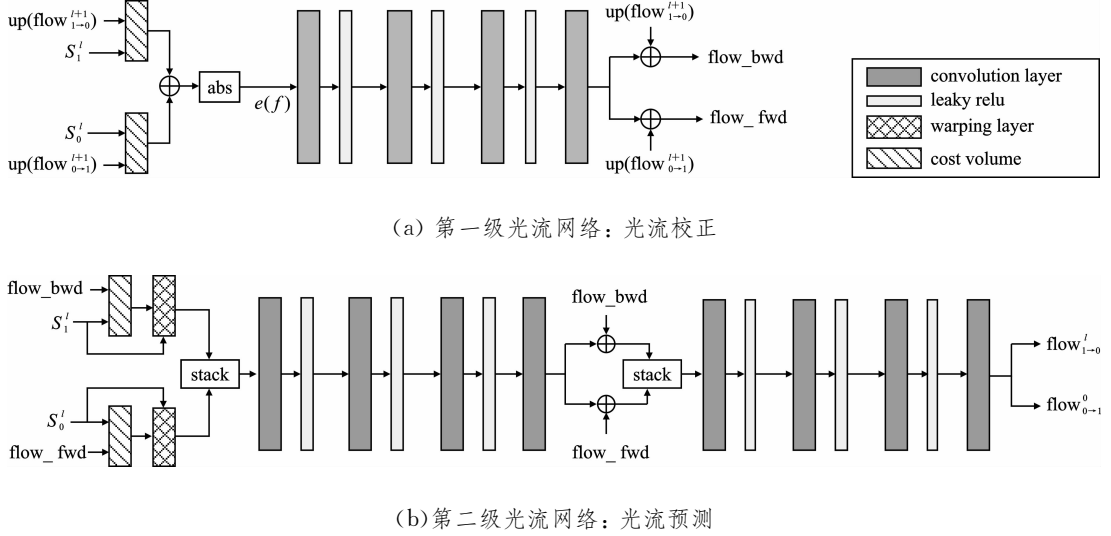


图2 第 l 层级联双向光流的结构图

光流预测的一个关键要素是匹配特征的相似度并衡量两个特征中相似物体的相对距离,即图2中的cost volume层. 这里特征匹配相似度使用文献[14-15]提出的cost volume层计算形变后特征的相似度,计算方法如公式(1)所示:

$$c'(F_0^l, F_1^l) = \frac{1}{N} (F_0^l(x))^T F_1^l(x+d) \quad (1)$$

式中,运算符 T 表示矩阵的转置, F_0^l 和 F_1^l 为两个输入特征, N 为特征的长度, d 为匹配距离,将其中一个特征平移 d 个像素后与另一个特征的转置相乘得到两个特征的相似性特征. 由于本文使用多尺度特征预测光流,在高尺度中使用短匹配距离也可捕获到帧间物体的大范围运动,这里将平移距离 d 为固定值,范围为 $[1, 8]$. Cost volume层输出特征的尺寸为 $d^2 \times H \times W$,其中 H, W 是输入特征的高和宽.

图2中的形变层(warping layer)用于得到在光流引导下平移后的特征,本文使用文献[12]提出的双线性像素形变方法,用光流表示每个像素点 x, y 方向上的移动距离,将像素平移后用双线性插值计算形变后每个点的像素值,形变过程如图3所示.

第一级光流校正网络的输入为两帧第 l 级的空间特征 S_0^l, S_1^l 和上一级预测的双向光流值 $\text{flow}_{0 \rightarrow 1}^{l+1}, \text{flow}_{1 \rightarrow 0}^{l+1}$,并将双向光流上采样2倍. 该过程假设上一级预测的光流存在错误,而错误的光流会增加后续

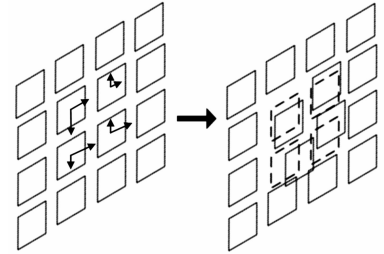


图3 特征形变层示意图

光流预测的难度;并假设在理想情况下使用精确光流将两个输入帧的特征形变至中间时刻应该完全相同,因此可使用两个形变特征的误差来表征光流预测的错误,计算方法如式(2)所示,其中 $\odot$ 表示上文所述的形变层.

$$e(f) = |S_0^l \odot \text{up}(\text{flow}_{0 \rightarrow 1}^{l+1}) - S_1^l \odot \text{up}(\text{flow}_{1 \rightarrow 0}^{l+1})| \quad (2)$$

将绝对值误差 $e(f)$ 输入至参数为 R_1 的卷积神经网络 $F[\cdot, R_1]$ 得到双向光流的校准值,与上一级双向光流相加得到校准后的光流 flow_fwd 和 flow_bwd ,如公式(3)所示.

$$\begin{bmatrix} \text{flow_bwd} \\ \text{flow_fwd} \end{bmatrix} = \begin{bmatrix} \text{up}(\text{flow}_{1 \rightarrow 0}^{l+1}) \\ \text{up}(\text{flow}_{0 \rightarrow 1}^{l+1}) \end{bmatrix} + F[e(f), R_1] \quad (3)$$

第二级光流预测网络用于生成第 l 级尺度空间的光流值 $\text{flow}_{0 \rightarrow 1}^l$ 和 $\text{flow}_{1 \rightarrow 0}^l$. 网络输入为两帧第 l 级

的空间特征 S_0^l, S_1^l 和校准后的光流 flow_fwd 和 flow_bwd . 首先使用前向光流 flow_fwd 和后向光流 flow_bwd 形变空间特征 S_0^l, S_1^l , 并使用 Cost volume 层得到特征 $S_0^l(S_1^l)$ 与形变后特征的平移相似度 Cor^l . 然后将两个相似度特征输入至参数为 R_2 的卷积神经网络 $F[\cdot, R_2]$ 得到相对上一级尺度校正光流的残差值, 获得高尺度特征忽略的运动细节, 并将残差光流与校正光流相加, 如公式(4)所示. 并再次输入至参数为 R_3 的卷积神经网络 $F[\cdot, R_3]$ 以增强光流的效果.

$$f = \begin{bmatrix} \text{flow_bwd} \\ \text{flow_fwd} \end{bmatrix} + F(\text{Cor}^l, R_2) \quad (4)$$

$$\begin{bmatrix} \text{flow}_{1 \rightarrow 0}^l \\ \text{flow}_{0 \rightarrow 1}^l \end{bmatrix} = F(f, R_3) \quad (5)$$

上述神经网络 F 中四个 CNN 层的卷积核通道数分别设置为 196, 64, 32 和 4; 激活函数使用 Leaky Relu ($\alpha=0.1$); warping layer 层和 cost volume 层均无可训练参数.

2.3 特征合成与帧重建

正文内容. 中间帧的特征可在双向光流的引导下形变两个输入帧的特征得到, 将空间特征 S_t^l 与纹理特征 C_t^l 沿通道合并为一个特征 F_t^l , 则位于 t 时刻中间帧的特征可由公式(6)得:

$$F_t^l = [t \cdot \text{flow}_{0 \rightarrow 1}^l] \odot F_0^l + [(1-t) \cdot \text{flow}_{1 \rightarrow 0}^l] \odot F_1^l \quad (6)$$

式中, 运算符 $\odot$ 表示形变操作.

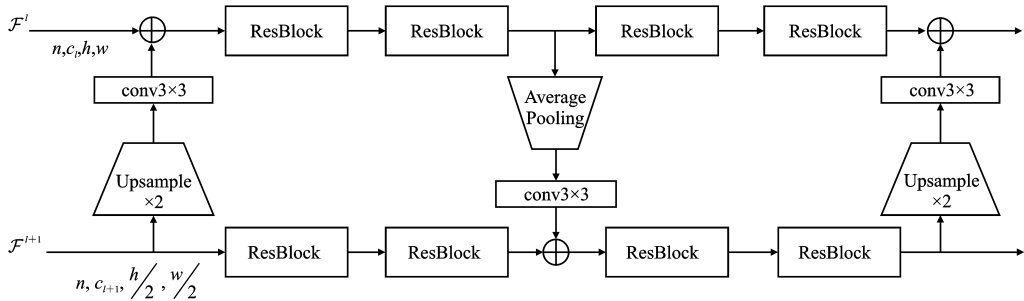


图6 模块 GridBlock 的结构示意图

2.4 网络训练

使用高帧率视频数据集 Vimeo90K^[11] 和 Adobe 240Fps^[16] 训练网络. Vimeo90K 包括 51,312 组训练数据, 每组图像尺寸均为 256×448 . Adobe 240Fps 包括 133 个分辨率为 $1\,280 \times 720$, 帧率为 240Fps 的视频文件, 总共 120k 个视频帧. 网络训练前将所有视频帧以相邻的 13 帧分为一组, 总共 64,711 组. 训练时使用随机剪裁和随机序列反序增广

特征示例如图 4 所示, 可看出用于光流预测的空间特征包含大量的形态信息, 空间特征提取网络更倾向于寻找物体的轮廓, 而纹理特征有丰富的细节信息.



(a) 输入帧 (b) 纹理特征 (c) 空间特征

图4 空间特征与纹理特征可视化

帧重建网络将结合多尺度空间与纹理信息合成所需的中间帧. 网络结构如图 5 所示.

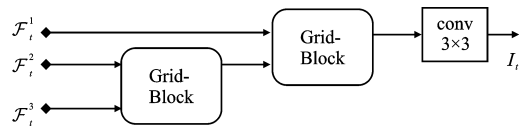


图5 帧重建网络结构图

帧重建网络使用公式(6)计算的 t 时刻特征重建出中间帧 I_t . 网络包括两个 GridBlock(如图 6 所示), 每个 GridBlock 有两个不同尺度的输入和一个低尺度的输出. GridBlock 的结构如图 6 所示, 网络采用高低尺度互补的形式包含两次上采样和一次下采样, 相邻的上下采样之间加入 2 个文献[13]提出的残差块(ResBlock), GridBlock 可实现高低尺度信息的融合, 提高低尺度特征的信息量, 降低重建难度.

数据集, 图像随机剪裁尺寸为 256×256 并保证每组的剪裁位置相同.

对于图像生成任务, 损失函数应衡量预测图像与真实图像之间的差异. 本文首先采用 l_1 损失(平均绝对误差)计算目标值 I_t^{GT} 与预测值 $\hat{I}_t$ 之间的误差:

$$l_o = \|\hat{I}_t - I_t^{GT}\|_1 \quad (7)$$

同时,本文引入了一种感知损失函数,其使用高维的特征相似度来评价预测图像的准确性:

$$l_p = \| \varphi(\hat{I}_t) - \varphi(I_t^{GT}) \|_1 \tag{8}$$

式中, $\varphi(\cdot)$ 是 ImageNet^[17] 预训练的 VGG16^[18] 模型中 conv4_3 输出的特征.

另外,本文对光流结果加入一个平滑度惩罚,促使光流预测的结果更加平滑:

$$l_s = \frac{1}{N} \sum_{l=1}^N (\| \nabla \omega_{0 \rightarrow 1}^l \|_1 + \| \nabla \omega_{1 \rightarrow 0}^l \|_1) \tag{9}$$

式中, $\nabla(\cdot)$ 表示哈密顿算子.

最后,结合公式(7)、(8) 和(9),网络的损失函数为:

$$l = \lambda_1 l_o + \lambda_2 l_p + \lambda_3 l_s \tag{10}$$

λ 权重分别设定为 10,320 和 0.05.

训练采用 AdaMax 优化器训练网络,学习率初始设置为 0.000 1,权重衰减设为 0.000 5, β_1 和 β_2 分别设置为 0.9 和 0.999,batchsize 设为 10.网络首先训练 30 轮,然后将学习率降低为 10^{-5} 并再训练 10 轮.

3 实验结果与分析

本文实验环境为 Ubuntu18.04,CPU 为 Dual Intel Xeon(R) CPU E5-2686 v4,2.30 GHz,GPU 为 NVIDIA GTX 1080Ti (Pascal),系统环境为 CUDA 10 和 cuDNN 7,并使用 Pytorch 框架实现算法.在 Vimeo90K^[11] 和 Adobe 240Fps^[16] 数据集上评估训练完毕的模型并验证各个组件的有效性(见 3.1 节).同时与近些年基于深度学习的算法相比较(见 3.2 节).本文使用 PSNR 和 SSIM 来衡量生成结果的准确率,另外比较算法在图像尺寸 640×480 上的运算时间.

3.1 组件消融实验

本小节将实验本文所提出的组件的有效性,包括有无空间特征(S_p)、纹理特征(C_p)和重建网络(G_b)的层数,实验结果如表 2 所示.

从实验可以看出结合空间与纹理信息能提供更好的质量.相较于仅有重建模块的网络,融合空间特征或纹理特征均增加了 2dB 的 PSNR 精度,结合空间特征和纹理特征能提高 3dB 的 PSNR 精度.随着 Gridblock 数量和金字塔级数的增加,虽然增大了较多的参数数量和计算时间,但生成质量的提升更高.综合运算时间、参数数量和精度的考虑,本文最终选择空间特征+纹理特征+两个 Gridblock(三级尺度金字塔)的网络结构.

表 2 消融实验结果

组件	参数数量 /百万	耗时 /秒	UCF101		Vimeo90K	
			PSNR	SSIM	PSNR	SSIM
S_p	2	9.51	31.56	0.921	32.21	0.934
C_p	2	9.53	33.86	0.938	34.12	0.953
G_b	2	9.62	34.01	0.941	34.15	0.952
✓	2	9.73	34.55	0.955	34.64	0.960
✓	1	7.43	33.92	0.942	34.21	0.954
✓	0	6.22	30.12	0.927	30.25	0.933

3.2 效果对比实验

本小节将本文提出的算法与近些年基于深度学习的算法相比较,包括 Separable adaptive convolution (SepConv)^[5],deep voxel flow (DVF)^[6], Super SloMo^[7],context-aware method (CtxSyn)^[9] 和 depth-aware method (DAIN)^[19].

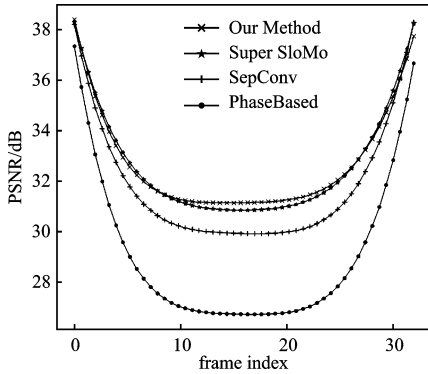
实验统计结果如表 3 所示,网络 DVF 虽然参数量最少但使用了较多人工设计的无参数层,运算时间最长,且在 Vimeo90K 数据集中表现较差.与 CtxSyn 和 Super SloMo 相比,本文算法在 Vimeo90K 数据集和 UCF101 数据集中测试结果更好,速度更快,参数量更低.本文算法在数据集 UCF101 中相比于 SepConv 和 DAIN 的精度只低了 0.44 dB;在 Vimeo90K 相比于 DAIN 的精度低了 0.07 dB,但参数量降低了 60%.由此,说明了本文提出的级联网络方法能够充分利用视频插帧任务的尺度空间特性,在有限的参数量与较低的计算开销下得到较好的性能.

表 3 算法性能比较

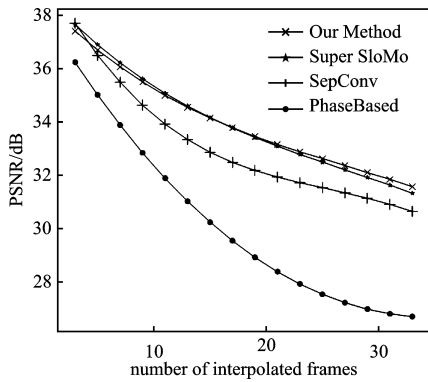
算法	参数数量 /百万	耗时 /秒	UCF101		Vimeo90K	
			PSNR	SSIM	PSNR	SSIM
DVF	1.6	0.47	34.12	0.963	31.54	0.946
SepConv	21.6	0.20	34.69	0.966	33.45	0.967
SuperSloMo	39.6	0.17	33.14	0.938	—	—
CtxSyn	17.5	0.17	34.01	0.941	33.76	0.955
DAIN	24.0	0.12	34.99	0.968	34.71	0.975
本文算法	9.73	0.13	34.55	0.955	34.64	0.960

为更好地验证方法的有效性,本文测试了算法在插入不同数目中间帧时的性能表现.选取 Super SloMo、SepConv 和文献[3]提出的相位方法作为对比,测试插入 30 帧时不同位置的精度(如图 7(a)) 和不同插帧数量的平均精度(如图 7(b)).从图 7(a) 可以看出,随着插帧位置逐渐移动至中间位置,生成精度逐渐降低,本文算法在中间位置附近度优势明

显,尤其是第 10 帧至第 25 帧.从图 7(b)中可以看出,随着帧间插帧数量的增大,平均精度逐渐下降,在插帧数量大于 17 时本文算法效果更好.随着插帧数量的增大,本文提出算法表现出了更好的稳健性.



(a) 帧间生成 30 帧时,每个位置的平均 PSNR



(b) 生成不同数量中间帧时的平均 PSNR

图 7 在 Adobe240-fps 数据集上生成多个帧的测试结果

4 结束语

本文提出一种基于深度学习和级联推理的视频插帧算法,利用深度学习可以自动从数据提取有效信息而无需手动设计的特点,实现网络端到端的训练.算法将视频插帧任务分解为空间信息和纹理信息重建,并提出一种级联光流预测网络,使用空间特征获取两帧间的双向光流,并在双向光流的引导下生成中间帧的空间和纹理特征;最后设计了一个多尺度重建网络,结合多尺度形变后的特征重建出中间帧.实验表明,本文提出的方法在确保生成质量的情况下实现了网络参数的轻量化和计算的高效性.

参考文献:

[1] LEE W H, CHOI K, RA J B. Frame rate up conversion based on variational image fusion[J]. IEEE

Transactions on Image Processing, 2014, 23(1): 399-412. DOI: 10.1109/TIP.2013.2288139.

[2] CHOI B D, HAN J W, KIM C S, et al. Motion-compensated frame interpolation using bilateral motion estimation and adaptive overlapped block motion compensation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2007, 17(4): 407-416. DOI: 10.1109/TCSVT.2007.893835.

[3] MEYER S, DJELOUAH A, MCWILLIAMS B, et al. PhaseNet for video frame interpolation[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 498-507. DOI: 10.1109/CVPR.2018.00059.

[4] NIKLAUS S, MAI L, LIU F. Video frame interpolation via adaptive convolution[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 670-679. DOI: 10.1109/CVPR.2017.244.

[5] NIKLAUS S, MAI L, LIU F. Video frame interpolation via adaptive separable convolution[C]//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 261-270. DOI: 10.1109/ICCV.2017.37.

[6] LIU Z W, YE H R A, TANG X O, et al. Video frame synthesis using deep voxel flow[C]//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 4463-4471. DOI: 10.1109/ICCV.2017.478.

[7] JIANG H Z, SUN D Q, JAMPANI V, et al. Super SloMo: high quality estimation of multiple intermediate frames for video interpolation[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 9000-9008.

[8] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[C]//Proceedings of the 18th International Conference International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany: Springer, 2015: 234-241. DOI: 10.1007/978-3-319-24574-4_28.

[9] NIKLAUS S, LIU F. Context-aware synthesis for video frame interpolation[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 1701-1710. DOI: 10.1109/CVPR.2018.00183.

[10] SOOMRO K, ZAMIR A R, SHAH M. UCF101: a dataset of 101 human actions classes from videos in the

- wild[J]. arXiv preprint arXiv:1212.0402, 2012.
- [11] XUE T F, CHEN B A, WU J J, et al. Video enhancement with task-oriented flow[J]. International Journal of Computer Vision, 2019, 127(8): 1106-1125. DOI: 10.1007/s11263-018-01144-2.
- [12] JADERBERG M, SIMONYAN K, ZISSERMAN A. Spatial transformer networks[C]//Advances in Neural Information Processing Systems. 2015: 2017-2025.
- [13] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770-778. DOI: 10.1109/CVPR.2016.90.
- [14] DOSOVITSKIY A, FISCHER P, ILG E, et al. FlowNet: learning optical flow with convolutional networks [C]//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015: 2758-2766. DOI: 10.1109/ICCV.2015.316.
- [15] ILG E, MAYER N, SAIKIA T, et al. FlowNet 2.0: evolution of optical flow estimation with deep networks [C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 2462-2470. DOI: 10.1109/CVPR.2017.179.
- [16] SU S C, DELBRACIO M, WANG J, et al. Deep video deblurring for hand-held cameras[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 1279-1288. DOI: 10.1109/CVPR.2017.33.
- [17] DENG J, DONG W, SOCHER R, et al. ImageNet: a large-scale hierarchical image database[C]//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL: IEEE, 2009: 248-255. DOI: 10.1109/CVPR.2009.5206848.
- [18] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
- [19] BAO W B, LAI W S, MA C, et al. Depth-aware video frame interpolation[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA: IEEE, 2019: 3703-3712.

作者简介:

张培健 男,(1995-),硕士研究生.研究方向为计算机视觉.
滕奇志(通讯作者) 女,(1961-),教授,博士生导师.研究方向为数字图像处理、模式识别.

E-mail: qzteng@scu.edu.cn.

何小海 男,(1964-),教授,博士生导师.研究方向为图像处理、模式识别和图像通信.

(上接第38页)

- [6] CHAI X L, GAN Z H, YUAN K, et al. A novel image encryption scheme based on DNA sequence operations and chaotic systems[J]. Neural Computing and Applications, 2019, 31(1): 219-237. DOI: 10.1007/s00521-017-2993-9.
- [7] ZHOU Y C, BAO L, CHEN C L P. A new 1D chaotic system for image encryption[J]. Signal Processing, 2014, 97: 172-182. DOI: 10.1016/j.sigpro.2013.10.034.
- [8] BELAZI A, KHAN M, EL-LATIF A A, et al. Efficient cryptosystem approaches: S-boxes and permutation-substitution-based encryption[J]. Nonlinear Dynamics, 2017, 87(1): 337-361. DOI: 10.1007/s11071-016-3046-0.
- [9] WU X J, WANG K S, WANG X Y, et al. Color image DNA encryption using NCA map-based CML and one-time keys[J]. Signal Processing, 2018(148): 272-287. DOI: 10.1016/j.sigpro.2018.02.028.
- [10] KANG X J, GUO Z H. A new color image encryption scheme based on DNA encoding and spatiotemporal chaotic system[J]. Signal Processing: Image Communication, 2020(80): 115670. DOI: 10.1016/j.image.2019.115670.

作者简介:

王 钊 男,(1995-),硕士研究生.研究方向为网络信息安全技术及应用. E-mail: 2374805061@qq.com.

张雪锋 男,(1975-),博士,教授.研究方向为信息安全、信号与信息处理.

姜 文 男,(1983-),博士,高级工程师.研究方向为基础设施与工业互联安全.