

基于视频先验信息的轻量化去噪卷积神经网络

申屠敏健¹, 朱强¹, 朱树元¹, 孟现东²

(1. 电子科技大学信息与通信工程学院, 四川成都 611731; 2. 鹏城实验室, 广东深圳 518000)

摘要: 提出了一种基于视频先验信息的轻量化去噪卷积神经网络. 先验信息从近邻的多帧视频图像中获取, 采用了基于预去噪的视频运动补偿方法消除噪声和运动偏移对信息获取准确度的影响. 为降低卷积神经网络复杂度, 构建了基于双路处理的卷积神经网络用于去除视频噪声, 特别是设计了双路稠密连接单元, 实现了网络的轻量化. 双路稠密连接单元通过高、低分辨率特征分解和特征拼接, 有效降低了网络复杂度. 实验结果表明: 采用本文方法去除视频噪声能够获得较好的客观评价结果和主观视觉结果. 此外, 在减少网络参数、降低浮点运算次数和提升运行速度方面均体现出了良好性能.

关键词: 视频去噪; 卷积神经网络; 先验信息; 轻量化网络; 特征

基金项目: 国家自然科学基金(No.U20A20184)

中图分类号: TP391

文献标识码: A

文章编号: 0372-2112(2023)06-1510-08

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20210679

A Priori Information-Based Lightweight Convolutional Neural Network for Video Denoising

SHENTU Min-jian¹, ZHU Qiang¹, ZHU Shu-yuan¹, MENG Xian-dong²

(1. School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China;

2. Peng Cheng Laboratory, Shenzhen, Guangdong 518000, China)

Abstract: A priori information-based lightweight convolutional neural network (CNN) for video denoising is proposed in this paper. More specifically, the priori information is obtained from the adjacent frames and the pre-denoising-based motion compensation is applied to effectively collect the priori information. Meanwhile, the dual-path processing-based CNN is designed to remove the video noise. Specifically, the dual-path cross connection unit (DCU) is proposed to simplify the feature extraction for the design of a lightweight network. With DCU, the high-resolution features and the low-resolution features are generated and they are concatenated by shorter connections, which achieve low complexity. The experimental results demonstrate that our proposed method offers good objective results as well as subjective results. Moreover, it effectively reduces network parameters and floating point operations, and achieves a fast CNN-based video denoising.

Key words: video denoising; convolutional neural network; priori information; lightweight network; feature

Foundation Item(s): National Natural Science Foundation of China (No.U20A20184)

1 引言

视频噪声的存在不仅会影响人们的视觉观感体验,而且会影响后续视频处理效率. 如何有效消除视频噪声从而提高视频质量,长期以来一直是视频处理领域的重点研究问题之一.

视频由连续图像按时间顺序组成,采用静态图像去噪算法^[1-6]对视频中的每帧图像去噪是实现视频去

噪的最直接方法. 静态图像去噪方法的实现建立在图像内部空间信息基础之上,图像的纹理和结构特征是最常用的空间信息. 与静态图像不同,视频中相邻图像之间存在极大时序关联性,充分利用这一关联性构建去噪算法,能够达到较好的去噪效果,当前主流视频去噪算法,如VBM3D^[7]、VBM4D^[8]和VNLB^[9]等,都建立在视频图像时序关联性的基础之上.

近年来,采用卷积神经网络构建去噪算法已成为图像去噪^[10]和视频去噪^[11]领域的主流方法,然而这些算法的计算代价过高,并且网络结构十分复杂,因此限制了它们在实际中的应用.面向实际应用设计轻量化的去噪卷积神经网络,解决计算代价和复杂度问题是十分必要的.卷积神经网络的轻量化要求在保证算法性能的前提下减少网络参数和浮点运算次数(FLOPs).为达到这一目的,在SqueezeNet中,Iandola等人^[12]通过使用更小尺寸的卷积以及减少卷积数量等策略降低网络参数量;Howard等人^[13]提出了一种针对移动设备的神经网络——MobileNets,该网络建立在深度可分离卷积的基础之上,以更少的网络参数达到了较好的性能;此外,在Xception中,Chollet^[14]在使用分组卷积实现算法时,所采用的卷积组数与输入特征通道数相同,从而大幅降低了卷积运算次数;Zhang等人^[15]在ShuffleNet中使用小尺寸1×1分组卷积、Channel Shuffle操作以及深度可分离卷积,减少了计算量并保障了推理精度;在GhostNet中,Han等人^[16]通过使用低复杂度线性变换减少相似特征,达到了减少浮点计算量的目的.上述轻量化方法多采用分组卷积或深度可分离卷积实现,这两种卷积并未融合所提取的全部特征,因此网络性能会受到影响.此外,这些方法仅在图像分类、目标检测和语义分割^[17]等领域得到了应用,目前仍缺少能够针对图像和视频进行去噪处理的轻量化卷积神经网络.

为实现低复杂度、高效视频去噪,本文基于视频时序先验信息,提出了一种轻量化视频去噪卷积神经网络,该方法充分利用近邻参考视频图像的时空关联信息消除视频噪声,并通过运动补偿消除帧间运动偏差以提升先验信息的利用效率.在去噪算法设计方面,本文提出了一种双路结构卷积神经网络,该网络利用输入视频产生高、低分辨率两路特征,通过分别处理两路不同分辨率的特征降低浮点运算次数,并应用稠密连接技术降低训练难度,同时利用残差学习保障网络达到良好性能.实验结果表明,本文所提出的方法不仅有效去除了视频噪声,而且在浮点运算次数、运行速度以及内存使用等方面都体现出了优势,实现了视频去噪卷积神经网络设计的轻量化.

2 本文方法

2.1 运动补偿

本文以近邻视频图像时空信息融合为基础,充分利用视频先验信息构建高效去噪模型.为消除视频帧间运动偏差对信息融合的影响,提升视频去噪性能,采用近邻图像运动补偿方法^[11]对帧间图像运动偏差进行补偿,并采用近邻参考图像预去噪方式消除噪声对补

偿精度的影响.整个运动补偿过程分三步实现:第一步,采用静态图像去噪算法对当前待处理视频图像的前后近邻图像去噪,实现参考图像预去噪;第二步,将当前图像划分为宏块,并在前后近邻参考图像中基于欧氏距离为每个宏块搜索相似块;第三步,根据当前待处理图像内宏块所处位置,利用第二步搜索得到的相似块构建参考图像,完成运动补偿.常用的基于块匹配的视频图像运动补偿算法所构建的参考图像易出现明显块效应,会对后续基于CNN的特征提取造成影响.为解决这一问题,本文采用基于块中心像素的参考图像构建方法^[11]实现近邻参考图像的构建.

2.2 网络构成

本文提出的视频去噪卷积神经网络采用了一种双路处理结构,网络结构如图1所示.整个网络由分路层、双路稠密连接单元(Dual-path Cross Connection Unit, DCU)、双路降维单元以及合路层组成.其中,分路层由两组步长分别为1和2的卷积组成;分路层后连续串行连接三个用于提取浅层特征的双路稠密连接单元,并通过稠密连接技术,使每个单元均实现特征复用,不过这也导致了特征通道的激增.为解决使用稠密连接带来的问题,本文构建了双路降维单元用于减少特征通道数量,实现特征降维.在对两路特征进行降维后,再串行连接三个用于提取深层特征的双路稠密连接单元,最终用合路层融合两路不同分辨率的深层特征.此外,本文使用残差学习技术,对合路层输出特征和当前视频图像进行全局跨越连接得到最终结果.在本文所提出的去噪网络中,高、低分辨率特征在双路稠密连接单元和降维单元中实现了信息互补,从而能够有效控制网络性能的降低.

2.3 算法实现

2.3.1 分路层和合路层

为充分利用视频时空先验信息构建去噪网络,本文将运动补偿后得到的多帧参考视频图像与当前图像通过拼接的方式融合,将融合后的信号输入分路层,产生高、低分辨率两组特征.分路层由两组步长分别为1和2的卷积构成.将这两组卷积分别作用于融合后的视频,产生与原视频图像分辨率相同的特征作为高分辨率特征,并产生分辨率为原分辨率四分之一的特征作为低分辨率特征.

合路层由两组卷积构成,实现两路特征的融合.其中,高分辨率特征经过卷积后产生与视频图像色彩维度相同的特征,低分辨率特征经过卷积后再进行上采样产生与视频图像色彩维度和分辨率均相同的特征,将两路特征相加得到融合特征.

2.3.2 双路稠密连接单元

本文所设计的双路稠密连接单元,其结构如图2所

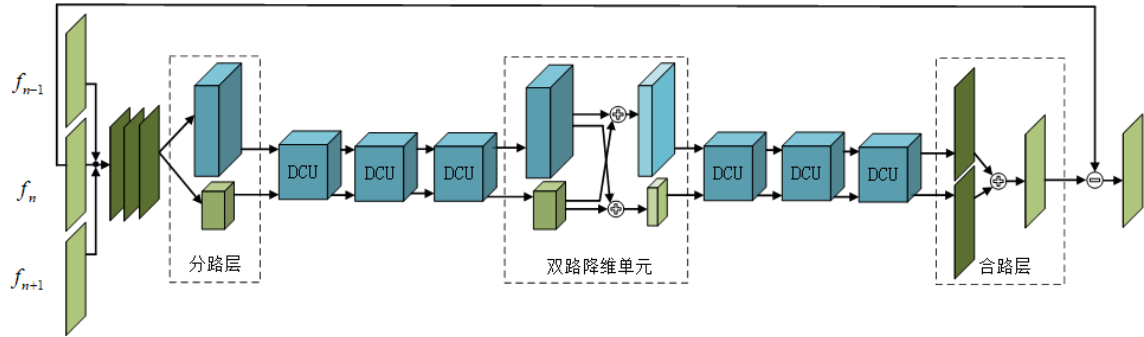


图1 本文所提出的视频去噪网络实现结构图

示,实现步骤:第一步,采用三组步长为1和一组步长为2的卷积并行处理两路不同分辨率的特征,其中,一组步长为2的卷积处理高分辨率输入特征并产生低分辨率特征,一组步长为1的卷积处理低分辨率输入特征并结合最近邻上采样产生高分辨率特征,另外两组步长为1的卷积分别处理两路高、低分辨率的输入特征,产生一路高分辨率特征和一路低分辨率特征;第二步,将第一步产生的具有相同分辨率的特征进行交互融合,得到融合特征;第三步,采用两组步长为1的卷积并行处理融合后的特征;第四步,通过稠密连接技术,将输入特征和第二步及第三步产生的高、低分辨率特征以通道拼接的方式进行连接,即实现特征复用,得到最终的输出特征。

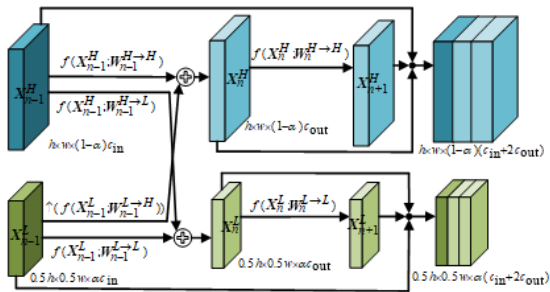


图2 双路稠密连接单元结构图

在双路稠密连接单元中使用稠密连接技术能够有效降低浮点运算次数,并减少参数量,同时还能提高网络训练速度.但特征复用会使特征通道维度激增,为解决该问题,本文在每个单元中仅使用一次特征复用,并在网络中设计和使用双路降维单元实现通道降维.

此外,本文在双路稠密连接单元中的每组卷积后都采用了 Batch Normalization^[18]和 ReLU^[19]激活函数,分别实现批量归一化和去线性化.

2.3.3 双路降维单元

为解决稠密连接所引起的特征通道激增问题,本文设计双路降维单元对特征通道进行降维,该单元的结构如图3所示.降维单元由四组卷积组成,通过对高、低分辨率特征分别进行卷积实现特征通道降维.将

卷积输出的特征进行融合得到降维后的高、低分辨率特征.该降维单元分别采用权重为 $W^{H \rightarrow H}$ 和 $W^{H \rightarrow L} \in \mathbf{R}^{k \times k \times c_{in} \times c_{out}}$ 的卷积对高分辨率特征进行降维,并分别采用权重为 $W^{L \rightarrow H}$ 和 $W^{L \rightarrow L} \in \mathbf{R}^{k \times k \times c_{in} \times c_{out}}$ 的卷积对低分辨率特征进行降维,这里, k , c_{in} 和 c_{out} 分别表示卷积大小、输入特征通道维度以及输出特征通道维度,其中 $c_{in} > c_{out}$.

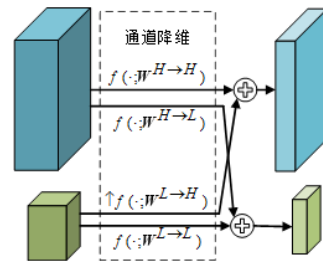


图3 双路降维单元结构图

2.4 轻量化效率分析

本文所设计的双路稠密连接单元采用六组卷积以双路并行的方式处理高、低分辨率特征,与直接采用两层串行连接的标准卷积结构相比,能够有效减少浮点运算次数和网络参数量.针对这两个指标,本文对这两种结构进行详细比较.

双路稠密连接单元的输入是分辨率不同的两路特征,一路是高分辨率特征 $X_{n-1}^H \in \mathbf{R}^{h \times w \times (1-\alpha)c_{in}}$,另一路是低分辨率特征 $X_{n-1}^L \in \mathbf{R}^{0.5h \times 0.5w \times \alpha c_{in}}$,其中 h , w 和 c_{in} 分别表示输入该单元特征的高度、宽度和通道数量, α 表示低分辨率特征通道数量在输入特征通道总数量中的比率 ($0 < \alpha < 1$).该单元内的特征处理过程包括高低两种分辨率特征交互、交互后特征处理以及特征复用.特征交互部分由四组不同的卷积构成,它们的权重分别为 $W_{n-1}^{H \rightarrow H}$, $W_{n-1}^{H \rightarrow L}$, $W_{n-1}^{L \rightarrow L}$ 和 $W_{n-1}^{L \rightarrow H}$,产生的高分辨率特征和低分辨率特征分别为 $X_n^H \in \mathbf{R}^{h \times w \times (1-\alpha)c_{out}}$ 和 $X_n^L \in \mathbf{R}^{0.5h \times 0.5w \times \alpha c_{out}}$,其中 c_{out} 表示经卷积处理后所产生特征的通道数量.交互后特征处理部分由两组处理不同分辨率特征的卷积构成,分别用于处理高分辨率特征并产生新的高分辨率特征 $X_{n+1}^H \in \mathbf{R}^{h \times w \times (1-\alpha)c_{out}}$,以及处

理低分辨率特征以产生新的低分辨率特征 $\mathbf{X}_{n+1}^L \in \mathbf{R}^{0.5h \times 0.5w \times ac_{out}}$, 该部分中两组不同卷积的权重分别为 $\mathbf{W}_n^{H \rightarrow H}$ 和 $\mathbf{W}_n^{L \rightarrow L}$. 双路稠密连接单元的处理过程为

$$\begin{cases} \mathbf{X}_n^H = f(\mathbf{X}_{n-1}^H; \mathbf{W}_{n-1}^{H \rightarrow H}) + \uparrow(f(\mathbf{X}_{n-1}^L; \mathbf{W}_{n-1}^{L \rightarrow H})) \\ \mathbf{X}_n^L = f(\mathbf{X}_{n-1}^L; \mathbf{W}_{n-1}^{L \rightarrow L}) + f(\mathbf{X}_{n-1}^H; \mathbf{W}_{n-1}^{H \rightarrow L}) \\ \mathbf{X}_{n+1}^H = f(\mathbf{X}_n^H; \mathbf{W}_n^{H \rightarrow H}) \\ \mathbf{X}_{n+1}^L = f(\mathbf{X}_n^L; \mathbf{W}_n^{L \rightarrow L}) \end{cases} \quad (1)$$

其中, $f(\cdot)$ 表示卷积操作; $\uparrow(\cdot)$ 为最近邻上采样. 此过程中浮点运算次数 $F(\cdot)$ 为

$$\begin{cases} F(f(\mathbf{X}_{n-1}^H; \mathbf{W}_{n-1}^{H \rightarrow H})) = h \times w \times k^2 \times (1-\alpha)^2 \times c_{in} \times c_{out} \\ F(f(\mathbf{X}_{n-1}^H; \mathbf{W}_{n-1}^{H \rightarrow L})) = \frac{h}{2} \times \frac{w}{2} \times k^2 \times \alpha \times (1-\alpha) \times c_{in} \times c_{out} \\ F(f(\mathbf{X}_{n-1}^L; \mathbf{W}_{n-1}^{L \rightarrow H})) = \frac{h}{2} \times \frac{w}{2} \times k^2 \times (1-\alpha) \times \alpha \times c_{in} \times c_{out} \\ F(f(\mathbf{X}_{n-1}^L; \mathbf{W}_{n-1}^{L \rightarrow L})) = \frac{h}{2} \times \frac{w}{2} \times k^2 \times \alpha^2 \times c_{in} \times c_{out} \\ F(f(\mathbf{X}_n^H; \mathbf{W}_n^{H \rightarrow H})) = h \times w \times k^2 \times (1-\alpha)^2 \times c_{out}^2 \\ F(f(\mathbf{X}_n^L; \mathbf{W}_n^{L \rightarrow L})) = \frac{h}{2} \times \frac{w}{2} \times k^2 \times \alpha^2 \times c_{out}^2 \end{cases} \quad (2)$$

其中, k 表示卷积核大小. 稠密连接不涉及浮点运算, 因此双路稠密连接单元的浮点运算次数 F_{DCU} 为

$$\begin{aligned} F_{DCU} &= F(f(\mathbf{X}_{n-1}^H; \mathbf{W}_{n-1}^{H \rightarrow H})) + F(f(\mathbf{X}_{n-1}^H; \mathbf{W}_{n-1}^{H \rightarrow L})) \\ &\quad + F(f(\mathbf{X}_{n-1}^L; \mathbf{W}_{n-1}^{L \rightarrow H})) + F(f(\mathbf{X}_{n-1}^L; \mathbf{W}_{n-1}^{L \rightarrow L})) \\ &\quad + F(f(\mathbf{X}_n^H; \mathbf{W}_n^{H \rightarrow H})) + F(f(\mathbf{X}_n^L; \mathbf{W}_n^{L \rightarrow L})) \\ &= \left[(1-\alpha)^2 + \frac{1}{2}\alpha(1-\alpha) + \frac{1}{4}\alpha^2 \right] hwk^2 c_{in} c_{out} \\ &\quad + \left[(1-\alpha)^2 + \frac{1}{4}\alpha^2 \right] hwk^2 c_{out}^2 \end{aligned} \quad (3)$$

此外, 两层连续串行连接的标准卷积所需要的浮点计算次数 F_V 为

$$F_V = h \times w \times k^2 \times c_{in} c_{out} + h \times w \times k^2 \times c_{out}^2 \quad (4)$$

由式(3)和式(4)可知, 双路稠密连接单元与直接串行连接两层标准卷积的结构之间的浮点计算次数差为

$$\begin{aligned} \Delta_F &= F_{DCU} - F_V \\ &= \left[\frac{3}{4}\alpha - \frac{3}{2} \right] ahwk^2 c_{in} c_{out} + \left[\frac{5}{4}\alpha - 2 \right] ahwk^2 c_{out}^2 \end{aligned} \quad (5)$$

由于 $0 < \alpha \leq 1$, 所以 $\frac{3}{4}\alpha - \frac{3}{2} < 0$, 并且 $\frac{5}{4}\alpha - 2 < 0$, 因此 $\Delta_F < 0$, 表明双路稠密连接单元能够节约浮点运算次数.

除去浮点运算次数, 双路稠密连接单元中所使用卷积的参数量为

$$\begin{aligned} P_{DCU} &= k^2 c_{in} c_{out} \left[(1-\alpha)^2 + 2\alpha(1-\alpha) + \alpha^2 \right] \\ &\quad + k^2 (1-\alpha)^2 c_{out}^2 + k^2 \alpha^2 c_{out}^2 \\ &= k^2 c_{in} c_{out} + k^2 (1-\alpha)^2 c_{out}^2 + k^2 \alpha^2 c_{out}^2 \end{aligned} \quad (6)$$

在同样输出通道数为 $c_{in} + 2c_{out}$ 的特征时, 直接串行两层标准卷积所需参数量 P_V 为

$$P_V = k^2 c_{in} c_{out} + k^2 (c_{in} + 2c_{out}) c_{out} \quad (7)$$

由式(6)和式(7)可知, 双路稠密连接单元与直接串行连接两层标准卷积的结构之间的参数量之差为

$$\begin{aligned} \Delta_P &= P_{DCU} - P_V \\ &= k^2 c_{out} \left[(1 + 2\alpha^2 - 2\alpha) c_{out} - (c_{in} + 2c_{out}) \right] \end{aligned} \quad (8)$$

由于 $0 < \alpha \leq 1$, 所以 $(1 + 2\alpha^2 - 2\alpha) c_{out} - (c_{in} + 2c_{out}) < 0$, 因此 $\Delta_P < 0$, 表明双路稠密连接单元所需参数量小于两层连续串行堆叠的标准卷积所需的参数量.

3 仿真实验

3.1 数据集

本文用 142 个 RGB 彩色视频构建训练集和验证集, 这些视频的分辨率均为 1920×1080 . 这两个数据集的构建过程分四步实现: 第一步, 对每个视频每间隔 2 秒提取连续 16 帧视频图像, 共产生 728 组图像; 第二步, 将每个图像用抗锯齿滤波器进行滤波; 第三步, 对滤波后的图像进行空间降采样, 得到分辨率为 960×540 的图像; 第四步, 选取 97% 的降采样图像组成训练集, 用剩余图像组成验证集.

本文采用三个数据集测试算法性能, 包括图 4 所示由 5 个彩色视频 Boat, Bus, Foreman, Tennis 和 Football 所组成的 Five-Seq 测试集(注: 这 5 个视频均选自 Derf's Test Media collection 数据集 <https://media.xiph.org/video/derf>); 另外两个测试集分别为 Set8 数据集^[20]和 DAVIS 数据集^[21].



图 4 Five-Seq 测试集

注: 从左至右: Boat, Bus, Foreman, Tennis, Football

3.2 实验设置

本文通过将具有零均值、不同标准差的高斯噪声加入原始无噪声视频而形成具有不同噪声强度的视频. 高斯噪声是日常视频信号采集和处理过程中最常见的一种噪声, 被广泛用于视频去噪、质量增强等算法的性能评估中. 为充分验证算法的鲁棒性, 本文采用标准差分别为 10, 20, 30 和 40 的高斯噪声进行性能测试.

本文方法首先采用图像去噪方法 DnCNN^[10]对当前待去噪视频图像的前、后各一帧图像进行预去噪, 再将

预去噪得到的两帧图像与当前图像进行运动补偿. 预去噪算法中所使用的参数为其默认设置, 运动补偿中所使用的搜索窗大小为 41×41 , 搜索块的大小为 20×20 , 搜索步长为 1.

本文在卷积神经网络训练过程中, 将训练集和验证集中所有图像按步长为 50 裁剪为 120×120 的图像块, 并将其像素值归一化在 $[0, 1]$ 范围内. 网络共训练 20 个周期, 学习率在训练阶段下降两次: 第一次下降在第 12 周期, 学习率从 10^{-3} 降为 10^{-4} ; 第二次下降在第 17 周期, 学习率从 10^{-4} 降为 10^{-7} . 本文训练过程中使用的优化器为 Adam^[22] 优化器. 在整个网络中, 第 1 个双路稠密连接单元中四组并行处理不同分辨率特征的卷积核大小均为 3×3 , 每组中的卷积数量都是 216, 两组用于并行处理交互特征的卷积核大小均为 3×3 , 每组中卷积数量都是 60. 低频特征通道数占比为 $\alpha = 0.5$. 所有卷积神经网络训练的初始化均采用 He 等人^[23] 提出的权重初始化方法, 并采用均方误差作为损失函数. 所有神经网络均在配有 NVIDIA 2080Ti (GPU) 和 Xeon E5-2680 (CPU) 的服务器上进行训练和测试.

3.3 结果分析

本文采用两种客观评价指标衡量算法性能, 分别为峰值信噪比 (PSNR) 和结构相似度 (SSIM), 其中 PSNR 通过经典的视频 PSNR 计算方法^[8,11] 得到, 即

$$\text{PSNR} = 10 \log_{10} \left(\frac{\sum_{(x,t) \in X \times T} (y(x,t) - \hat{y}(x,t))^2}{|X||T|} \right) \quad (9)$$

其中, X 和 T 分别表示视频的空间坐标和时间坐标; y 和 \hat{y} 分别表示原始视频和去噪后的视频. 此外, 本文通过比较不同方法的 FLOPs、网络参数量 (Parameters)、处理速度 (FPS) 和内存使用量 (Memory) 来衡量算法复杂度.

本文将所提出的方法与四种主流去噪方法进行对比, 包括: BM3D^[1]、DnCNN^[10]、VNLnet^[11] 和 FastDVDnet^[20], 其中 VNLnet 和 FastDVDnet 是基于卷积神经网络的视频去噪方法, 被直接用于实现视频整体去噪, 而两种静态图像去噪方法 BM3D 和 DnCNN 则被应用于每帧图像进而实现视频去噪.

表 1 列出了不同去噪方法在用于去除高斯噪声时得到的结果. 由表 1 的结果可以发现, 在所有对比方法中, 本文方法在多数情况下均达到了较好的去噪效果. 除了比较客观评价指标, 本文还展示了不同方法在去除噪声时得到的视觉结果 (如图 5 和图 6 所示), 两个结果分别提取自 Five-Seq 数据集中 Tennis 视频第 120 帧和 DAVIS 测试集中 Chamaleon 视频第 12 帧. 从这组结果可以发现, 本文方法和 FastDVDnet 复原出了更多的图像纹理细节, 而其他方法在去噪后得到的结果比较模糊, 无法良好呈现出细节内容. 通过以上对比也证明了本文方法达到了较好的去噪效果.

除了对比不同方法之间的去噪性能, 本文还针对同样使用了前、后两帧视频图像作为先验信息实现视频去噪的 VNLnet 方法进行了网络参数量、运行速度和内存使用量的对比, 结果如表 2 所示, 其中 FPS 为每秒处理帧数, 是连续处理 500 组分辨率为 224×224 的图像的平均速度. 表 2 的结果表明本文方法各项指标均优

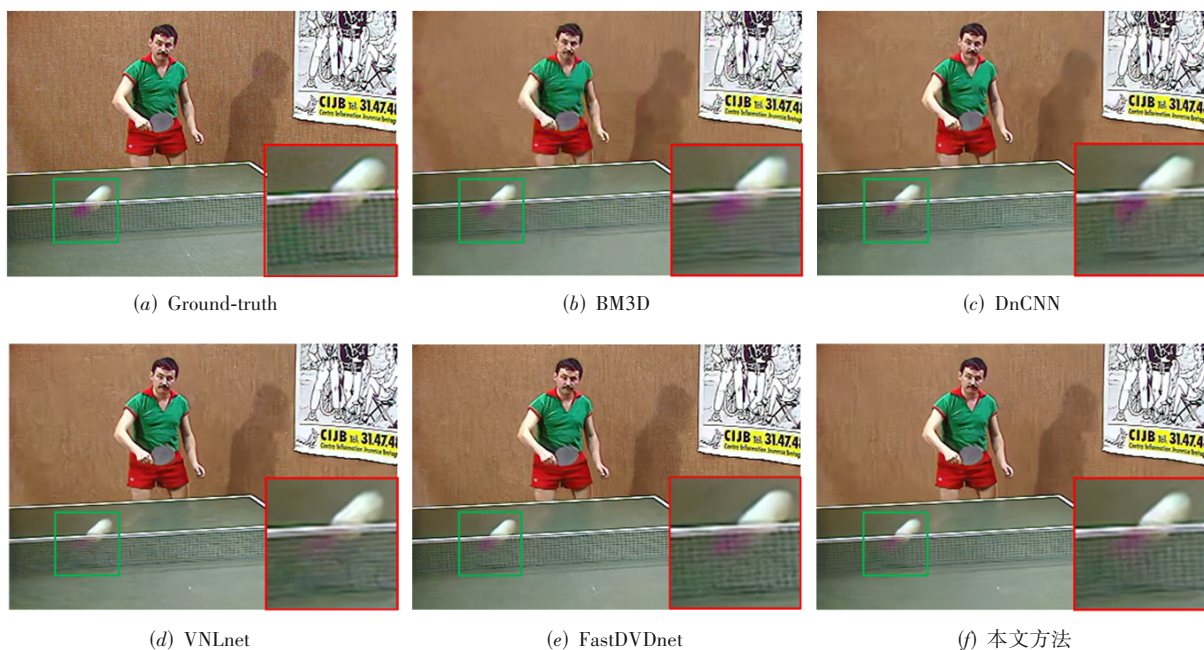


图 5 各算法在 Tennis 序列上去除标准差为 30 的高斯噪声时的视觉结果对比

表 1 去噪性能比较(PSNR/SSIM)

标准差	对比方法	Five-Seq	Set8	DAVIS
10	BM3D	35.37/0.943 3	36.07/0.948 7	38.18/0.961 1
	DnCNN	35.45/0.945 2	36.32/0.951 4	38.41/0.962 6
	VNLnet	36.20/0.954 8	36.66/0.955 1	38.94/0.966 5
	FastDVDnet	35.87/0.956 5	36.21/0.953 4	38.56/0.965 5
	本文方法	36.62/0.958 6	37.00/0.958 6	39.38/0.969 9
20	BM3D	31.57/0.881 0	32.36/0.898 9	34.48/0.919 3
	DnCNN	31.73/0.886 5	32.76/0.907 1	34.80/0.924 7
	VNLnet	32.73/0.907 8	33.20/0.914 3	35.49/0.933 8
	FastDVDnet	32.97/0.919 6	33.24/0.918 9	35.60/0.931 2
	本文方法	33.17/0.914 8	33.30/0.914 3	36.10/0.942 2
30	BM3D	29.47/0.828 5	30.27/0.854 9	32.39/0.880 9
	DnCNN	29.71/0.833 8	30.70/0.907 1	32.72/0.889 3
	VNLnet	30.77/0.864 0	31.21/0.876 4	33.47/0.902 3
	FastDVDnet	31.07/0.879 0	31.49/0.888 1	33.86/0.909 8
	本文方法	31.21/0.875 2	31.75/0.890 0	34.20/0.916 6
40	BM3D	27.81/0.770 2	28.61/0.803 8	30.57/0.833 8
	DnCNN	28.31/0.786 6	29.34/0.830 4	31.36/0.857 5
	VNLnet	29.41/0.822 1	29.80/0.841 5	32.07/0.874 2
	FastDVDnet	29.80/0.845 6	30.27/0.860 0	32.67/0.890 3
	本文方法	29.87/0.839 1	30.41/0.860 3	32.83/0.891 9

于 VNLnet 方法,由此表明本文方法实现了视频去噪网络的轻量化。

表 2 网络参数量、运行速度和内存使用量对比

比较指标	VNLnet	本文方法
FLOPs (G)	226.85	52.64
Parameters (K)	4 518.05	2 340.94
FPS (GPU)	38.46	84.67
FPS (CPU)	1.38	3.71
Memory (MB)	1 654.32	378.55

为充分验证本文所提出的双路稠密连接单元和双路降维单元的性能和轻量化作用,以及视频先验信息对去噪性能的影响,本文通过三个消融实验分别验证其有效性. 消融实验 1 中用四组 3×3 卷积替换双路稠密连接单元,四组卷积两两串行堆叠后再并行处理两路不同分辨率特征. 除替换双路稠密连接单元,该消融实验还采用两组并行 3×3 卷积对双路降维单元进行了替换. 消融实验 2 中用两组并行 3×3 卷积替换双路降维单元,并进行了性能对比. 在消融实验 3 中,本文不再将前、后两帧视频图像作为先验信息用于当前帧的去噪,仅单独使用当前视频图像实现去噪.

本文将标准差为 30、测试集为 Five-Seq 的三组消融实验结果展示在表 3 中. 从表 3 中的结果可以看到,双路稠密连接单元和双路降维单元的使用不仅可以带来去噪性能的提升,而且能够有效降低算法复杂度. 此外,视频先验信息的使用也有助于去噪性能的提升.

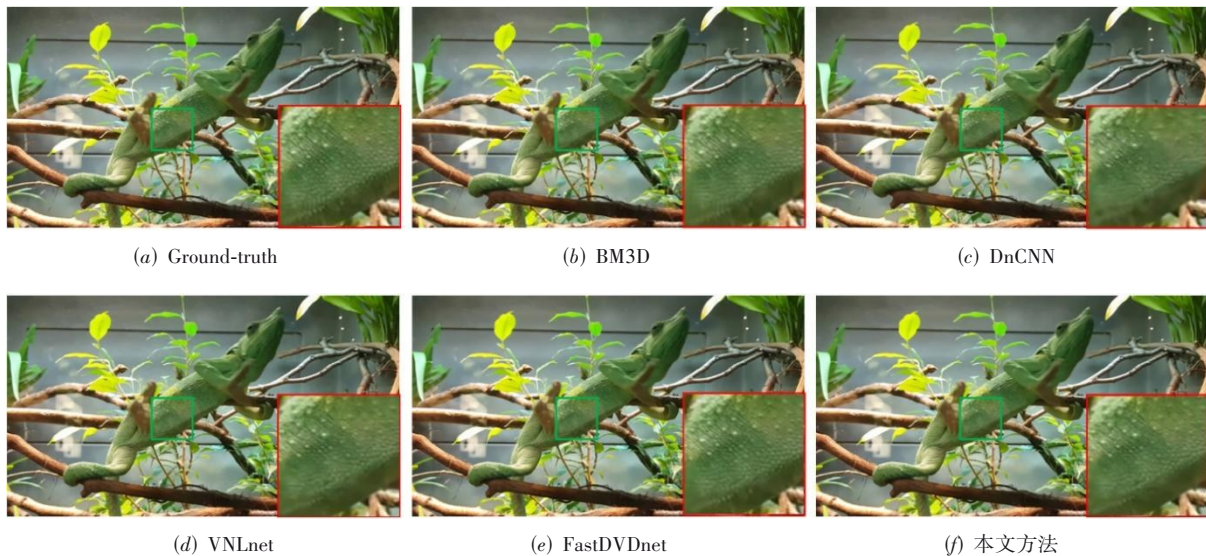


图 6 各算法在 Chamaleon 序列上去除标准差为 20 的高斯噪声时的视觉结果对比

表3 消融实验结果对比

比较指标	消融实验1	消融实验2	消融实验3	本文方法
PSNR	31.13	31.18	29.87	31.21
SSIM	0.873 8	0.874 1	0.841 1	0.875 2
FLOPs (G)	86.53	88.01	52.27	52.64
Parameters (K)	2 758.75	3 307.82	2 329.27	2 340.94
Memory (MB)	766.15	442.29	378.55	378.55
FPS (GPU)	68.37	62.46	86.54	84.67
FPS (CPU)	3.58	2.47	3.79	3.71

4 结论

本文提出了一种基于先验信息和卷积神经网络的轻量化视频去噪方法,利用前后视频图像之间的时空先验信息构建去噪网络,并针对卷积神经网络参数量大、浮点运算次数多以及占用内存大的问题设计了轻量化的实现结构,在达到较高去噪性能的同时,降低了复杂度.实验结果表明,本文方法在客观评价指标、主观视觉结果和算法复杂度指标上均有明显优势.

参考文献

- [1] DABOV K, FOI A, KATKOVNIK V, et al. Image denoising by sparse 3-D transform-domain collaborative filtering [J]. *IEEE Transactions on Image Processing*, 2007, 16(8): 2080-2095.
- [2] GU S H, ZHANG L, ZUO W M, et al. Weighted nuclear norm minimization with application to image denoising [C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 2862-2869.
- [3] MAIRAL J, BACH F, PONCE J, et al. Non-local sparse models for image restoration[C]//2009 IEEE 12th International Conference on Computer Vision. Kyoto: IEEE, 2009: 2272-2279.
- [4] DONG W S, ZHANG L, SHI G M, et al. Nonlocally centralized sparse representation for image restoration[J]. *IEEE Transactions on Image Processing*, 2013, 22(4): 1620-1630.
- [5] BUADES A, COLL B, MOREL J M. Nonlocal image and movie denoising[J]. *International Journal of Computer Vision*, 2008, 76(2): 123-139.
- [6] WEISS Y, FREEMAN W T. What makes a good model of natural images? [C]//2007 IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis: IEEE, 2007: 1-8.
- [7] DABOV K, FOI A, EGIAZARIAN K. Video denoising by sparse 3D transform-domain collaborative filtering[C]// *Proceedings of the 2007 European Signal Processing Conference*. Poznan: IEEE, 2007: 145-149.
- [8] MAGGIONI M, BORACCHI G, FOI A, et al. Video denoising, deblocking, and enhancement through separable 4-D nonlocal spatiotemporal transforms[J]. *IEEE Transactions on Image Processing*, 2012, 21(9): 3952-3966.
- [9] ARIAS P, MOREL J M. Towards a Bayesian Video Denoising Method[M]// *Advanced Concepts for Intelligent Vision Systems*. Cham: Springer International Publishing, 2015: 107-117.
- [10] ZHANG K, ZUO W M, CHEN Y J, et al. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising[J]. *IEEE Transactions on Image Processing*, 2017, 26(7): 3142-3155.
- [11] DAVY A, EHRET T, MOREL J M, et al. A non-local CNN for video denoising[C]//2019 IEEE International Conference on Image Processing. Taipei: IEEE, 2019: 2409-2413.
- [12] IANDOLA F N, MOSKEWICZ M W, ASHRAF K, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size[EB/OL]. (2016)[2021]. <http://arxiv.org/abs/1602.07360>.
- [13] HOWARD A, ZHU M, CHEN B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017) [2021]. <http://arxiv.org/abs/1704.04861>.
- [14] CHOLLET F. Xception: Deep learning with depthwise separable convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 1800-1807.
- [15] ZHANG X, ZHOU X, LIN M, et al. ShuffleNet: An extremely efficient convolutional neural network for mobile devices[EB/OL]. (2017) [2021]. <http://arxiv.org/abs/1707.01083>.
- [16] HAN K, WANG Y H, TIAN Q, et al. GhostNet: More features from cheap operations[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020: 1577-1586.

- [17] 孟球, 徐磊, 郭嘉阳. 一种基于改进的 MobileNetV2 网络语义分割算法[J]. 电子学报, 2020, 48(9): 1769-1776.
MENG L, XU L, GUO J Y. Semantic segmentation algorithm based on improved MobileNetV2[J]. Acta Electronica Sinica, 2020, 48(9): 1769-1776. (in Chinese)
- [18] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]//Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37. New York: ACM, 2015: 448-456.
- [19] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with deep convolutional neural networks[C]//Proceedings of the 2012 International Conference on Neural Information Processing Systems. Doha: Curran Associates Inc., 2012: 1097-1105.
- [20] TASSANO M, DELON J, VEIT T. FastDVDnet: Towards real-time deep video denoising without flow estimation[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020: 1351-1360.
- [21] PONT-TUSET J, PERAZZI F, CAELLES S, et al. The 2017 DAVIS challenge on video object segmentation [EB/OL]. (2018)[2021]. <https://arxiv.org/abs/1803.00557>.
- [22] KINGMA D, BA J. Adam: A method for stochastic optimization[EB/OL]. (2017) [2021]. <https://128.84.21.199/abs/1412.6980v4>.
- [23] HE K M, ZHANG X Y, REN S Q, et al. Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification[C]//2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 1026-1034.

作者简介



申屠敏健 男, 1995年9月出生, 浙江东阳人. 电子科技大学信息与通信工程学院硕士研究生. 主要研究方向为深度学习与视频去噪.
E-mail: shentu_mj@163.com



朱强 男, 1997年2月出生, 甘肃定西人. 电子科技大学信息与通信工程学院博士研究生. 主要研究方向为图像及视频超分辨率.
E-mail: zhuqiang@std.uestc.edu.cn



朱树元 男, 1980年1月出生, 河南平顶山人. 电子科技大学信息与通信工程学院教授. 主要研究方向为图像视频处理、图像视频编码等.
E-mail: eezsy@uestc.edu.cn