

基于细粒度特征提取的数字媒体视频运动目标自适应跟踪

高海静

(陕西工业职业技术学院 信息工程学院, 陕西 咸阳 712000)

摘要: 为了有效跟踪数字媒体视频中的运动目标, 采取一种基于细粒度特征提取的自适应跟踪方法。这种方法涉及了特征的强化提取以及融合。经实验, 该算法在约26次迭代后达到最佳状态, 准确率为97.62%, 相较于粒子滤波算法和均值漂移算法分别领先11.33%和38.26%; 其最优鲁棒性表现为13.24%, 相较于两种算法减小了21.44%~41.73%。该算法还能在相似目标或存在遮挡的情况下, 保持对目标的有效跟踪。综上, 该基于细粒度特征提取的自适应跟踪方法提供了一种准确、鲁棒的数字媒体视频运动目标跟踪解决方案。

关键词: 目标追踪; 细粒度特征; 粗粒度特征; 特征融合; 多重卷积

中图分类号: TP391.41 **文献标识码:** A **文章编号:** 1003-7241(2025)08-0079-05

Adaptive Tracking of Moving Objects in Digital Media Video Based on Fine-grained Feature Extraction

GAO Haijing

(Dept. of Information Engineering, Shaanxi Polytechnic Institute, Xianyang 712000, China)

Abstract: In order to effectively track moving objects in digital media videos, an adaptive tracking method based on fine-grained feature extraction is proposed. This method involves enhanced feature extraction and fusion. The experimental results show that the algorithm reaches the best state after about 26 iterations, and the accuracy rate is 97.62%, which is 11.33% and 38.26% higher than that of particle filter and mean shift respectively. Its optimal robustness is 13.24%, which is reduced by 21.44%-41.73% compared with the two algorithms. The algorithm can also keep the effective tracking of the target in the case of similar target or occlusion. In summary, the adaptive tracking method based on fine-grained feature extraction provides an accurate and robust tracking solution for digital media video moving objects.

Keywords: target tracking; fine-grained characteristics; coarse-grained characteristics; feature fusion; multiple convolution

0 引言

在现代数字媒体环境中, 视频运动目标跟踪已成为一项关键技术, 对于智能监控、无人驾驶、人机交互以及虚拟现实等领域有着重要的应用价值^[1]。国外的Wei等提出了一种结合了滤波信道互补以及自适应网络的目标追踪算法, 克服了背景干扰^[2]。Jahan等提出了一种基于滤波器的状态协方差矩阵, 有效提高在非线性追踪当中的准确率表现^[3]。国内的吴悦等则针对当中的多目标追踪问题, 提出了一种基于端到端框架的多目标追踪模型, 提高了追踪精度和可靠性^[4]。段晓磊等则通过对目标行动轨迹的预测, 提高了追踪的准确率^[5]。可以看出, 由于照明变化、目标尺度和姿态的变化以及遮挡等多种复杂场景因素, 使得精确和稳定的目标跟踪在实际应用中面临巨大挑战^[6]。为了解决这些问题, 本研究提出了一种基于细粒度特征提取的数字媒体视频运动目标自适应跟踪

方法。通过精细地提取和处理目标特征, 该方法可以更准确地识别和跟踪目标, 同时可以更有效地适应各种复杂的场景变化。该方法的创新性主要体现在引入了一种双线程注意力模块, 以增强网络的特征提取能力, 使其可以利用更具有深度的特征信息。这项技术的贡献和重要性主要体现在以下几个方面。首先, 它对于提高数字媒体视频的目标跟踪精度和稳定性具有重要的实际意义。其次, 它为了解决复杂场景下的目标跟踪问题提供了一种新的解决方案, 有助于推动目标跟踪技术的进一步发展。最后, 它的成功实施可能会为智能监控、无人驾驶、人机交互以及虚拟现实等领域带来重要的应用价值。

1 基于细粒度特征提取的数字运动目标自适应追踪模型

1.1 基于细粒度特征强化提取网络的搭建

在数字媒体视频运动目标跟踪领域, 基于细粒度特征提取的自适应追踪模型显得尤为重要。该模型通过深

*基金项目: 全国高等职业院校信息技术课程教学改革研究项目(KT2024170)

收稿日期: 2023-12-17

度提取和处理目标特征,以提高跟踪的准确性和稳定性。然而,面对复杂场景的挑战,如何进一步提取和利用细粒度特征仍然是一个问题^[7]。为此,研究提出细粒度特征强化提取网络。基于现实中,图像不仅存在细粒度特征区别,同时存在粗粒度特征区别的实际情况,引入了粗粒度图像特征,通过对目标细粒度以及粗粒度特征的结合使用,旨在从丰富的视觉信息中获取更有效的特征表

示,以增强目标追踪模型的性能。这些方法的提出和实施,期望为数字媒体视频运动目标跟踪的精度和稳定性提供参考,以推动这一领域的技术发展。研究在细粒度识别的基础上,考虑到实际场景中不仅仅会存在诸如纹理等细粒度特征,还会存在轮廓、形状等粗粒度特征,提出了一种多粒度特征强化提取网络,其结构如图1所示。

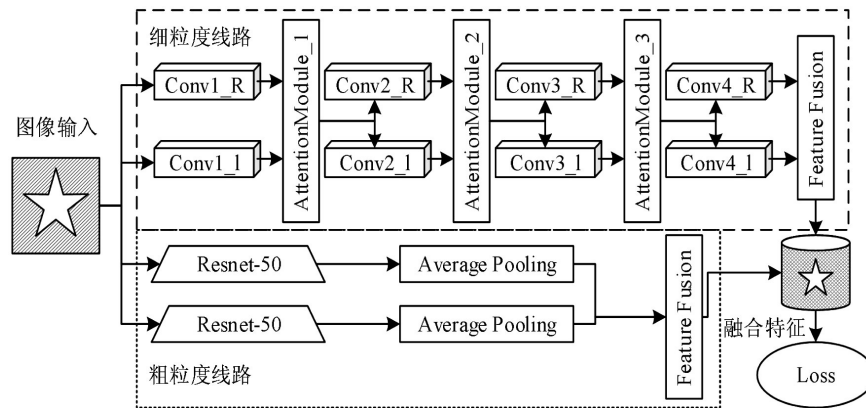


图1 多粒度特征强化提取网络

在图1所示的多粒度特征强化提取网络中,主要有一条双线程细粒度提取线路和一条双线程粗粒度提取线路。需要注意的一点在于,对于输入网络的图像数据,需要属于同一识别目标。当网络中输入一目标图像时,首先该图像进入细粒度特征提取模块,通过多条联合获取到具备目标重点学习的细粒度特征。然后该图像进入粗粒度特征提取模块,通过两条较为简单的粗粒度特征提取网络,获取到较为明显的全局适用特征。最后,局部细粒度特征与全局适用特征通过特征融合模块进行融合。研究中所采用的特征融合方式为Add结合Concatenate方法^[8]。Add是一种简单的特征融合方式,通常在特征维度相同时使用。它通过在相同位置的特征上进行相加操作,使得融合后的特征保持原始的特征维度不变。这种方式的优点是可以保持网络的计算复杂度不变,同时可以获得不同特征的共享信息。然而,它的缺点是可能会导致信息混淆,特别是当不同的特征在同一位置具有相似的数值时。Concatenate是通过将不同的特征在一个特定的维度上进行拼接。这样可以保留各自特征的独立性,避免了Add方式可能出现的信息混淆问题^[9]。然而,Concatenate方式的缺点是会增加网络的计算复杂度,因为特征维度会随着特征的拼接而增加。假设两条输入分别为 X_1, \dots, X_j 和 Y_1, \dots, Y_j ,其对应的卷积核为 K_1, \dots, K_j ,那么Add特征融合的输出则如式(1)所示。

$$T_{\text{add}} = \sum_{i=1}^j (X_i + Y_i) \cdot K_i = \sum_{i=1}^j X_i \cdot K_i + \sum_{i=1}^j Y_i \cdot K_i \quad (1)$$

则Concatenate特征融合方法的输出如式(2)所示。

$$T_{\text{con}} = \sum_{i=1}^j X_i \cdot K_i + \sum_{i=1}^j Y_i \cdot K_{i+j} \quad (2)$$

从式(1)与式(2)中可以较为明显地看出,Add融合方法更像是特征信息之间的叠加,而Concatenate则是一种对特征信息的拼接。Add特征融合方法更像是特征信息之间的叠加,可以将不同粒度的特征直接相加。这种方法能够保留细粒度和粗粒度特征的信息,避免了特征叠加产生干扰。而Concatenate特征融合方法则是一种对特征信息的拼接,将不同粒度的特征按照通道的方式进行拼接。这种方法能够有效增加特征的信息量,提高模型的训练性能。因此,研究采用联合方法进行特征融合,在一定程度上克服两种方法各自的缺点。一方面保证了细粒度以及粗粒度所提取出的特征都被保留,避免发生特征叠加产生干扰;另一方面有效增加了特征的信息量,从而提高模型的训练性能。在数字运动目标自适应追踪中,细粒度特征强化提取网络已经证明了其在处理复杂场景下的优势。然而,为了进一步提升跟踪精度和稳定性,需要更深入地探究并改进这一网络的运作机制。特别是在特征提取和分析的过程中,如何更有效地增强网络的学习能力是需要进一步研究的重要问题。通过深入研究和改进细粒度特征强化提取网络的运作机制,可以提高数字运动目标自适应追踪的精度和稳定性,为实际应用提供更好的解决方案。

1.2 面向数字运动目标自适应追踪的改进

细粒度图像分类是一项非常重要的计算机视觉识别

技术。这一研究目标的主要驱动力在于对大类别的粗粒度特征进行更细致的划分。为了在数字媒体视频运动目标跟踪领域尽可能地利用细粒度图像分类的优势,此处引入了一种多线程注意力框架用于提取图片中的细粒度共享特征^[10]。具体而言,该细粒度支路由多个注意力模块构成,这一模块主要由两个部分组成:多重卷积模块和注意力单元。多重卷积模块由两个不互通参数的残差单元构成,其之间存在互连关系,目的在于获取图片的多重特征信息,并将不同层次的特征作为输入送入中央的注意力单元。至于注意力单元,通过低层次特征对高层次特征的组合,以及高层次特征对低层次特征的指导,结合共享参数的训练策略,产生一张带有权重信息的共享注意力权重地图,从而有效地弥合两种异质模态之间的差距^[11]。如图2所示即为研究在细粒度特征提取线路中所引入的多线程注意力框架示意图。

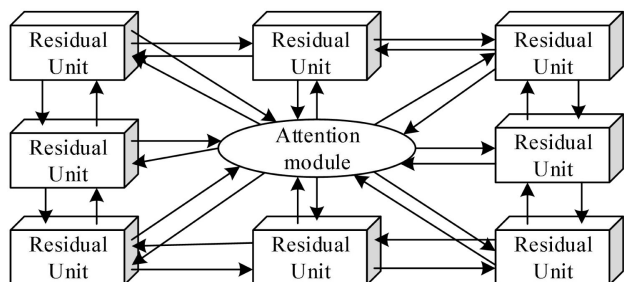


图2 多线程注意力框架示意图

如图2所示的多线程注意力框架中,研究所采用的Residual Unit模块的结构中,1级卷积层采用了Stage1和Stage2,2~4级卷积层则使用了Stage3~5,不同卷积层之间

的间隔为注意力单元,该结构设计参考了ResNet50的原始结构^[12]。这种结构设计的目的旨在通过增加卷积数,更深度地实现网络对于特征的学习。同时,在细粒度图像分类中,可能存在多种不同的模态,例如不同光照条件下的图像或者是不同感兴趣区域的图像。多线程注意力框架能够通过共享注意力机制,将不同模态下的图像信息进行融合,弥合不同模态之间的差异,提高分类性能。并且,由于在实际中,图像通常存在多样性和复杂性,例如姿态变化、遮挡、噪声等。多线程注意力框架能够在多个层次上对图像进行关注,从而提高模型对图像变化的鲁棒性,使得模型更加适应复杂的图像场景。为了进一步扩展该模块的功能,对其进行改进,使其具备双输入的能力。如图3所示,即为研究所引入的双输入注意力单元的示意图。

图3即为研究所引入的双输入注意力单元,其主要由一条同时具备上采样与下采样功能的线路与两条卷积线路组成。其中,卷积线路的主要目的在于对目标的细粒度特征以及粗粒度特征进行卷积处理。而中间的线路的主要目的在于提取细粒度特征与粗粒度特征的注意力交集。其采用另一种结合上采样与下采样的结构,首先通过下采样缩小图像特征,然后再通过上采样将图像特征变大,最终进入软掩模结构,通过多个卷积层以及池化层,最终提取出高层特征,并增强模型的感受野。该操作可以将高层特征中的激活像素投影到注意力模块中。进而加强模型对于特征的利用,实现更佳的动态目标追踪效果。

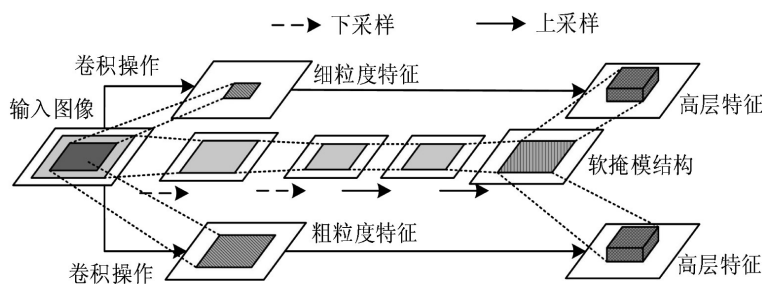


图3 双输入注意力单元框架示意图

2 基于细粒度特征提取的数字运动目标自适应追踪模型的测试

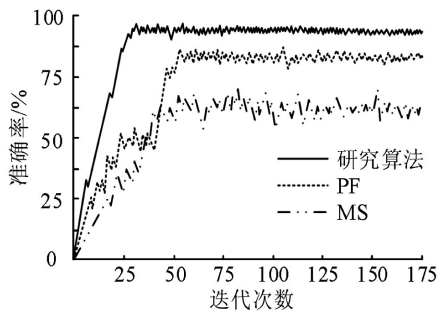
为了对研究所提出的基于细粒度特征提取的数字运动目标自适应追踪模型的性能进行评估,进行仿真实验,并引入粒子滤波(particle filter, PF)算法和均值漂移(mean shift, MS)算法与研究所提出的算法进行比对。研究选择采用2021年视觉目标跟踪(visual object tracking, VOT)数据集进行仿真实验,将其中随机的80%用于训练,20%用于测试。该数据集是一种用于评估和比较短期单目标

视觉跟踪算法的基准测试集。VOT项目由多个运动目标追踪挑战组成,这些挑战的目标是通过提供具有严格注释的数据集,推动计算机视觉研究领域的视觉跟踪算法的发展。VOT数据集中的每个视频序列都包含一些帧,每个帧中都有一个或多个注释的对象。注释通常是对象的边界框或更精细的标记,如对象的像素级分割。这些注释使得对视觉追踪算法的训练以及测试成为了可能。并且该数据集中还包含评估框架,可以计算各种性能指标,如准确性和鲁棒性。研究所采用的软硬件环境如表1所示。

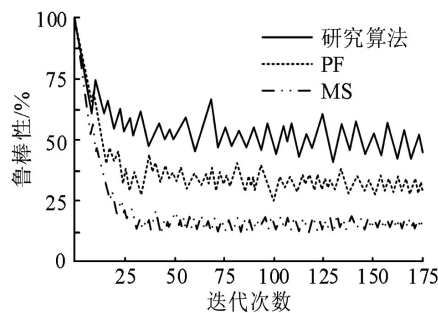
表1 软硬件环境

类别	名称	参数取值
硬件	Intel Core i7-9700K CPU	3.6 GHz, 8核8线程
	NVIDIA GeForce RTX 2060 GPU	6 GB, GDDR6
	Kingston DDR4 RAM	3 200 MHz, 16 GB (8 GB×2)
	Samsung 980 PRO SSD	NVMe M.2, 2 TB
软件	Windows 10 Pro	—
	Ubuntu 20.04 LTS	—
	Python	3.8
	NumPy 数据库	—
	Pandas 数据库	—
	Scipy 数据库	—
	TensorFlow 框架	2.6
	PyTorch 框架	1.7
	MATLAB	R2023a

首先对3种图像追踪模型的准确率以及鲁棒性进行测试,测试结果如图4所示。从图4中可以看出,研究所提出的模型在准确率以及鲁棒性方面具备更佳的表现,其训练效果也表现良好,能够以更少的训练次数达到最优状态。研究算法能够在约26次迭代达到最佳状态,同时其准确率为97.62%,相较于PF算法以及MS算法分别领先了11.33%和38.26%;其最优的鲁棒性表现为13.24%,相较于PF算法以及MS算法分别减小了21.44%和41.73%。较高的准确率以及较低的鲁棒性确保了研究算法在数字运动目标的追踪中能够保持较高的精度,并且能具有较低的错误率。



(a) 准确率

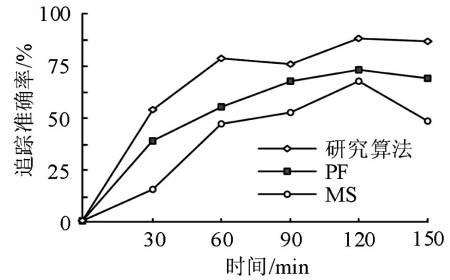


(b) 鲁棒性

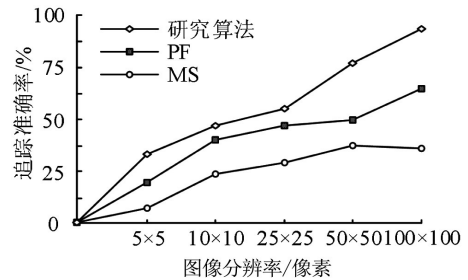
图4 3种算法的准确率以及鲁棒性测试

对3种图像追踪模型的细粒度图像追踪准确率进行测试,如图5所示即为细粒度图像追踪准确率测试图。从

图5(a)中可以看出,随着时间的增加,PF算法以及MS算法出现了一定程度的准确率下降问题,而研究所提出算法仍然能保持其较好的准确率,在长时间使用中具备更大的优势。从图5(b)中可以看出,研究所提算法由于其结合了细粒度以及粗粒度的关系,其在较小像素图像中能够保持较高的追踪识别准确率,相对于PF算法以及MS算法,在细粒度图像识别追踪的应用中具备更大的优势。



(a) 追踪准确率随时间变化



(b) 追踪准确率随图像像素变化

图5 3种算法的细粒度图像追踪准确率

为对研究所提出的基于细粒度特征提取的数字运动目标自适应追踪模型的实际使用效果进行验证,选择某视频监控片段,以图像中的车辆为追踪目标进行实际追踪效果验证。最后对3种图像追踪模型的实际图像追踪识别率进行测试,如图6所示为测试结果图,从图6中可以明显看出研究所提出模型具备更优的追踪性能。3种算法均能在识别目标不存在大部分遮挡的情况下完成对目标的识别。其中的PF算法在目标的特征被遮挡后,便失去了对目标的追踪;MS算法在出现相似目标后,由于对目标特征的提取深度不够,出现了追踪目标混淆的问题;而研究所提出算法能够在相似目标以及存在遮挡的情况下,保持对目标的追踪。

3 结束语

在数字媒体视频运动目标跟踪领域,本次研究提出了基于细粒度特征提取的自适应跟踪算法。其目标是提高跟踪的准确率,优化鲁棒性。为了验证其性能,研究中进行了迭代和比较实验。结果显示,该算法在约26次迭代后达到最佳状态,准确率为97.62%,优于PF算法和MS算法,分别领先11.33%和38.26%。同时,其鲁棒性最优

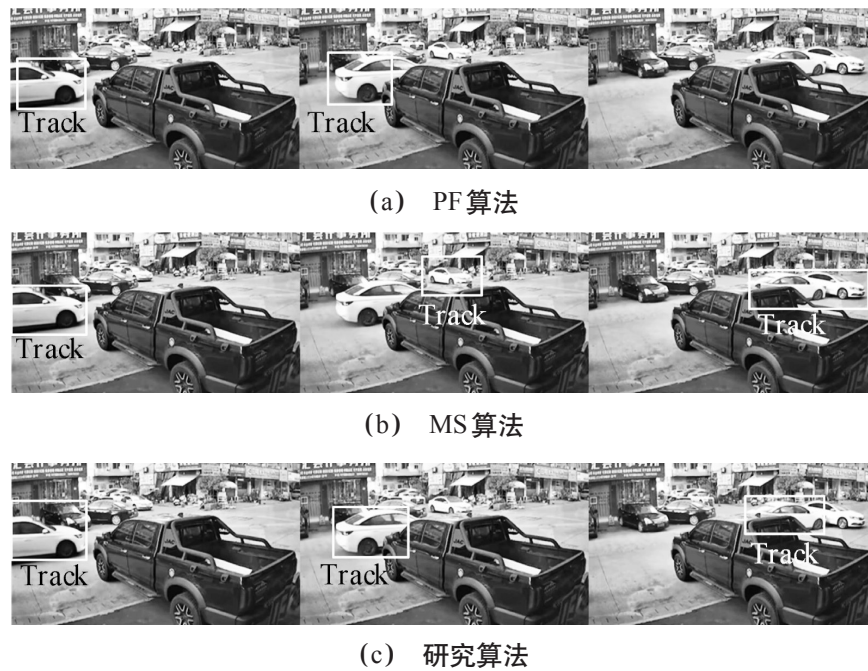


图6 实际图像追踪测试

表现为13.24%，分别比PF算法和MS算法减少了21.44%和41.73%。并且研究所提出算法在存在相似目标或遮挡的条件下，仍能够保持对目标的有效追踪。值得注意的是，该算法的追踪准确率并不受运行时间以及图像像素的影响，证明了其出色的稳定性和适应性。该研究的贡献在于提出并验证了一种基于细粒度特征提取的自适应跟踪算法，它可以在复杂的环境中，如存在相似目标或遮挡的情况下，保持对目标的有效追踪。然而，本研究也存在一些不足，其计算复杂度较高。另外，算法在处理大规模数据时的性能尚未得到充分验证。

参考文献：

[1] 周志杨, 刘光辉, 杨蕾, 等. 基于图像检测技术的室内人员动态服装热阻系统设计[J]. 计算机测量与控制, 2022, 30(1): 34-46.

[2] WEI B, CHEN H, DING Q, et al. SiamOAN: Siamese object-aware network for real-time target tracking[J]. Neurocomputing, 2022, 471(30): 161-174.

[3] JAHAN K, RAO S K. Measure of nonlinearity with application to bearings-only target tracking[J]. International Journal of E-Collaboration, 2021, 17(3): 46-61.

[4] 吴悦, 雒江涛, 张攀, 等. 基于Transformer和位置约束的端到端多目标追踪算法[J]. 重庆邮电大学学报(自然科学版), 2023, 35(3): 563-570.

[5] 段晓磊, 刘翔, 陈强, 等. 基于粒子滤波与LSTM网络对未标记AGV的追踪方法[J]. 传感器与微系统, 2020, 39(2): 37-43.

[6] 王祖武, 丁健, 魏文力, 等. 基于无人机双目图像的线目标测量的研究[J]. 计算机测量与控制, 2020, 28(1): 61-104.

[7] 李严. 基于多特征融合的健美操手臂动作轨迹识别[J]. 自动化技术与应用, 2023, 42(9): 90-93.

[8] 鲁家皓, 张捷, 胡国胜, 等. 基于多特征融合的图像匹配研究[J]. 微型电脑应用, 2022, 38(3): 1-8.

[9] 王丽亚, 刘昌辉, 蔡敦波, 等. 基于字符级联合网络特征融合的中文文本情感分析[J]. 微电子学与计算机, 2020, 37(1): 80-86.

[10] 杨彩霖. 自适应无监督聚类算法的运动图像关键帧跟踪[J]. 微型电脑应用, 2020, 36(12): 134-136.

[11] 赵建强, 朱万彤, 陈诚. 基于多重卷积神经网络模型的命名实体识别[J]. 计算机技术与发展, 2023, 33(1): 187-192.

[12] NAR A, YLDRM M, EROLU Y. Classification of pneumonia cell images using improved resnet50 model[J]. Traitement du Signal, 2021, 38(1): 165-173.

作者简介: 高海静(1980—), 女, 硕士, 副教授, 研究方向: 数字媒体技术, 程序开发。