

doi:10.3969/j.issn.1003-3106.2022.11.006

引用格式:徐鑫伟,兰太吉,薛旭成,等.基于局部运动约束的改进 PWStableNet 电子稳像网络[J].无线电工程,2022,52(11):1944-1952. [XU Xinwei, LAN Taiji, XUE Xucheng, et al. Improved PWStableNet Electronic Image Stabilization Network Based on Local Motion Constraints[J]. Radio Engineering, 2022, 52(11): 1944-1952.]

基于局部运动约束的改进 PWStableNet 电子稳像网络

徐鑫伟^{1,2}, 兰太吉^{1*}, 薛旭成^{1,2}, 吴梦飞^{1,2}

(1. 中国科学院长春光学精密机械与物理研究所, 吉林 长春 130033;

2. 中国科学院大学, 北京 100049)

摘要: 为了改善 PWStableNet 稳像神经网络输出的视频图像局部易发生失真的缺点, 提出了 2 种新的约束方式作为改进, 以损失函数的形式来抑制输出视频中的过量运动, 阻止局部失真的发生。主要是 2 种约束方案的设计: 一是对翘曲场进行频域分析, 鼓励低频信息抑制高频信息; 二是对生成的稳定帧进行分析, 计算局部运动的剧烈程度, 鼓励平稳运动抑制剧烈运动。经过实验数据对比, 在主观效果上, 改进后网络输出的帧间差分图更接近于真稳定视频, 过量的局部运动得到抑制; 在客观评价指标上, 改进后的网络多数性能指标高于改进前, 性能有效程度提升平均值为 7.73%, 说明改进后神经网络的综合质量优于原网络。

关键词: 电子稳像; 卷积神经网络; 像素翘曲; 频域分析; 约束局部运动

中图分类号: TP391.4

文献标志码: A

开放科学(资源服务)标识码(OSID):



文章编号: 1003-3106(2022)11-1944-09

Improved PWStableNet Electronic Image Stabilization Network Based on Local Motion Constraints

XU Xinwei^{1,2}, LAN Taiji^{1*}, XUE Xucheng^{1,2}, WU Mengfei^{1,2}

(1. Changchun Institute of Optics, Fine Mechanics and Physics, Changchun 130033, China;

2. Chinese Academy of Sciences, Beijing 100049, China)

Abstract: In order to improve the local distortion of the video image output by PWStableNet image stabilization neural network, two new constraint methods are proposed as improvements to suppress excessive motion in the output video by the form of loss function and prevent the occurrence of local distortion. The main work is the design of two constraint schemes. One scheme analyzes the warpage field in frequency domain to encourage low-frequency information and suppress high-frequency information; the other analyzes the generated stable frame, calculates the intensity of local motion, and also encourages stable motion and suppresses violent motion. Through the comparison of experimental data, in terms of subjective effect, the inter-frame difference diagram output by the improved network is closer to the true stable video, and the excessive local motion is suppressed; in terms of objective evaluation indicators, most performance indicators of the improved network are higher than those before the improvement, and the average improvement of performance effectiveness is 7.73%, which shows that the comprehensive quality of the improved neural network is better than that of the original network.

Keywords: electronic image stabilization; convolution neural network; pixel warpage; frequency domain analysis; constrained local motion

0 引言

视频稳定是业余和专业视频拍摄的共同需求。在业余方面, 特定的视频稳定设备通常是十分昂贵的, 而随着以手机为主的手持式摄像机的普及, 业余用户每天拍摄出来的视频占比越来越大, 不具有稳定器的摄像设备在拍摄过程中经常会受到高频抖动的干扰, 导致拍摄的视频带有比较明显的抖动, 视觉

质量大幅降低。将带有抖动的视频重新稳定下来是一项非常有意义的研究^[1]。

电子稳像技术以仅需要计算、无需额外设备就能达到稳像效果的优势成为了当今主流的稳像方法。一些传统电子稳像方法^[2-3]需要提取特征轨迹

收稿日期: 2022-05-05

基金项目: 国家自然科学基金(62005280)

Foundation Item: National Natural Science Foundation of China(62005280)

或使用光流法估计摄像机的运动,然后对其进行平滑以实现稳定,最后根据固定网格计算出一个全局单应矩阵或几个同形矩阵以将晃动的帧重新翘曲成稳定的帧。由于相机运动而产生的视差和深度的变化,使得这些传统方法难以产生稳定的视频。此外,这些方法通常依赖特征点的提取,鲁棒性较差,并在处理低质量视频(如夜景视频、模糊视频、有水印的视频和带有噪声的视频)时可能会稳像失败。

近年来,卷积神经网络(Convolutional Neural Networks, CNN)在计算机视觉任务上取得了巨大的成功,比如图像识别^[4]、目标检测^[5]和分割^[6]。特别是一些传统的视频处理问题已经可以用深度神经网络重新解决,如图像去噪去模糊^[7]、图像效果增强^[8]等。通过创建一个实际的数据集在构建好的神经网络上面进行训练, CNN 也可以实现视频稳定的效果。

Wang 等^[9]设计了一种用于多网格翘曲变换学习的网络,该网络能够获得与传统方法相当的性能,并且对低质量的视频更具鲁棒性。Xu 等^[10]设计了一种对抗性网络,用来估计一个仿射矩阵,通过仿射矩阵可以生成稳定的帧。然而,这些方法在固定网格的基础上估计一个或一组仿射矩阵,对于不同深度变化的场景不够灵活和准确。与传统的稳定方法相比,基于 CNN 的稳定方法具有更强的鲁棒性,因为它能够很好地提取高维抽象特征,而不是依赖于特征点提取和匹配,特征点提取和匹配的准确性很大程度上取决于视频的质量。在基于仿射变换或基于网格变换方面,它们与传统的方法非常相似,在强视差、深度变化以及有遮挡的情况下,这些方法是非常脆弱的,很容易把视频帧像素翘曲到错误的位置。

Zhao 等^[11]提出了一种新的名为 PWStableNet 的 CNN 来消除各种不稳定视频的抖动,其学习的对象是为每个不稳定帧像素提供理想位置的像素翘曲场,通过像素翘曲场,可以将抖动视频帧中的每个像素转换到稳定位置。与传统方法的每帧估计一个单应性变换或基于固定网格估计几个单应的变换方法

相比,像素翘曲场为每个像素提供翘曲方向。每帧计算 2 个像素翘曲场分别代表像素横向翘曲方向和纵向翘曲方向,可以为稳定视图中的每个像素提供理想的单独位置,一旦某个位置发生了计算错误,对整体的影响很小,所以对退化模糊的图像具有很强的鲁棒性并且可以很好地处理视差,也可以处理不同对象之间的不连续深度变化。该网络能够有效处理各种低质量、高深度的抖动视频,是当前鲁棒性最好的深度学习稳像方法,同时对比其他的深度学习方法在运算速度上有所提升。

PWStableNet 在处理有噪声、有模糊、带水印和大深度等低质量抖动视频有鲁棒性和通用性很强的优点,但该网络也存在缺乏有效局部运动抑制,易产生过度翘曲,造成局部失真的问题。本文针对此处进行改进创新,在训练过程中以损失函数的形式对生成的像素翘曲矩阵在空间上添加频域损失计算来抑制高频成分的出现,并利用 MeshFlow 网格流模型^[12]对局部区域运动剧烈程度进行评估,添加局部运动损失阻止过度运动。实验测试结果表明,改进后的 PWStableNet 能够有效抑制局部失真并且在峰值信噪比(Peak Signal to Noise Ratio, PSNR)、帧间结构相似度(Structural Similarity Index, SSIM)和稳定度(Stability)等客观指标上有所提升。

1 PWStableNet 的基本结构

PWStableNet 的基本结构如图 1 所示,网络结构的输入是给定的一组共 $2w + 1$ 张连续、不稳定的视频帧,主体部分是一个多级级联的编码解码卷积网络用来估计帧间运动的抖动,网络输出一个像素级的翘曲场,该翘曲场由 2 个翘曲矩阵组成,指示了稳定帧每个像素应该由哪个不稳定帧像素移动而来。通过这些翘曲场,可以将抖动帧中的像素转换到新的位置组成稳定帧。与以往基于固定网格估计单应矩阵或少量同形矩阵的方法相比, PWStableNet 提出的像素级翘曲场可以为稳定视图中的每个像素提供合适的单独位置,并能很好地处理不同物体之间的视差甚至不连续的深度变化。

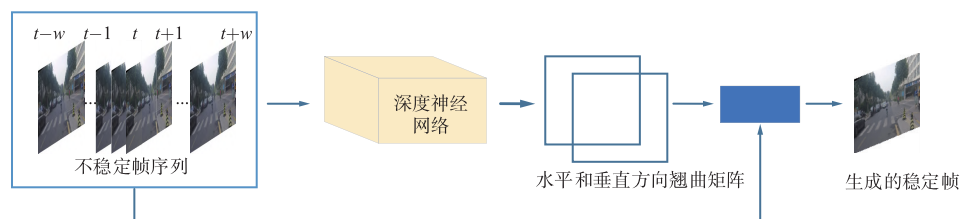


图 1 PWStableNet 的基本结构

Fig. 1 Basic structure of PWStableNet

1.1 系统输入

神经网络输入需要足够量的信息,要想将一帧图像稳定下来需要用到相邻视频帧的信息,因此网络以待稳定帧前后 w 帧作为输入。如图 1 所示,待稳定帧为第 T 帧,网络输入是 $t-w \sim t+w$ 帧,共 $2w+1$ 帧。输入帧图像半径 w 的大小决定信息量的大小,会对稳定效果和处理速度产生一定的影响。

1.2 像素翘曲场

像素翘曲场是与图片像素宽、高相同,通道数为 2 的一组矩阵,用来映射稳定帧和不稳定帧之间的关系。像素翘曲场的映射关系如图 2 所示。

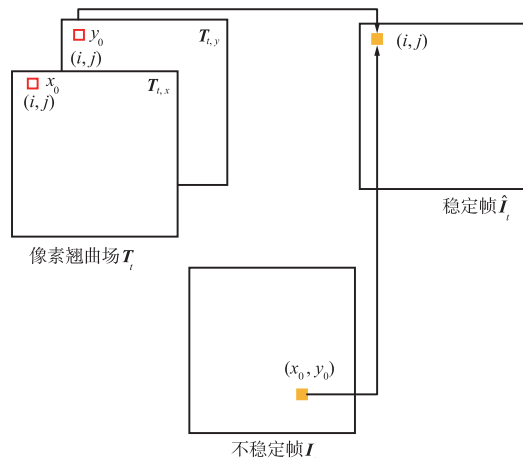


图 2 像素翘曲场的映射关系

Fig. 2 Mapping relationship of pixel warpage field

第 t 帧稳定前图像记为 I_t , 稳定后图像记为 $\hat{I}_t$, 像素翘曲场记为 T_t , 该帧稳定前的某个像素坐标编

号记为 (x_0, y_0) , 该像素稳定后坐标编号记为 (i, j) , 像素翘曲场包含的 2 个翘曲矩阵 $T_{t,x}$ 和 $T_{t,y}$ 对应像素 (x_0, y_0) 到 (i, j) 的关系:

$$\begin{cases} T_{t,x}(i, j) = x_0, \\ T_{t,y}(i, j) = y_0. \end{cases}$$

由上式可以看出, 2 个翘曲矩阵分别记录了源像素的水平坐标和垂直坐标, 稳定帧 $\hat{I}_t$ 的像素取自于不稳定帧 I_t , 例如 $\hat{I}_t(i, j)$ 的像素值取自 $I_t(x_0, y_0)$ 。 (x_0, y_0) 作为坐标值应该与 (i, j) 一样同为整数, 但是翘曲场 $T_{t,x}(i, j)$ 和 $T_{t,y}(i, j)$ 中存储的数据通常是浮点类型, $\hat{I}_t(i, j)$ 的像素值由 (x_0, y_0) 附近 4 个最邻近像素双线性插值取得。

1.3 神经网络主体结构

神经网络主体结构如图 3 所示, 可以看到有三级级联的网络结构, 每级结构相似, 主要是编码器-解码器框架, 类似于 U-Net^[13] 的结构。以单一级结构来看, 其编码器由若干个抽象层组成, 这些抽象层能够提取大大小小的特征信息, 在较深的层上有较多的通道。它的解码器由与编码器对应的信道编号相同、大小相同的反卷积层组成, 并通过跳跃连接与对应的反卷积层相连接。以级间结构来看, 每级结构都与相邻的级自上而下保持连接, 意义在于当前级的各层能够学习其前一级的剩余特征, 在更细节位置得到更精确的估计。每级结构都输出一个翘曲场 T^m (m 表示级序数), 这些学习了更多图像细节的翘曲场相叠加得到该帧最终输出的翘曲场 T 。

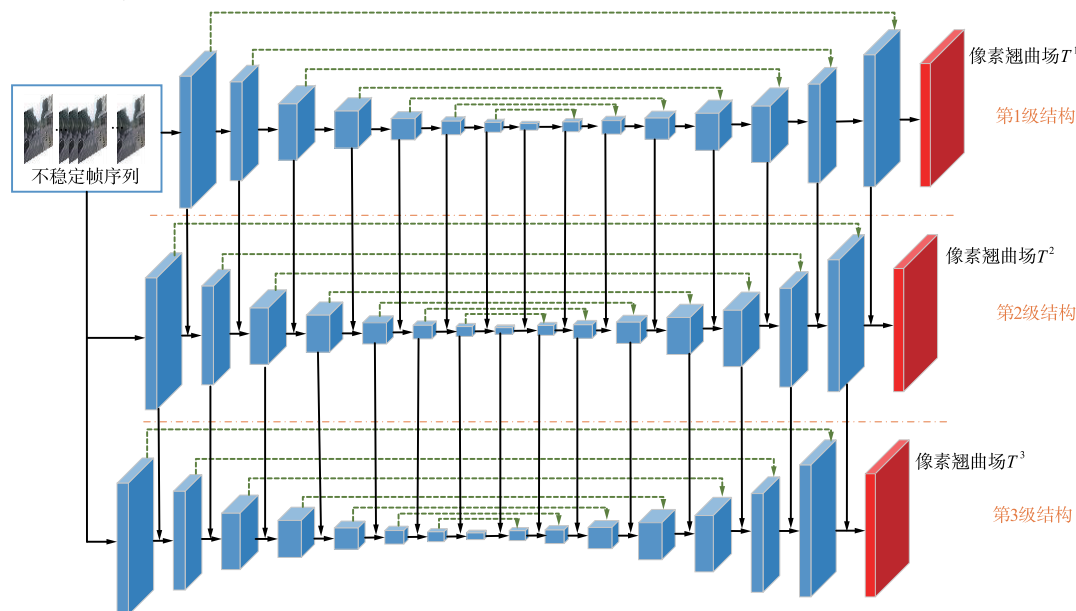


图 3 神经网络主体结构

Fig. 3 Main structure of neural network

2 改进与创新

在性能测试时发现,原生的 PWStableNet 网络模型生成的视频整体上稳定,但是偶尔在局部会产生不稳定区域,出现不正常的扭曲失真,经过分析并查阅文献后发现了导致该问题的一部分原因。像素级翘曲矩阵为每个像素安排位置,拥有极高的自由度,但是原网络对像素自由运动只做了简单的帧间均方差进行约束,缺乏更加直接有效的约束来防止过量运动。其后果就是输出的翘曲场很容易发生个别像素点的误匹配,把像素移动到过远或过近的位置,增加了输出图像的噪声,一定区域内误匹配数量过多则会局部图像失真,降低生成的稳定图像的质量。

2.1 对翘曲场增加频域约束

在查阅相关文献后,受到基于光流的视频稳定神经网络^[14]的启发,在网络训练过程中采用对生成的翘曲矩阵加强空间平滑度的方式,来抑制局部失真和减少噪声。加强空间平滑度的方法常见的有总变分约束和 Yu 等^[15]提出的线性翘曲场约束。然而,这 2 种方法存在着一定的局限性,总变分约束对局部噪声的抑制作用较强,而对畸变的抑制作用较弱;线性翘曲场约束难以控制,因为强约束限制了像素级翘曲场处理大规模非线性运动的灵活性,而弱约束不能很好地抑制局部形变^[14]。因此,加强翘曲场空间平滑度的方法应该满足以下 3 个条件:一是能保证像素级翘曲矩阵的处理抖动视频帧的灵活性;二是能抑制输出图像的局部形变;三是能减弱输出图像的整体和局部噪声。在频域利用傅里叶谱约束的方法能够同时满足以上条件,适合作为加强翘曲场空间平滑度的方法。

具体实现方法是在训练过程中对翘曲场中的每个翘曲场进行傅里叶变换得到其频谱,再对频谱进行加权运算得到新的频谱,新频谱的 L2 范数就作为此帧翘曲场频域损失函数的值,计算公式为:

$$L_{t,f} = \|\hat{G} \cdot \mathcal{F}W_t\|_2, \quad (1)$$

式中, t 表示帧序号; $L_{t,f}$ 为该帧频域损失; $\hat{G}$ 为加权滤波函数; $\mathcal{F}W_t$ 为翘曲矩阵经过傅里叶变换后的频谱。加权滤波函数 $\hat{G}$ 采用如图 4 所示的倒置高斯映射光谱:

$$\hat{G} = \frac{\max(G) - G}{\max(G)}, \quad (2)$$

式中, G 为 $\mu = 0, \sigma = 10$ 的二维高斯分布; $\hat{G}$ 由其最大值反转并归一化而来(如图 4 所示)。模型鼓励生成低频成分占比高的翘曲场,翘曲场低频成分占比越高,与倒置高斯函数加权相乘后的新频谱能量越低,计算后的损失函数数值也就越低。

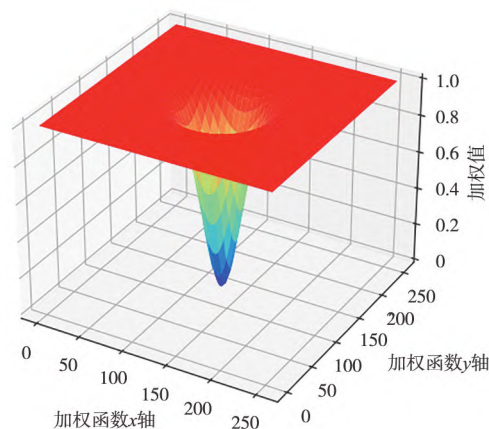


图 4 倒置高斯加权函数 $\hat{G}$

Fig. 4 Inverted Gaussian weighting function $\hat{G}$

2.2 对稳定帧增加局部运动约束

频域约束是通过对翘曲矩阵进行整体性约束来抑制局部失真的发生,属于间接方法,图像的局部失真程度可以采用更为直接的方式进行评估。Liu 提出的 MeshFlow 模型^[12]和 SteadyFlow 模型^[16]都能很好地检测局部区域内的帧间移动。MeshFlow 模型基于 KLT 稀疏光流特征匹配^[17],SteadyFlow 模型基于稠密光流,二者效果上相似,但是 MeshFlow 计算复杂度远低于 SteadyFlow。为了提升计算效率加快运行速度,本文采用 MeshFlow 模型来计算局部区域的帧间运动,评估局部区域不稳定程度。

MeshFlow 运动模型如图 5 所示。在每帧图像上放置一个 8×8 规则单元的均匀网格,经过计算后每个网格顶点都会被赋予一个运动矢量,用来表示该网格顶点附近区域的局部运动情况。假设 $\hat{p}$ 和 p 是相邻帧之间匹配的特征对的坐标位置,其中 $\hat{p}$ 处于第 $t-1$ 帧, p 处于第 t 帧,在特征位置处的运动矢量 v_p 可以计算为: $v_p = p - \hat{p}$ (如图 5(a)中的虚线箭头所示),并认为特征点附近的网格顶点应该有一个类似的运动。因此定义一个以 p 为中心的区域(如图 5(b)中的虚线圆所示),并将 v_p 分配给区域内的网格顶点(如图 5(c)若干箭头所示)。

所有匹配的特征都会将它们的运动传播到附近

的网格顶点,每个网格顶点会接收多个来自不同特征点的运动矢量。因此需要筛选出其中一个来代表此顶点运动矢量。中值滤波器经常用于光流估计,并被视为高质量流估计的诀窍^[18]。这里借用了该研究的稀疏运动规则化的类似思想,采用中值滤波筛选出代表顶点区域的运动矢量。所有带有运动矢量的网格顶点组成了一个稀疏运动场,表示了帧间的运动情况。

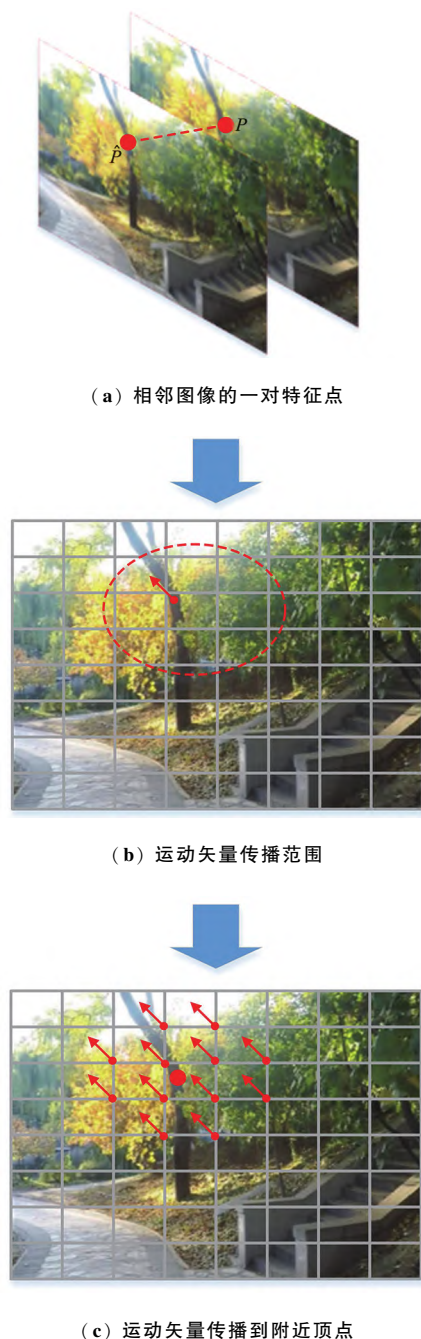


图5 MeshFlow 运动模型

Fig. 5 MeshFlow motion model

网络运动矢量分解如图6所示,每个网格顶点运动矢量一般可以分解为2个部分:一个是表示整个帧全局运动的全局矢量;另一个是表示局部细节自由运动的局部自由矢量。

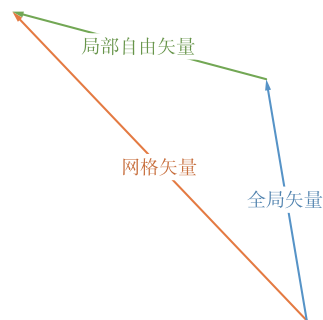


图6 网络运动矢量分解

Fig. 6 Mesh motion vector decomposition

目的是计算出局部区域的不稳定程度,全局矢量与局部区域不稳定度无关,应该采用代表局部自由运动的局部自由矢量来衡量局部运动剧烈程度。局部自由矢量可以由网格矢量与全局矢量相减得到,因此需要计算出全局矢量。前面提到 MeshFlow 模型的特征点由 KLT 光流特征匹配^[17]生成,可以借鉴传统方法利用这些特征点计算出全局仿射变换,把网格顶点坐标带入到仿射变换模型中得到新网格顶点的位置:

$$\begin{bmatrix} X' \\ Y' \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & b_1 \\ m_{21} & m_{22} & b_2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}, \quad (3)$$

式中, $[X', Y']$ 表示仿射变换后的网格顶点坐标; $[X, Y]$ 表示仿射变换前网格顶点的坐标; $m_{11}, m_{12}, m_{21}, m_{22}, b_1, b_2$ 均为全局仿射变换的参数。全局矢量则定义为:

$$\mathbf{M}_{t,d} = \begin{bmatrix} X' \\ Y' \end{bmatrix} - \begin{bmatrix} X \\ Y \end{bmatrix}. \quad (4)$$

在训练过程中根据网络输出的翘曲场 T 生成一个稳定帧,这个稳定帧与上一个稳定帧组成相邻帧,构建出 MeshFlow 网络模型并计算所有网格顶点的局部自由矢量。用这些局部自由矢量组成的稀疏运动场的 L2 范数来衡量此帧局部区域运动的剧烈程度,一个视频所有帧局部损失相加作为该视频的损失函数:

$$L_{t,m} = \sum_{n=1}^N \|\mathbf{M}_{t,n,g} - \mathbf{M}_{t,d}\|_2, \quad (5)$$

式中, t 表示帧序号; $L_{t,m}$ 为该帧的局部运动损失; N

表示一帧网格顶点数量; n 表示网格顶点序数; $M_{t,n,g}$ 表示第 n 个根据特征点直接求出的网格顶点矢量; $M_{t,d}$ 为根据仿射变换求出的全局矢量。

3 结果分析

在公开数据集 DeepStab^[9] 上进行训练, 该数据集共有 61 对稳定、不稳定视频数据, 数据集中的稳定视频称为真稳定视频, 以 45 对视频作为训练集, 8 对视频作为验证集, 8 对视频作为测试集。网络模型学习率采用阶梯式下降的方法, 初始状态学习率为 0.01, 之后每经过 10 轮, 学习率下降 10 倍。本文所有实验数据结果均在同一硬件平台上运行, CPU 型号为 Intel i9-12900k, GPU 型号为 Nvidia RTX-3090, 在此平台上进行测试输出视频的平均帧率为 79 帧/秒, 能够达到拍摄视频实时处理的需求。

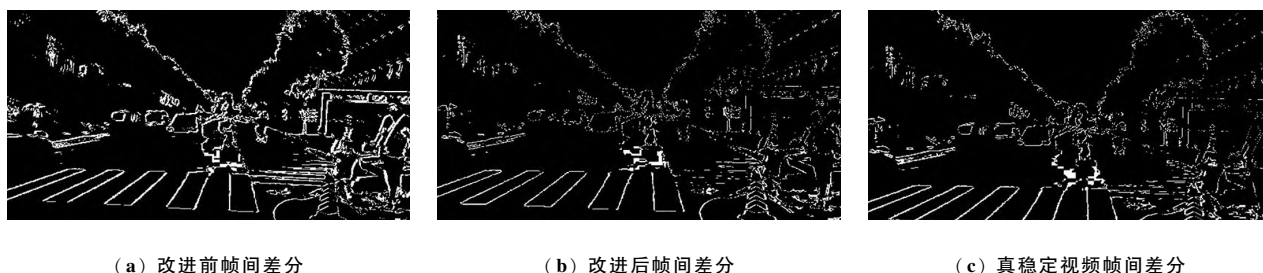


图 7 帧间差分对比

Fig. 7 Comparison of inter-frame difference

3.2 客观评价

一般电子稳像方法输出的帧边界会有一部分无信息区域, 处理这些区域的方法通常是直接裁剪, 为了方便比较, 把所有方法输出的视频都裁剪到一个公共区域。为了评估, 改进了网络的有效性, 除了损失函数组成不同, 网络结构以及其他超参数均保持一致。本文改进后的稳像网络除了与改进前网络进行对比, 还与原不稳定视频和基于光流特征匹配的传统方法^[2]进行对比, 共 4 组对照数据。测试对象为数据集 DeepStab^[9] 上的 8 个不稳定视频, 根据视频内容的特征分为“常规”“运动”“低照度”和“视差”4 种类型, 对应摄像机在日常使用过程中经常遇到的情况。为了定量比较提出的网络和以前的方法, 计算了 3 个客观指标: PSNR, SSIM 和稳定度, 下面简要地说明这 3 个指标。

① PSNR: 能够反映稳定后视频图像结构信息的完整程度, 其值越高, 表示受到噪声影响越小, 稳定后图像质量越高, 失真程度越小^[19]。

3.1 主观效果

本文的改进之处在于引入了更多的约束来改善 PWStableNet 易发生局部失真的缺点, 卷积网络结构未发生改变, 处理低质量视频的优秀能力得以保留, 同样具有很强的鲁棒性。

用肉眼难以直接察觉到稳定后视频微小的局部运动, 但是可以通过帧间差分图像来实现局部运动可视化。网络改进前、网络改进后和真稳定视频的帧间差分图如图 7 所示, 白色区域表示了相邻帧之间发生了运动的物体的基本轮廓, 白色区域的密集程度能够反映帧间运动的剧烈程度。由图 7 可以看出, 改进后白色区域的密集程度接近于真稳定视频, 明显低于改进前, 说明改进后的局部运动受到了抑制, 减少了局部失真发生的可能性, 主观效果优于改进前。本文改进后在局部区域运动上更符合稳定视频的要求。

② SSIM: 能够衡量相邻两帧之间图像相似程度, 一般认为稳定的视频相邻帧之间结构是相似的, 帧间结构相似度高说明视频越稳定。

③ 稳定度: 用于衡量视频的的稳定程度, 一般认为稳定的视频内的运动缓慢的低频成分占比较大, 稳定度计算了视频内的低频成分所占的比例。计算方法采用与文献[10]相同的方法, 将每个帧分割成 4×4 个网格, 记录每个网格顶点经过所有帧的历史路径, 并将每个顶点路径视为一维时域信号进行频域分析。每个顶点的稳定度计算为最低频率分量 (除去直流分量, 第 2~6 低的频率) 在全频率的占比。该视频的稳定的度的值为计算所有网格顶点稳定度的平均值。

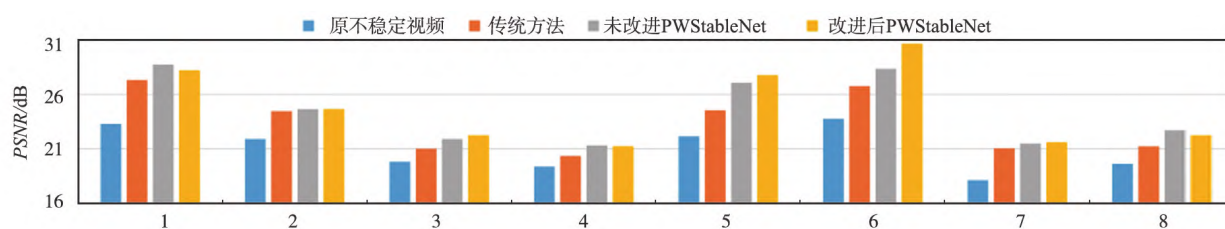
客观指标对比结果如图 8 所示, 第 1, 2 号视频对应“常规”类型; 第 3, 4 号视频对应“运动”类型; 第 5, 6 号视频对应“低照度”类型; 第 7, 8 号视频对应“视差”类型, 相同类型的视频指标数据相近。从 PSNR 指标上进行分析, 改进后的网络与改进前性

能持平,二者均高于传统方法;从 SSIM 指标上进行分析,本文方法性能略高于传统方法与改进前网络;从稳定度上进行分析,不同类型的视频上呈现出不同的结果。在“常规”类型的视频中,本文方法性能略高于传统方法和改进前网络;在“运动”类型的视频中,3 种方法性能接近;在“低照度”和“视差”类

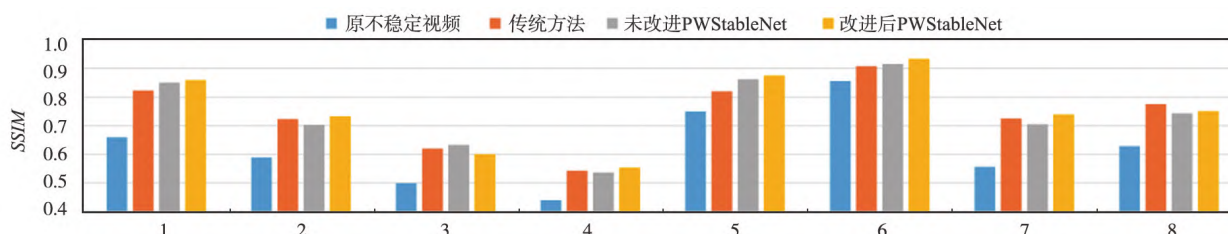
型中,传统方法数据指标高于 2 种基于神经网络的方法,其原因就在于传统方法处理这 2 种类型的视频时稳像失败,生成了含有大量空白区域的帧,空白区域网格顶点的运动矢量为零导致了稳定度异常提高。总体上看,本文改进后网络的客观评价指标更好。



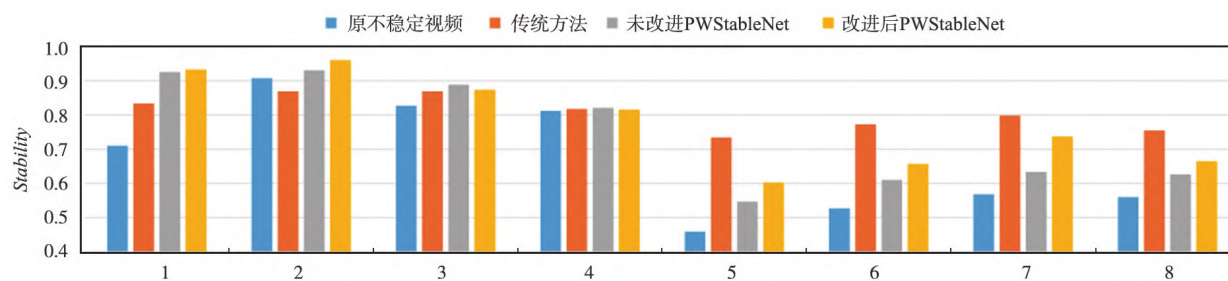
(a) 测试视频编号及类型



(b) 测试视频峰值信噪比 (PSNR) 对比



(c) 测试视频帧间结构相似度 (SSIM) 对比



(d) 测试视频稳定度 (Stability) 对比

图 8 性能指标对比

Fig. 8 Performance indices comparison

本文方法是基于 PWStableNet 改进而来,与改进前网络相对比进行针对性的分析是评判改进有效性的重要内容。数据内容为有效稳像占比,如表 1 所示,用于衡量稳像方法的有效程度,以不稳定视频指标到真稳定视频指标之差为尺度 1,以不稳定视频指标到稳像方法指标为有效稳像,即:

$$R = \frac{H_o - H_u}{H_s - H_u} \times 100\%, \quad (6)$$

式中, R 为有效稳像占比; H_o 为稳像方法的指标; H_u 为不稳定视频的指标; H_s 为真稳定视频的指标。

表 1 对比了网络改进前后输出的结果在不同视

频的 3 个指标上的具体数据。由表 1 中数据可以看出,提升的指标(加粗的数值)占多数约 71%,以平均值来看,3 种指标均有不同程度的提升,总的提升均值为 7.73%。这些客观评价指标说明了在引入 2 个新约束后的 PWStableNet 网络在综合性能上优于改进前。

表 1 改进前后有效稳像对比

Tab. 1 Comparison of effective image stabilization before and after improvement

编号	数据指标	改进前	改进后	提升比例
1(常规)	PSNR/dB	51.81	46.86	-4.94
	SSIM/%	64.07	66.95	2.88
	Stability/%	84.86	87.82	2.96
2(常规)	PSNR/dB	56.47	56.26	-0.21
	SSIM/%	73.88	93.64	19.75
	Stability/%	41.14	93.74	52.59
3(运动)	PSNR/dB	64.02	73.78	9.76
	SSIM/%	82.99	62.77	-20.22
	Stability/%	59.14	44.38	-14.76
4(运动)	PSNR/dB	88.34	84.30	-4.04
	SSIM/%	44.12	52.17	8.05
	Stability/%	9.38	3.99	-5.39
5(低照度)	PSNR/dB	62.30	71.00	8.70
	SSIM/%	66.23	73.79	7.56
	Stability/%	25.05	40.74	15.69
6(低照度)	PSNR/dB	38.71	58.19	19.48
	SSIM/%	55.40	71.88	16.48
	Stability/%	32.37	50.64%	18.27
7(视差)	PSNR/dB	54.18	55.79	1.61
	SSIM/%	55.45	68.21	12.76
	Stability/%	20.28	52.49	32.21
8(视差)	PSNR/dB	46.55	39.64	-6.91
	SSIM/%	52.75	55.79	3.03
	Stability/%	17.53	27.64	10.11
平均值	PSNR/dB	57.80	60.73	2.93
	SSIM/%	61.86	68.15	6.29
	Stability/%	36.22	50.18	13.96

4 结束语

本文对现有的视频稳像网络 PWStableNet 进行改进,解决改进前网络易发生局部失真的问题。对卷积网络引入了 2 个新的约束:像素翘曲场的频域约束和局部自由运动矢量的约束,来减少稳定视频在局部区域的多余运动,抑制局部失真,生成更稳定的视频。这 2 种约束的优点在于以直接和间接 2 种方式进行评估,更加细致、全面地计算出局部自由运动大小。经过实验对比验证,本文改进后的网络生成的视频在帧间差分具有更好的效果,同时在

PSNR,SSIM 和稳定度等客观性能指标上较其他算法更优。随后的研究重点内容将集中在优化网络结构、提升计算效率上,以获得更好的运算速度和稳像效果。

✦

参考文献

- [1] LIU S C, YUAN L, TAN P, et al. Bundled Camera Paths for Video Stabilization[J]. ACM Transactions on Graphics (TOG), 2013, 32(4): 1-10.
- [2] 刘广龙. 基于特征光流的电子稳像技术研究[D]. 哈尔滨: 哈尔滨工业大学, 2015.
- [3] 熊炜, 王传胜, 管来福, 等. 基于特征跟踪和网络路径运动的视频稳像算法[J]. 计算机工程与科学, 2020, 42(5): 843-850.
- [4] CHEN Y, WAN L, LI Z, et al. Neural Feature Search for RGB-infrared Person Re-identification[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE, 2021: 587-597.
- [5] CHEN X, XIE C, TAN M, et al. Robust and Accurate Object Detection via Adversarial Learning[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE, 2021: 16622-16631.
- [6] HE Y, YANG D, ROTH H, et al. DiNTS: Differentiable Neural Network Topology Search for 3D Medical Image Segmentation[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE, 2021: 5841-5850.
- [7] DONG H, PAN J, XIANG L, et al. Multi-scale Boosted Dehazing Network with Dense Feature Fusion[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020: 2157-2167.
- [8] XU K, YANG X, YIN B, et al. Learning to Restore Low-light Images via Decomposition-and-Enhancement [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020: 2281-2290.
- [9] WANG M, YANG G Y, LIN J K, et al. Deep Online Video Stabilization with Multi-grid Warping Transformation Learning[J]. IEEE Transactions on Image Processing, 2019, 28(5): 2283-2292.
- [10] XU S Z, HU J, WANG M, et al. Deep Video Stabilization Using Adversarial Networks[J]. Computer Graphics Forum, 2018, 37(7): 267-276.
- [11] ZHAO M, LING Q. PWStableNet: Learning Pixel-wise Warping Maps for Video Stabilization[J]. IEEE Transactions on Image Processing, 2020, 29: 3582-3595.
- [12] LIU S, TAN P, YUAN L, et al. Meshflow: Minimum Latency Online Video Stabilization [C]//European Conference on Computer Vision. Amsterdam: Springer, 2016: 800-815.

- [13] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional Networks for Biomedical Image Segmentation [C]//International Conference on Medical Image Computing and Computer-assisted Intervention. Munich: Springer, 2015: 234–241.
- [14] YU J, RAMAMOORTHY R. Learning Video Stabilization Using Optical Flow [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020: 8159–8167.
- [15] YU J, RAMAMOORTHY R. Robust Video Stabilization by Optimization in CNN Weight Space [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, 2019: 3800–3808.
- [16] LIU S, YUAN L, TAN P, et al. SteadyFlow: Spatially Smooth Optical Flow for Video Stabilization [C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 4209–4216.
- [17] TOMASI C, KANADE T. Shape and Motion from Image Streams Under Orthography: A Factorization Method [J]. Int J Comput. Vis., 1991, 9: 137–154.
- [18] SUN D, ROTH S, BLACK M J. Secrets of Optical Flow Estimation and Their Principles [C]//2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco: IEEE, 2010: 2432–2439.
- [19] 李天宇, 吴浩, 毛艳玲, 等. 一种改进 CycleGAN 的低照度图像增强网络 [J]. 无线电工程, 2022, 52 (2): 799–806.

作者简介



徐鑫伟 男, (1997—), 就读于中国科学院长春光学精密机械与物理研究所机械电子专业, 硕士研究生。主要研究方向: 光电成像与图像处理。



(* 通信作者) 兰太吉 男, (1988—), 博士, 副研究员。主要研究方向: 成像电子学。



薛旭成 男, (1980—), 博士, 研究员, 博士生导师。主要研究方向: 相机电子学、模数混合集成电路设计。



吴梦飞 男, (1992—), 硕士研究生。主要研究方向: 低照度图像处理。