计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2024,33(6):143−152 [doi: 10.15888/j.cnki.csa.009533]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

局部与全局相融合的孪生低照度视频增强网络①

竺钰成, 杨　羊

(浙江理工大学 信息科学与工程学院, 杭州 310018)
通信作者: 杨　羊, E-mail: yangyang0070@zstu.edu.cn

摘　要: 在低照度环境下拍摄到的视频往往有对比度低、噪点多、细节不清晰等问题, 严重影响后续的目标检测、分割等计算机视觉任务. 现有的低照度视频增强方法大都是基于卷积神经网络构建的, 由于卷积无法充分利用像素之间的长程依赖关系, 生成的视频往往会有部分区域细节丢失、颜色失真的问题. 针对上述问题, 提出了一种局部与全局相融合的孪生低照度视频增强网络模型, 通过基于可变形卷积的局部特征提取模块来获取视频帧的局部特征, 并且设计了一个轻量级自注意力模块来捕获视频帧的全局特征, 最后通过特征融合模块对提取到的局部特征和全局特征进行融合, 指导模型能生成颜色更真实、更具细节的增强视频. 实验结果表明, 本方法能有效提高低照度视频的亮度, 生成颜色和细节都更丰富的视频, 并且在峰值信噪比和结构相似性等评价指标中也都优于近几年提出的方法.

关键词: 卷积神经网络; 低照度视频增强; 孪生网络; 自注意力机制; 特征融合

引用格式: 竺钰成,杨羊.局部与全局相融合的孪生低照度视频增强网络.计算机系统应用,2024,33(6):143–152. http://www.c-s-a.org.cn/1003-3254/9533.html

Siamese Low-light Video Enhancement Network with Fusion of Local and Global Features

ZHU Yu-Cheng, YANG Yang

(School of Information Science and Engineering , Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: Videos captured in low illumination environments often carry problems such as low contrast, high noise, and unclear details, which seriously affect computer vision tasks such as target detection and segmentation. Most of the existing low-light video enhancement methods are constructed based on convolutional neural networks. Since convolution cannot make full use of the long-range dependencies between pixels, the generated video often suffers from loss of details and color distortion in some regions. To address the above problems, this study proposes a Siamese low-light video enhancement network coupling local and global features. The model obtains local features of video frames through a deformable convolution-based local feature extraction module and designs a lightweight self-attention module to capture the global features of video frames. Finally, the extracted local and global features are fused by a feature fusion module, which guides the model to generate enhanced videos with more realistic colors and details. The experimental results show that the proposed method can effectively improve the brightness of low-light videos and generate videos with richer colors and details. It also outperforms the methods proposed in recent years in evaluation metrics such as peak signal-to-noise ratio and structural similarity.

Key words: convolutional neural network (CNN); low-light video enhancement; Siamese network; self-attention mechanism; feature fusion

① 收稿时间: 2023-12-27; 修改时间: 2024-01-29; 采用时间: 2024-02-07; csa 在线出版时间: 2024-04-19
CNKI 网络首发时间: 2024-04-23

在夜间监控等场景中, 由于光照条件差, 相机传感器接收到的光子数量通常较少, 有时甚至无法接收到足够的光子, 这会导致监控视频缺失许多细节, 严重影响视频质量[1]. 类似的问题也会在自动驾驶等领域中出现. 为了提高视频质量, 一些方法尝试通过升级硬件来解决, 例如使用星光级相机[2], 但是这些方法成本较高, 不适合大规模应用. 因此, 越来越多的研究关注于通过算法来实现低照度视频增强.

过去, 大部分工作都集中在低照度图像增强方面[3–5]. 然而, 将低照度图像增强方法直接应用于视频领域时, 可能会产生时域一致性问题, 导致增强后的视频出现闪烁和伪影. 传统的低照度视频增强方法基于直方图均衡[6]和 Retinex 理论[7], Liu 等人[8]提出一种用于低照度视频增强的多尺度类 Retinex 算法, 在 HIS 颜色空间利用引导滤波代替多尺度 Retinex 算法中的高斯滤波, 使增强后的视频拥有更好的细节和清晰度. Dong 等人[9]对反变换后的低照度视频做图像去雾操作来实现低照度视频的增强, 并利用相邻帧之间的相关性来提高时域一致性. 然而, 这些方法可能会带来运动模糊和图像失真等问题.

随着深度学习技术的不断发展, 研究人员提出了基于深度学习的低照度视频增强方法. 在数据集收集方面, 与低照度图像数据集不同, 由于在拍摄视频时无法使用相机的长曝光功能, 因此很难获得真实的动态低照度和正常照度视频数据集. 为解决上述问题, Chen 等人[10]收集了一个静态低照度视频数据集 dark row video (DRV), DRV 数据集收集了包括道路、行人、车辆、花草、店铺等多种场景的 202 个低照度视频图像序列及其对应的长曝光图像, 其中包含用于训练的静态低照度视频图像序列和用于测试的动态低照度视频图像序列. Jiang 等人[11]设计了一种新的视频采集设备, 能同时收集低照度和正常光条件下的视频对, 建立了首个动态低照度视频数据集. 与仅处理低照度图像增强不同, 低照度视频增强方法还需确保视频的时域一致性, 以防止增强后的视频出现闪烁和伪影等问题. 为应对这一问题, Lai 等人[12]提出了一种带有 Conv-LSTM 模块的深度网络, 先利用 FlowNet2[13]在训练阶段估计光流, 再利用光流扭曲技术和设计相应的损失函数从视频序列中学习时域一致性. Zhang 等人[14]用单张图像来训练基于图像的低照度视频增强模型, 通过对单张图像进行图像分割、光流预测等操作, 设计相应的损失函数来保证增强后视频的时域一致性. Li 等人[15]基于显示查找表技术设计了一个可学习的强度感知模块来保持帧间亮度一致, 防止生成的视频产生闪烁效应. 除了上述提到的方法外, Lv 等人[16]设计了一个多尺度特征提取、融合网络来实现低照度视频增强, 并在人工合成的低照度视频数据集上进行训练. Wang 等人[17]设计了一个端到端的低照度视频增强框架, 提出了一个自监督去噪模块, 同时基于 Retinex 理论实现低照度视频亮度的增强. Zheng 等人[18]提出了一种语义指导的零射低照度视频增强网络, 该网络能在没有配对的低照度和正常照度视频对数据集的情况下进行训练并实现低照度视频增强. Triantafyllidou 等人[19]提出了一种基于双 CycleGAN 的数据生成机制 SIDGAN, 能够生成动态的低照度和正常照度视频对. Ye 等人[20]使用递归监督的金字塔残差稠密结构提取多尺度空间中的上下文特征, 然后通过空间-时间特征重构子网络恢复目标帧的质量, 并保持时域一致性. Liu 等人[21]提出了一种基于合成事件引导的低照度视频增强方法, 通过使用多帧融合的事件来指导低照度视频增强. Liang 等人[22]提出了一种利用事件引导的低照度视频增强方法, 通过建立视觉信号间的时空一致性, 实现低照度条件下视频质量的提升. 虽然现有的低照度视频增强方法取得了一些进展, 但这些方法几乎都是基于卷积神经网络, 而卷积神经网络受限于卷积的局部感受野, 无法捕捉视频帧的全局特征. 在增强低照度视频时, 这可能会导致某些低照度区域的细节丢失和色彩失真等问题.

本文针对上述问题提出了一个局部与全局相融合的孪生网络模型. 通过基于可变形卷积的局部特征提取模块, 提取视频帧的局部特征. 并设计了一个轻量级自注意力模块, 以捕获像素之间的长程依赖关系, 获取视频帧的全局特征. 最后, 通过特征融合模块融合所提取到的局部特征和全局特征, 使模型能输出颜色更真实、细节更丰富的低照度视频增强结果. 同时, 为了解决视频的时域一致性问题, 本文采用孪生网络结构来保证增强视频的时域一致性.

1 本文方法

现有的基于深度学习的低照度视频增强方法主要是基于卷积神经网络构建的. 在进行低照度视频增强时, 卷积由于自身局部感受野的限制, 难以捕获

到视频帧的全局特征, 增强后的视频往往会产生部分区域细节丢失、颜色失真的情况. 本文提出了一种局部与全局相融合的孪生低照度视频增强网络. 整体的孪生网络结构如图 1 所示, 在训练前先通过跨帧滑动窗口模块对输入视频帧进行处理, 随后将提取到的视频帧序列输入两个相同且权重共享的局部与全局相融合网络进行权重更新. 单个局部与全局相融合网络如图 2 所示, 整体为 U 型网络结构, 先使用深度可分离卷积初步提取视频帧的浅层特征, 随后通过局部特征提取模块 (LFEM) 提取局部特征信息, 轻量级自注意力模块 (LSAM) 提取全局特征信息, 最后通过特征融合模块 (FFM) 融合提取到的局部特征和全局特征, 指导网络生成颜色更相似、细节更丰富的视频.

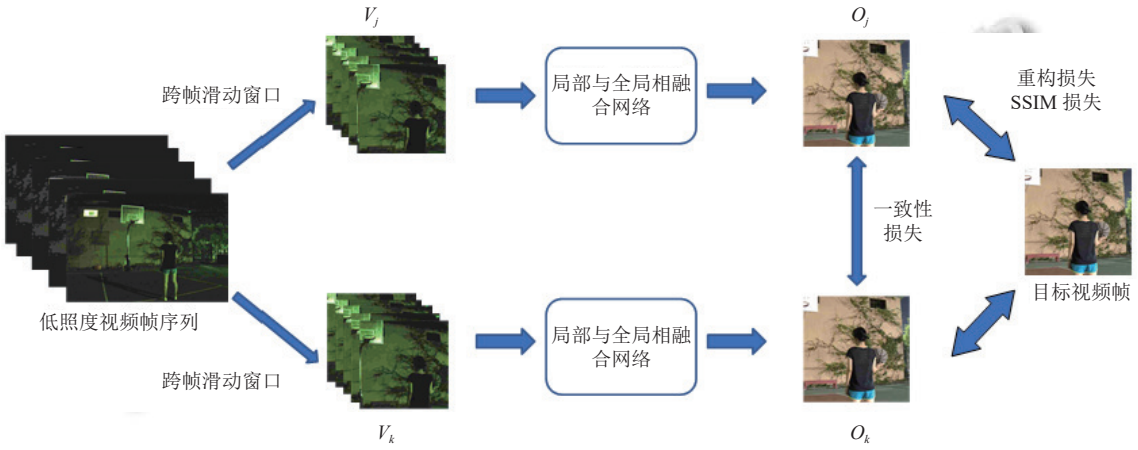


图 1 孪生网络结构

1.1 孪生网络结构

使用单个局部与全局相融合网络进行低照度视频增强可能会产生时域一致性问题. 为了确保生成视频的时域一致性, 本文采用了孪生网络结构, 如图 1 所示, 孪生网络结构含有两个相同且共享权重的局部与全局相融合网络, 接收两组通过跨帧滑动窗口模块提取的视频帧作为输入, 分别用 V_l 和 $V_{l'}$ 表示. 这两组视频帧经过局部与全局相融合网络模型进行增强, 生成增强的视频帧 O_l 和 $O_{l'}$. 在损失函数设计方面, 我们使用时域一致性损失函数来保证视频在时域上保持一致性, 减少视频的闪烁和伪影. 此外, 本文了还使用重构损失和结构相似性损失作为损失函数, 以促进网络生成更明亮、色彩正确和细节丰富的视频.

1.2 跨帧滑动窗口模块

本文采用了静态低照度视频数据集 DRV 作为实验数据集, 由于缺乏动态视频数据, 模型难以充分利用视频时间维度上的信息. 为了解决这个问题, 本文引入了一种跨视频帧的滑动窗口模块, 从静态视频中模拟具有时空信息的动态视频序列. 如图 3(a) 所示, 该模块通过在不同视频帧之间随机选择一个方向进行窗口滑动, 并截取窗口内的视频帧来模拟包含时空信息的低照度动态视频序列, 截取的视频帧序列如图 3(b) 所示. 这些模拟的动态视频序列被用作网络的训练数据, 使模型能够学习到更全面的时空特征表示, 从而提升对动态低照度视频的增强效果.

1.3 局部与全局相融合网络

局部与全局相融合网络如图 2 所示, 网络整体结构为包含编码器、解码器和跳跃连接的 U 型网络. 在编码器阶段, 输入网络的视频帧序列首先经过深度可分离卷积初步提取浅层特征, 再通过由局部特征提取模块、轻量级多头自注意力模块和特征融合模块组成的特征提取与融合模块进行特征提取与融合, 下采样使用步长为 2 的卷积层. 解码器阶段同样使用特征提取与融合模块提取与融合深层特征, 同时使用跳跃连接将编码器阶段提取到的浅层特征与解码器阶段的特征信息进行融合, 上采样采用了 Patch Expand[23]方法, 先通过线性层将特征图的特征维度增大 1 倍, 再经过重排操作将特征图的分辨率增大 1 倍, 特征图的维度缩小为输入时的 1/4.

1.3.1 局部特征提取模块 (LFEM)

局部特征指的是边缘、角落和纹理等特征信息, 它

在增强过程中有助于保留视频帧的细节. 相对于传统卷积, 可变形卷积能够自适应的调整卷积的采样位置, 适应不同物体的形状和尺寸, 实现高效的特征提取. 这种能力有助于模型更好地捕获低照度视频中的细节和纹理信息, 从而提高增强效果. 因此, 本文采用可变形卷积来提取视频帧中的局部特征. 局部特征提取模块如图 2(a) 所示, 由两个可变形卷积层、一个残差连接以及一个批归一化层组成, 并使用 GELU 作为激活函数.

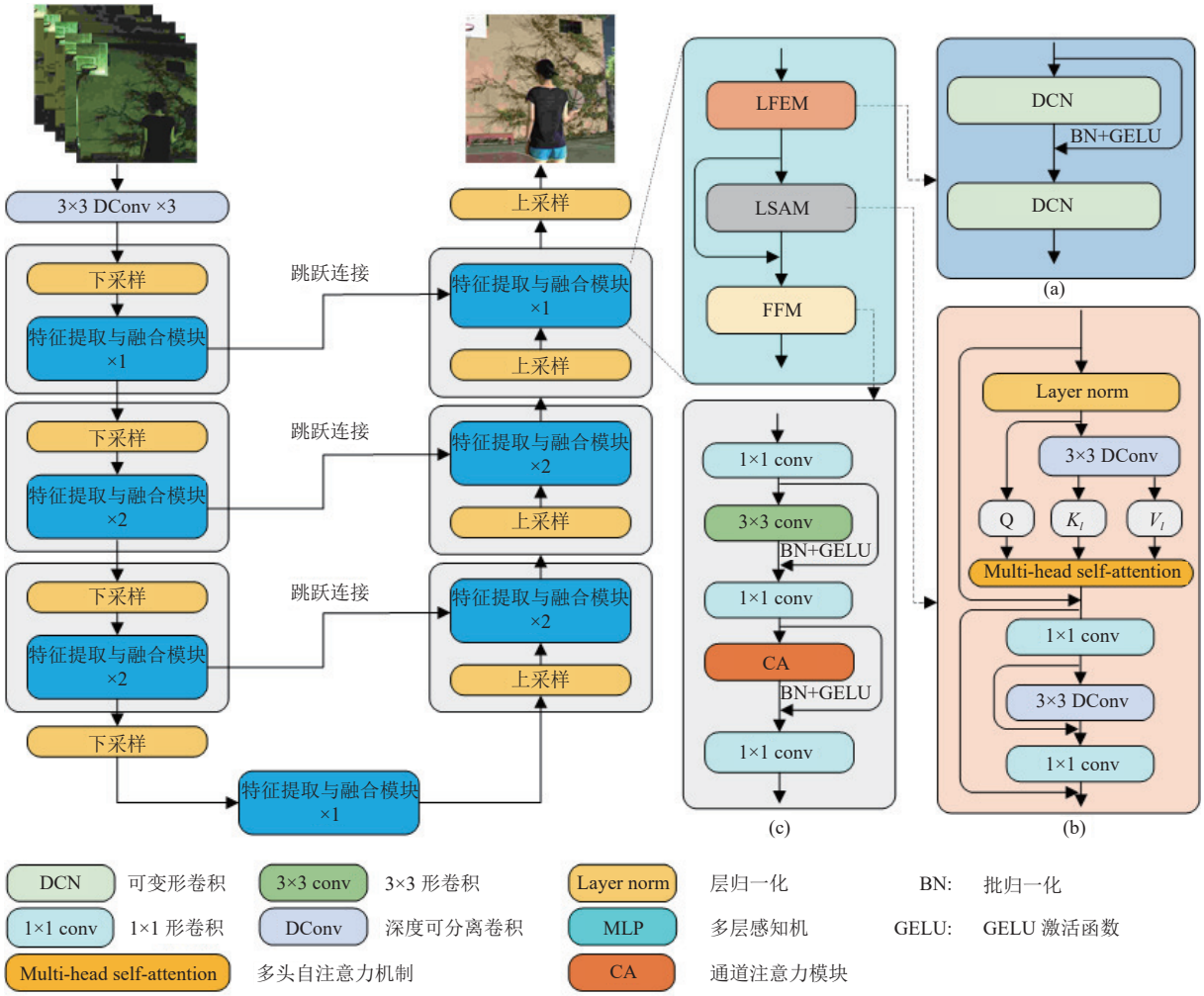


图 2 局部与全局相融合网络

1.3.2 轻量级多头自注意力模块 (LSAM)

尽管卷积在提取局部特征方面非常有效, 但其提取全局特征 (如整体颜色分布、空间布局或视频帧中特定结构的存在) 的能力有限. 如果网络无法正确地聚合全局信息并捕捉区域之间的长距离依赖关系, 可能会导致视频帧中某些区域的颜色不正确或细节不完整. 为了解决这个问题, 本文使用自注意力机制来学习视频帧内各区域之间的长距离依赖关系, 并提取全局特征.

然而, 传统的自注意力机制算法计算复杂度较高, 严重限制了输入视频帧的分辨率大小. 为了减小计算量, 使网络能输入更大分辨率的视频帧, 本文设计了一种轻量级自注意力机制, 如图 2(b) 所示. 与传统的自注意力机制相比, 本文的自注意力机制在计算键矩阵 K 和价值矩阵 V 时, 使用卷积核为 $n\times n$ 、步长为 n 的深度可分离卷积代替普通卷积, 通过生成尺寸更小的键矩阵 K_l 和价值矩阵 V_l 来减小计算量. 针对原来的多层感知机模块, 本文采用深度可分离卷积和 1×1 卷积代替, 有助于帮助模型从内部的中间特征中帮助捕捉局部和全局结构信息. 对于输入特征图 F_l, 对应的轻量级自注意力机制模块输出可以表示为:

$$Q = L(F_l) \tag{1}$$

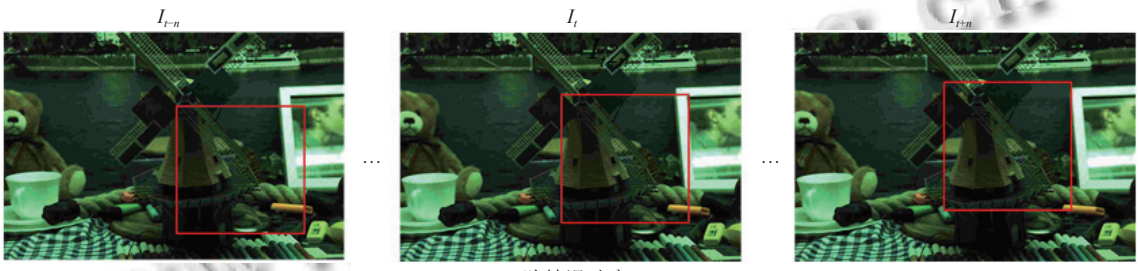
$$K_l,\ V_l = \mathrm{DConv}(L(F_l)) \tag{2}$$

$$LightSelf\text{-}Attention(Q,K_l,V_l)=Softmax\left(\frac{QK_l^{\mathrm{T}}}{\sqrt{d_k}}+B\right)V_l \tag{3}$$

其中, L表示全连接层, DConv表示深度卷积. B是可学习的相对位置偏置, 用于提供空间位置信息, d_k是K_l的维度.

1.3.3 特征融合模块 (FFM)

局部特征和全局特征分别代表视频帧的不同信息. 局部特征主要关注细节和纹理, 而全局特征则关注整体结构和场景. 将这两种特征融合起来可以将局部细节与全局语义相结合, 更好地捕捉视频帧的整体语义信息, 有助于模型更好地理解和重构视频帧的内容, 从而提高增强效果. 因此, 本文提出了一个特征融合模块 FFM. 如图 2(c) 所示, 该模块由通道注意力机制和残差模块组成. 通道注意力机制用于融合通道方向的特征信息, 残差模块包含卷积、批归一化和 GELU 激活函数, 用于增强网络的反向传播能力和泛化能力. 将局部特征和全局特征作为输入进行特征融合, 以更好地指导网络实现低照度视频增强.



(a) 跨帧滑动窗口

(b) 截取的视频帧序列

图 3 跨帧滑动窗口机制

1.4 损失函数

为了获得颜色一致、细节清晰的且具有良好时域一致性的视频, 本文设计了如下的联合损失函数:

$$L=\alpha\times L_r+\beta\times L_c+\gamma\times L_{\mathrm{ssim}} \tag{4}$$

其中, L_r表示重构损失, L_c表示时域一致性损失, L_{ssim}表示结构相似性损失, α、β、γ为可调的系数, 本文分别设置为 0.65, 0.05 和 0.15.

为了最小化输出视频帧与目标视频帧之间的差异, 本文使用了重构损失函数. 其定义如下:

$$L_r=\|I_n-I_t\|_1+\sum_{l=N}\|\phi_l(I_n)-\phi_l(I_t)\|_1 \tag{5}$$

其中, I_n表示网络的输出视频帧, I_t表示目标视频帧, ϕ_l表示从经过预训练的 VGG[24]网络模型的第l层提取的特征, N表示层数, 本文取 VGG 网络的 Conv1_1、Conv2_1、Conv3_1 和 Conv4_1 层计算损失值. 重构损失不仅在输出视频帧和目标视频帧级别上做了约束, 还使用经过预训练的 VGG 网络在视频帧的特征图级别上进行了约束, 能更好地指导网络生成和目标视频帧近似的输出视频帧.

针对输出视频的时域一致性问题, 本文设计了一个时域一致性损失函数, 其定义如下:

$$L_c=\|I_j-I_k\|_1+\sum_{l=N}\|\phi_l(I_j)-\phi_l(I_k)\|_1 \tag{6}$$

其中, I_j和I_k分别表示孪生网络输出的两个视频帧, 通过对孪生网络输出视频帧及其特征图进行约束, 减小输出视频帧之间的差异来保证视频的时域一致性.

除此之外, 为了保证输出图像和标签图像之间的结构相似性, 本文还使用了 SSIM 损失, 从亮度、对比

度和结构 3 个方面来减小输出视频帧和目标视频帧之间的差异.

$$L_{\text{ssim}} = 1 - \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(2\sigma_{xy} + C_2)} \tag{7}$$

其中, x和y分别表示输出视频帧和目标视频帧, μ表示均值, σ_x^2和σ_y^2表示方差, σ_{xy}表示协方差, C为常数.

1.5 训练过程

本文使用 PyTorch 框架来实现提出的方法. 网络总计训练了 2 500 个 epoch. 优化器选择 AdamW, 设置训练批量大小为 2, 初始学习率为 10^{-3}, 并使用余弦退火算法在训练过程中周期性地自动调整学习率, 以提高模型的性能和收敛速度. 在输入网络之前, 本文对视频帧进行了随机翻转、旋转和转置等操作, 以实现数据增强的目的.

2 实验与结果分析

2.1 数据集和评价指标

本文的实验采用了开源低照度视频数据集 DRV. 该数据集使用索尼 RX100 VI 相机, 在连拍模式下捕捉原始图像序列, 每秒约 16–18 帧的速度. 数据集包含 202 个静态低照度视频图像序列及其对应的长曝光图像对, 涵盖了道路、行人、车辆、花草、店铺等多种室内外场景. 为了公平起见, 本文采用了与 Chen 等人[10]相同的训练集、验证集和测试集划分方法, 其中 129 个序列用于训练, 24 个序列用于验证, 剩下的 49 个序列和 22 个动态低照度视频图像序列用作测试集. 为了评估本文提出的方法的有效性, 在衡量静态视频质量方面, 采用峰值信噪比 (PSNR) 和结构相似性 (SSIM) 这两个评价指标. 在评价生成视频的时域一致性时, 使用了平均亮度方差 (AB(var)) 评价指标和平均绝对亮度差 (MABD) 评价指标.

2.2 对比实验分析

我们从定量和定性两个角度出发, 将本文提出的方法和其他低照度视频增强方法进行对比实验.

2.2.1 静态视频增强定性评估实验

为了比较本文所提出的方法与其他方法在静态视频增强方面的差异, 我们和 FastLLVE 等其他 6 种低照度视频增强方法进行了比较. 如图 4 所示, 在静态低照度视频增强方面, 通过本文方法生成的视频帧在细节和颜色效果上表现更出色, 充分证明了在增强低照度视频方面, 本文提出的局部与全局相融合的孪生网络的有效性.

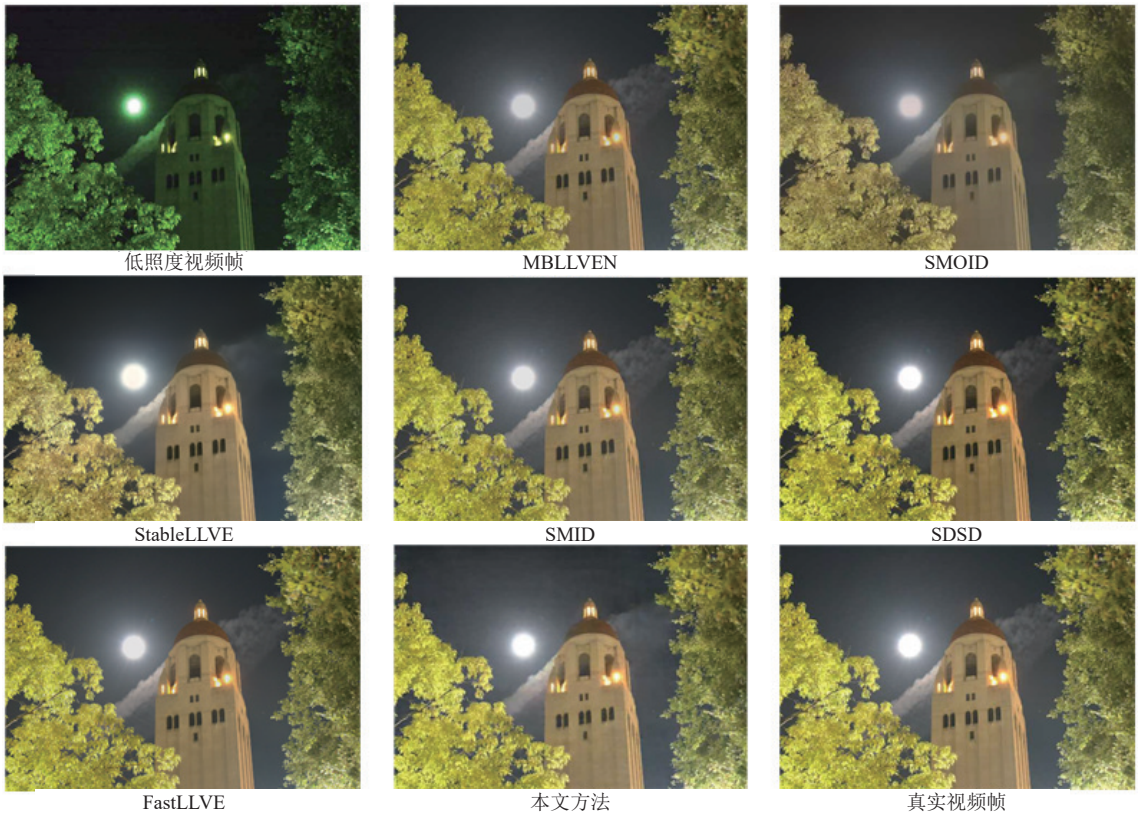
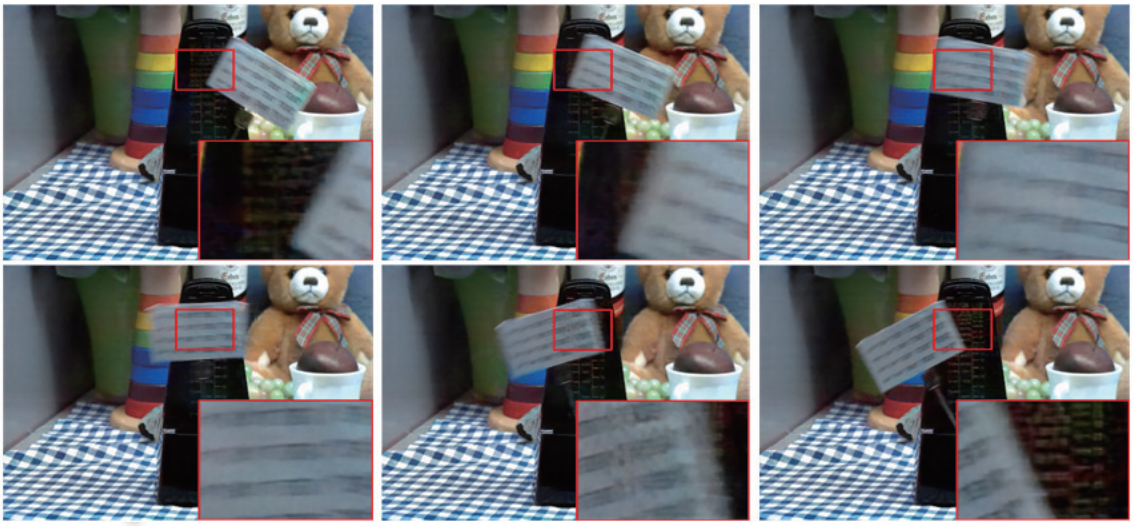


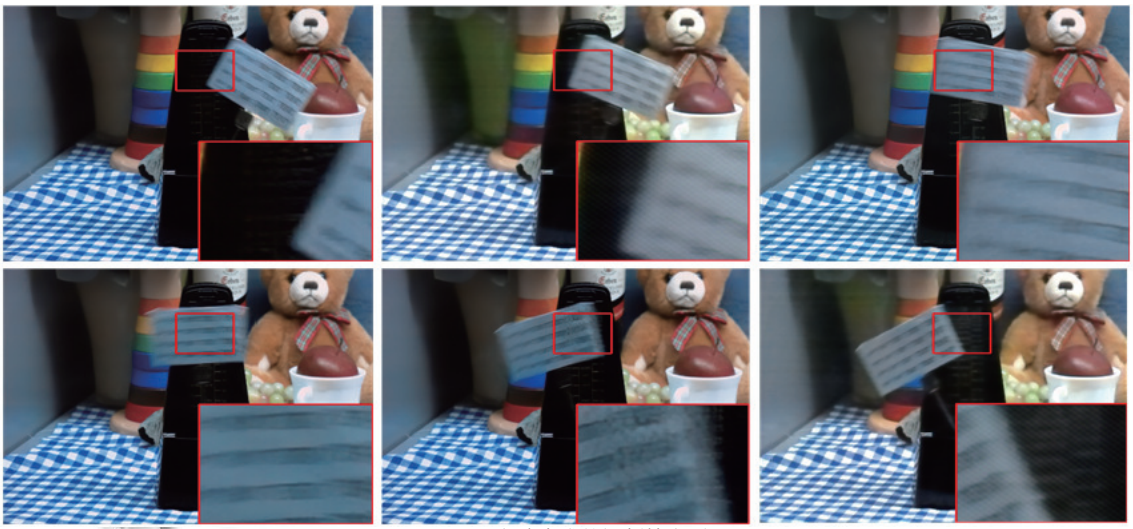
图 4 静态视频增强定性评估

2.2.2 视频时域一致性定性评估实验

为了验证所提出的方法在保证视频时域一致性方面的有效性, 将所提出方法和 SID 方法进行了对比实验, 从同一个增强视频中抽取相同的 6 帧进行比较. 如图 5 所示, 其中图 5(a) 为本方法生成的增强视频帧, 图 5(b) 为 SID 方法生成的增强视频帧. 可以看出, 本文提出的方法生成的视频具有更清晰的纹理细节. 此外, 各帧之间的亮度也更加一致, 这表明本文使用的孪生网络结构和时域一致性损失在保持视频时域一致性方面发挥了重要作用.



(a) 本方法生成的视频帧序列



(b) SID 方法生成的视频帧序列

图 5 视频时域一致性评估

2.2.3 静态视频增强定量评估实验

为了更直观地展示局部与全局相融合的孪生低照度视频增强网络在增强静态视频方面的有效性, 该实验将本方法与其他低照度视频增强方法进行了定量比较. 从不同方法输出的增强视频中随机选取 5 个连续帧, 计算平均 SSIM 和 PSNR. 结果如表 1 所示, 本方法在 PSNR 和 SSIM 方面明显优于传统方法 VBM4D+Rawpy[25]和 KPN+Rawpy[26]. 此外, 与基于深度学习的方法 MBLLVEN[16]、SID[27]、SMOID[11]、StableLLVE[14]、SMID[10]、SDSDNet[17]和 FastLLVE[15]相比, 在 PSNR 方面, 本方法比排名第 2 的 FastLLVE 方法高了 0.11 dB, 在 SSIM 评价指标中, 本方法比第 2 名的 SDSDNet 方法高了 0.025.

2.2.4 视频时域一致性定量评估实验

在视频时域一致性效果的对比实验中, 本文采用平均亮度方差 (AB(var)) 和平均绝对亮度差 (MABD)

两个指标验证提出方法的有效性, 结果如表 2 所示. 实验表明本方法在 AB(var) 评价指标上取得了最好的结果, 比排名第 2 的方法 FastLLVE 低了 0.07, 在 MABD 这一指标中比效果最好的方法 SMID 高了 0.08, 但是比其他方法的效果都要好. 因此, 可以证明本方法中提出的孪生网络结构和时域一致性损失能有效提保证视频的时域一致性.

表 1 静态视频增强定量评估

方法	PSNR (dB)↑	SSIM↑
VBM4D+Rawpy	14.47	0.318
KPN+Rawpy	18.77	0.542
MBLLVEN	26.42	0.801
SID	27.74	0.804
SMOID	26.16	0.793
StableLLVE	27.37	0.795
SMID	28.11	0.816
SDSDNet	28.07	0.824
FastLLVE	28.14	0.821
本文算法	**28.25**	**0.849**

表 2 视频时域一致性定量评估

方法	AB(var)↓	MABD↓
SID	2.73	4.62
MBLLVEN	2.41	4.01
SMOID	1.56	3.59
StableLLVE	1.45	2.98
SMID	1.37	**2.17**
FastLLVE	1.31	2.34
SDSDNet	1.34	2.28
本文算法	**1.24**	2.25

2.3 消融实验分析

设计消融实验系统的验证提出的模块以及损失函数对低照度视频增强任务的有效性.

2.3.1 验证 LFEM、LSAM 和 FFM 的有效性

为了验证提出的 LFEM、LSAM 和 FFM 对模型性能的影响. 我们对模型进行了消融实验, 通过删除原模型中的 LFEM、LSAM 或 FFM, 以 PSNR 和 SSIM 作为评价指标, 比较与原模型的增强效果. 实验结果如表 3 所示, 可以看出原模型在 PSNR 和 SSIM 指标上都达到了最佳值, 去除 LFEM、LSAM 或 FFM 都会导致 PSNR 和 SSIM 显著下降. 实验结果表明, LFEM、LSAM 和 FFM 都会影响模型对低照度视频增强的效果. 这 3 个模块相辅相成, 通过融合局部和全局特征来提升低照度视频增强效果.

2.3.2 验证损失函数有效性

为了验证由重构损失、时域一致性损失和结构相似性损失组成的联合损失函数的有效性, 进行了如下实验, 实验结果如表 4 所示. 可以看出, 使用由重构损失、时域一致性损失和结构相似性损失组成的联合损失函数时在每个指标中都取得了最优值, 删除结构相似性损失后, 对评估视频质量的 PSNR 和 SSIM 两个指标产生较大影响, 分别降低了 1.01 dB 和 0.027; 继续删除时域一致性损失, 发现衡量视频时域一致性的指标 AB(var) 和 MABD 受到较大影响, 分别提高了 0.13 和 0.67. 实验结果证明本文提出的联合损失函数能够有效提升模型性能, 保证输出视频的时域一致性.

表 3 验证 LFEM、LSAM 和 FFM 的有效性

方法	PSNR (dB)↑	SSIM↑
去除局部特征融合模块	26.84	0.810
去除轻量级自注意力模块	24.98	0.764
去除特征融合模块	25.72	0.783
原模型	**28.25**	**0.849**

表 4 验证损失函数有效性

损失函数	PSNR (dB)	SSIM	AB(var)	MABD
L_r	27.01	0.817	1.42	3.28
$L_r + L_c$	27.24	0.822	1.29	2.61
$L_r + L_c + L_{\text{ssim}}$	**28.25**	**0.849**	**1.24**	**2.55**

3 结论与展望

本文提出了一种局部与全局相融合的孪生低照度视频增强网络模型. 该模型采用孪生网络结构来保证输出视频的时域一致性, 使用跨帧滑动窗口模块从静态视频中模拟具有时空信息的动态视频序列用于训练. 在单个局部与全局相融合网络中, 采用可变形卷积提取视频的局部特征信息, 再使用轻量级自注意力机制捕获视频帧不同区域之间的长程依赖关系, 并通过特征融合模块将提取到的特征信息进行融合, 以此来指导网络生成颜色更相似、细节更丰富的视频. 我们通过对比实验验证了本文提出方法的性能优于近几年提出的方法, 并通过消融实验证明了本文提出的联合损失函数、局部特征提取模块、轻量级自注意力模块和特征融合模块在本方法中的有效性.

参考文献

1 Li CY, Guo CL, Han LH, *et al*. Low-light image and video enhancement using deep learning: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(12): 9396–9416. [doi: 10.1109/TPAMI.2021.3126

387]
2 于晨曦. 浅析星光级摄像机的发展与应用. 中国新通信, 2020, 22(3): 46–47. [doi: 10.3969/j.issn.1673-4866.2020.03.037]
3 Jobson DJ, Rahman Z, Woodell GA. A multiscale retinex for bridging the gap between color images and the human observation of scenes. IEEE Transactions on Image Processing, 1997, 6(7): 965–976. [doi: 10.1109/83.597272]
4 Lore KG, Akintayo A, Sarkar S. LLNet: A deep autoencoder approach to natural low-light image enhancement. Pattern Recognition, 2017, 61: 650–662. [doi: 10.1016/j.patcog.2016.06.008]
5 Nakai K, Hoshi Y, Taguchi A. Color image contrast enhacement method based on differential intensity/saturation gray-levels histograms. Proceedings of the 2013 International Symposium on Intelligent Signal Processing and Communication Systems. Naha: IEEE, 2013. 445–449.
6 Arici T, Dikbas S, Altunbasak Y. A histogram modification framework and its application for image contrast enhancement. IEEE Transactions on Image Processing, 2009, 18(9): 1921–1935. [doi: 10.1109/TIP.2009.2021548]
7 Land EH. The retinex theory of color vision. Scientific American, 1977, 237(6): 108–128. [doi: 10.1038/scientificamerican1277-108]
8 Liu HJ, Sun XK, Hua H, *et al*. Low-light video image enhancement based on multiscale Retinex-like algorithm. Proceedings of the 2016 Chinese Control and Decision Conference (CCDC). Yinchuan: IEEE, 2016. 3712–3715.
9 Dong X, Wang G, Pang Y, *et al*. Fast efficient algorithm for enhancement of low lighting video. Proceedings of the 2011 IEEE International Conference on Multimedia and Expo. Barcelona: IEEE, 2011. 1–6.
10 Chen C, Chen QF, Do M, *et al*. Seeing motion in the dark. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019. 3184–3193.
11 Jiang HY, Zheng YQ. Learning to see moving objects in the dark. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2019. 7323–7332.
12 Lai WS, Huang JB, Wang O, *et al*. Learning blind video temporal consistency. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 179–195.
13 Ilg E, Mayer N, Saikia T, *et al*. FlowNet 2.0: Evolution of optical flow estimation with deep networks. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017. 1647–1655.
14 Zhang F, Li Y, You SD, *et al*. Learning temporal consistency for low light video enhancement from single images. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 4965–4974.
15 Li WH, Wu GY, Wang WY, *et al*. FastLLVE: Real-time low-light video enhancement with intensity-aware look-up table. Proceedings of the 31st ACM International Conference on Multimedia. Ottawa: ACM, 2023. 8134–8144.
16 Lv FF, Lu F, Wu JH, *et al*. MBLLEN: Low-light image/video enhancement using CNNs. Proceedings of the 2018 British Machine Vision Conference. Newcastle: BMVA Press, 2018. 220.
17 Wang RX, Xu XG, Fu CW, *et al*. Seeing dynamic scene in the dark: A high-quality video dataset with mechatronic alignment. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 9680–9689.
18 Zheng S, Gupta G. Semantic-guided zero-shot learning for low-light image/video enhancement. Proceedings of the 2022 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops. Waikoloa: IEEE, 2022. 581–590.
19 Triantafyllidou D, Moran S, McDonagh S, *et al*. Low light video enhancement using synthetic data produced with an intermediate domain mapping. Proceedings of the 16th European Conference on Computer Vision. Glasgow: Springer, 2020. 103–119.
20 Ye J, Qiu CZ, Zhang ZY. Spatio-temporal propagation and reconstruction for low-light video enhancement. Digital Signal Processing, 2023, 139: 104071. [doi: 10.1016/j.dsp.2023.104071]
21 Liu L, An JF, Liu JZ, *et al*. Low-light video enhancement with synthetic event guidance. Proceedings of the 37th AAAI Conference on Artificial Intelligence. Washington: AAAI, 2023. 1692–1700.
22 Liang J, Yang Y, Li B, *et al*. Coherent event guided low-light video enhancement. Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision. 2023. 10615–10625.
23 Cao H, Wang Y, Chen J, *et al*. Swin-Unet: Unet-like pure Transformer for medical image segmentation. Proceedings of

the 2022 European Conference on Computer Vision. 2022. 205–218.

24 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. Proceedings of the 3rd International Conference on Learning Representations. San Diego: ICLR, 2015. 1–14.

25 Maggioni M, Boracchi G, Foi A, *et al*. Video denoising, deblocking, and enhancement through separable 4-D nonlocal spatiotemporal transforms. IEEE Transactions on Image Processing, 2012, 21(9): 3952–3966. [doi: 10.1109/TIP.2012.2199324]

26 Mildenhall B, Barron JT, Chen JW, *et al*. Burst denoising with kernel prediction networks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 2502–2510.

27 Chen C, Chen QF, Xu J, *et al*. Learning to see in the dark. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 3291–3300.

(校对责编: 孙君艳)