

doi: 10.7690/bgzd.2024.05.016

一种基于 SiameseRPN 的模糊视频目标跟踪算法

赵 曜, 周 勇, 龚 俊, 李 锐

(中国兵器装备集团自动化研究所有限公司特种计算机事业部, 四川 绵阳 621000)

摘要: 为提高追踪准确率, 提出一种针对模糊视频的轻量级视频跟踪算法。采用主流对抗生成网络模型 DeblurGAN V2 进行去模糊处理, 交由基于改进 SiameseRPN 目标跟踪算法进行目标跟踪。为实现轻量化, 将目标跟踪的特征网络替换为 EfficientNet 网络, 改进注意力机制为 ECANet 以捕捉多通道信息, 并在 GoPro 数据集上进行测试。测试结果表明: 相比 SiameseRPN 算法, 该算法能够实现更高的追踪准确率, 帧率能达到实时性要求, 具有一定的借鉴意义。

关键词: 模糊视频; 目标跟踪; 去模糊; 算法轻量化

中图分类号: TP391.4 **文献标志码:** A

A Blur Video Target Tracking Algorithm Based on SiameseRPN

Zhao Yao, Zhou Yong, Gong Jun, Li Rui

(Department of Special Computer, Automation Research Institute Co., Ltd. of
China South Industries Group Co., Ltd. Mianyang 621000, China)

Abstract: In order to improve the tracking accuracy, a lightweight video tracking algorithm for fuzzy video is proposed. The mainstream countermeasure generation network model DeblurGAN V2 is used for deblurring, and the improved SiameseRPN target tracking algorithm is used for target tracking. In order to achieve lightweight, the feature network of target tracking is replaced by EfficientNet network, and the attention mechanism is improved to ECANet to capture multi-channel information, which is tested on GoPro data set. The test results show that compared with the Siamese RPN algorithm, the proposed algorithm can achieve higher tracking accuracy, and the frame rate can meet the real-time requirements, which has certain reference significance.

Keywords: blurred video; object tracking; deblurring; lightweight algorithm

0 引言

视觉目标跟踪 (visual object tracking, VOT) 作为计算机视觉领域中的关键研究方向, 已被广泛应用于安防^[1]、机器人控制、自动驾驶等领域。视觉跟踪算法的输入为一段视频帧序列, 通过对视频序列某一帧上的目标进行框选完成初始化后, 在后续视频中持续预测目标位置并用边界框标定目标, 实现对视频中单个或多个目标的持续追踪。在机器人领域中, 高速机动造成视频画面模糊及丢帧, 而现有的视频目标跟踪算法都基于清晰、流畅的航拍视频进行训练, 不能妥善应对高模糊视频。且平台搭载边缘计算平台算力有限, 不能进行高复杂度算法推理。笔者在已有目标跟踪算法基础上, 提出轻量级针对模糊视频的目标跟踪算法, 具有现实意义。

目标追踪一般分为相关滤波和孪生网络 2 类。最先出现的相关滤波算法是一种在线推理算法, 常见的有 CSK^[2]和 KCF^[3]等算法。这些算法通过计算图像的手工特征进行跟踪, 结构简单且能达到较高

追踪速度, 但准确率较低。基于孪生网络的目标跟踪算法参数较少、运行速度较快, 最早的 SiameseFC^[4]网络一经提出就展现出较快的运行速度和跟踪速度。后续在 SiameseFC 算法的基础上, 有学者提出 SiameseRPN^[5]算法, 将经典目标检测算法 Fast R-CNN^[6]的 RPN 思想融入其中, 通过孪生网络分别实现分类网络和回归网络, 能有效解决目标跟踪时物体严重形变问题。笔者基于结构简单的 SiameseRPN 算法, 提出针对模糊视频的目标跟踪算法, 以实现追踪准确率和追踪速度的平衡。为应对视频中的模糊现象, 笔者提出在跟踪算法运行之前, 引入去模糊算法先行对视频进行去模糊处理。

1 结合去模糊预处理的 SiameseRPN

1.1 模型原理

为实现模糊视频中的单目标追踪, 笔者在 SiameseRPN 网络中的前期处理中加入 DeblurGAN V2^[7]以进行视频的去模糊, 即先进行图像去模糊,

收稿日期: 2024-01-22; 修回日期: 2024-02-19

第一作者: 赵 曜(1998—), 男, 河南人, 硕士。

再进行目标追踪。笔者在用 DeblurGAN V2 网络中去模糊时，复用其中已提取的特征信息加入后续跟踪网络中，尽可能多地保留特征。DeblurGAN V2

在完成去模糊处理后经由反卷积网络层将图片大小重新恢复成 255×255×3 尺寸，然后送入 SiameseRPN 网络进行目标追踪。算法总体框架如图 1 所示。

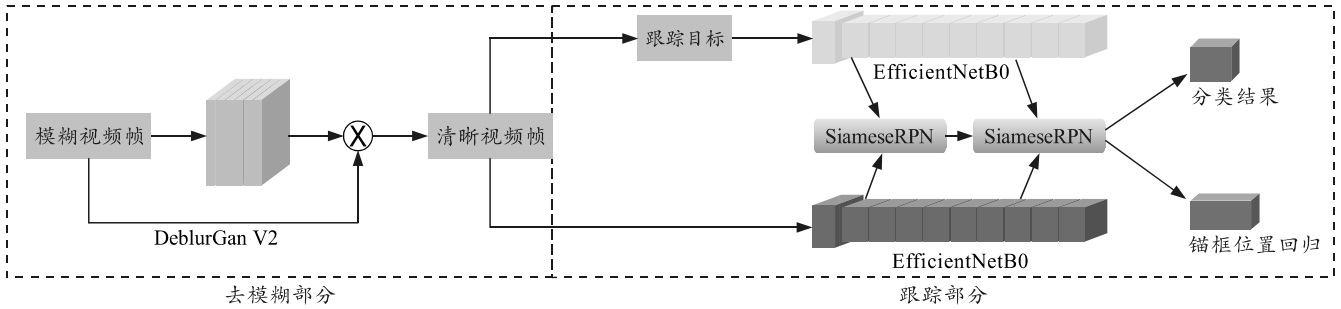


图 1 算法总体框架

在目标跟踪部分，相较于原来的 SiameseRPN 使用 AlexNet 前 5 层进行了特征提取，笔者换用最新 EfficientNetB0^[8]代替 AlexNet^[9]主干网络提取信息。相比 AlexNet，EfficientNet 有着简洁的结构，更高的识别率和更快的识别速度，在快速计算的同时能够尽可能多地保留图像的特征。

1.2 改进算法的注意力机制

笔者采用一种针对深度 CNN 的高效通道注意力模块 ECANet^[10]。该模块在避免降维的情况下能有效跨通道捕捉信息。

ECANet 结构主要分为 3 方面：1) 将特征图进行压缩，通过全局平均池化把特征图从(N, C, H, W)转换为(N, C, 1, 1)并获得聚合特征；2) 计算自适应卷积核的大小；3) 采用 Sigmoid 激活函数将权重映射在(0~1)之间。将 reshape 过后的权重值和原有的特征图做乘法运算，即可得到不同权重的特征图。

$$k = \psi(C) = \left\lfloor \log_2(C) / \gamma + b / \gamma \right\rfloor_{\text{odd}} \quad (1)$$

1.3 损失函数

损失函数，在分类分支中使用交叉熵损失，在回归分支中使用 smooth L₁ 损失。整体的损失函数为：

$$L_{\text{loss}} = L_{\text{cls}} + \lambda L_{\text{reg}} \quad (2)$$

式中 λ 为调节损失平衡的超参数。L_{cls} 和 L_{reg} 的表达式分别为：

$$L_{\text{cls}}(p_i, p_i^*) = -\log[p_i^* p_i + (1 - p_i^*)(1 - p_i)] \quad (3)$$

$$L_{\text{reg}} = \sum_{i=0}^3 L_{\text{smooth}L_1}(\delta[i], \sigma) \quad (4)$$

式中：p_i 为预测锚框为目标的概率；p_i^{*} 为目标的概率；δ[i] 为回归中输入的尺度信息的正则化表达；σ 为权重，可以调节损失函数在不同样本中的敏感程

度；L_{smoothL₁} 为 smoothL₁ 损失函数；i 为第 i 个网络输入。

2 试验数据集及平台

GoPro 数据集^[11]是业界常见的模糊图像数据集，笔者在 GoPro 数据集的基础上，通过手工标注目标的方法来进行标准锚框的选定，用以构建算法所需数据的训练集和测试集，进行后续的训练及测试。训练采用华为的服务器平台，平台为全国国产化，操作系统、CPU、训练卡和深度学习框架均为华为系列产品，组成部分如表 1 所示。

表 1 训练软硬件环境

环境组成部分	名称
操作系统	Openuler22.03
处理器	鲲鹏 910
NPU	Ascend 910
编程语言	Python 3.10
训练框架	Mindspore3.0

3 试验结果与分析

为了评估算法性能，利用重叠率和距离精度作为客观分析指标，重叠率表示算法预测目标位置与人工标注框之间的重叠面积与整体面积的比值。距离精度表示算法预测框的中心位置与人工标注框中心位置的欧氏距离，当该距离小于某一给定阈值时，该帧被视作成功跟踪，成功跟踪帧数与总帧数的比值被称作准确率。满足阈值条件的帧数比率越高，算法跟踪性能越好^[12]。

图 2 给出了各算法在欧氏距离阈值为 15pixel 时跟踪器的精确率、IOU 大于阈值 0.6 时跟踪器的成功率和基于模糊数据集的平均处理速度（即算法每秒可以处理的视频帧数量）。处理速度超过 30 帧/s 的算法具有实时性。

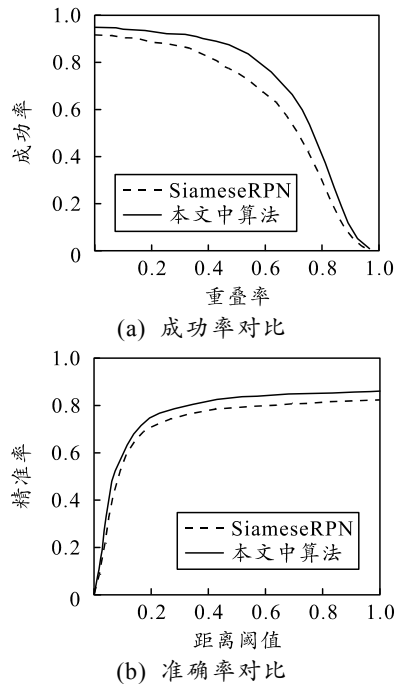


图 2 成功率和准确率对比

针对模糊视频，笔者提出的算法在准确率和成功率上均高于原先的 SiameseRPN 网络，帧率上有所降低，如表 2 所示。但算法处理速度仍然 > 30 帧/s，具有实时性。

表 2 算法效果对比

算法	准确率	成功率	帧率
SiameseRPN	0.783	0.731	75
本文中算法	0.821	0.815	42

4 结束语

笔者针对模糊视频中的目标跟踪问题，先利用对抗生成网络去模糊算法 DeblurGAN V2 进行去模糊处理，再在已有的目标跟踪算法基础上，更换神经网络特征提取网络，在保证轻量化的同时提高了特征提取效率。同时引入了新的注意力机制，使神经网络能充分利用通道间的信息，增强特征提取能力。测试结果表明：相比原先 SiameseRPN 算法，该算法目标追踪准确率更高，帧率较低但能达到实时性要求。

因为业界并没有针对模糊视频的目标跟踪提出标准的数据集，该算法能否较好应对被目标大规模形变以及外形相似目标跟踪仍需要验证。

参考文献：

- [1] 钟乐海, 韩正勇, 罗金生. 公共安全智能监控平台关键技术[J]. 兵工自动化, 2021, 40(11):60-6577.
- [2] MONTERO A, LANG J, LAGANIÈRE R. Exploiting the circulant structure of tracking-by-detection with kernels[C]//Proc. Int. Conf. Comput. Vis. Workshops. NY: IEEE, 2015: 587-594.
- [3] HENRIQUES J F, CASEIRO R, MARTINS P, et al. High-speed tracking with kernelized correlation filters[J]. IEEE transactions on pattern analysis and machine intelligence, NY: IEEE, 2014, 37(3): 583-596.
- [4] BERTINETTO L, VALMADRE J, HENRIQUES J F, et al. Fully-convolutional siamese networks for object tracking[C]//Computer Vision-ECCV 2016 Workshops Amsterdam, The Netherlands Springer. 2016: 850-865.
- [5] LI B, YAN J, WU W, et al. High Performance Visual Tracking with Siamese Region Proposal Network[C/OL]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 8971-8980. DOI:10.1109/CVPR.2018.00935.
- [6] GIRSHICK R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. Santiago, Chile: IEEE, 2015: 1440-1448.
- [7] KUPYN O, MARTYNIUK T, WU J, et al. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better[C]//Proceedings of the IEEE/CVF international conference on computer vision. Seoul, South Korea: IEEE, 2019: 8878-8887.
- [8] TAN M, LE Q V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks[C/OL]. Long Beach, California, USA: PMLR, 2019. <https://arxiv.org/abs/1905.11946>.
- [9] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Commun. ACM. New York: Association for Computing Machinery, 2017, 60(6): 84-90.
- [10] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. NJ: IEEE, 2020: 11534-11542.
- [11] NAH S, KIM T H, LEE K M. Deep Multi-Scale Convolutional Neural Network for Dynamic Scene Deblurring[C]//CVPR. Honolulu, HI, USA: IEEE, 2017.
- [12] 王殿伟, 许志杰, 覃泳睿. 一种基于改进 SiameseRPN 的全景视频目标跟踪算法[J]. 激光与光电子学进展, 2020, 57(24): 107-117.