

引用格式:于永辉,蔡佳航,刘斌,虞海江,杨文武.灯光剧烈变化环境自适应的二维人体目标检测[J].中国传媒大学学报(自然科学版), 2023,30(04):08-16.

文章编号:1673-4793(2023)04-0008-09

灯光剧烈变化环境自适应的二维人体目标检测

于永辉¹,蔡佳航¹,刘斌²,虞海江³,杨文武^{1*}

(1.浙江工商大学计算机科学与技术系,杭州 310018;2.南昌大学信息工程学院,南昌 330031;3.中科院软件所,北京 100190)

摘要:不同于人类视觉能够适应各种灯光变化环境,现有的二维人体目标检测算法在剧烈灯光变化场景中其检测性能会明显下降。针对这一问题,本文提出了一种灯光剧烈变化环境自适应的二维人体目标检测方法。首先,基于具有剧烈灯光变化的舞台演出环境,本文采集并构建了一个包含各种灯光颜色和丰富灯光变化的人体图片基准数据集(命名为“StageHuman”),以用于验证当前二维人体目标检测算法的缺陷与不足。其次,提出一种基于风格迁移的数据增强策略,将特定场景图片中的剧烈灯光变化风格迁移到大规模公开数据集COCO的人体图片中,再利用风格迁移后的大规模数据集来训练深度神经网络模型,从而提升模型在剧烈灯光变化环境下的二维人体检测性能。最后,通过大量的实验对比与分析,验证了本文方法能够有效提升深度神经网络模型在剧烈灯光变化环境下的鲁棒性和检测精度,并且该有效性不依赖于具体的风格迁移算法,而主要取决于所迁移的灯光变化风格的多样性和完整性。

关键词:人体目标检测;深度学习;风格迁移;数据增强

中图分类号:TP391 文献标识码:A

Adaptive 2D human object detection in environments with dramatic lighting changes

Yu Yonghui¹, Cai Jiahang¹, Liu Bin², Yu Haijiang³, Yang Wenwu^{1*}

(1. Department of Computer Science and Technology, Zhejiang Gongshang University, Hangzhou 310018, China; 2. School of Information Engineering, Nanchang University, Nanchang 330031, China; 3. Institute of Software Chinese Academy of Sciences, Beijing 100190, China)

Abstract: Unlike human vision, which can adapt to various lighting environments, the performance of existing 2D human object detection algorithms will be significantly reduced in the scene of drastic lighting changes. In order to solve this problem, in this paper a two-dimensional human object detection method was proposed, which adapted to the environment of drastic lighting changes. Firstly, based on the stage performance environment with drastic lighting changes, a human body image benchmark dataset (named “StageHuman”) containing various light colors and rich light changes was collected and constructed to verify the defects and deficiencies of the current two-dimensional human object detection algorithm. Secondly, a data enhancement strategy based on style transfer was proposed, which migrated the dramatic lighting change style in a specific scene image to the body picture of a large-scale open data

基金项目:浙江省自然科学基金(LY21F020010)

作者简介(*为通讯作者):于永辉(1999-),男,硕士研究生,主要从事计算机视觉技术研究。Email:2336788251@qq.com;杨文武(1981-),男,博士,教授,主要从事计算机动画、计算机视觉和人体运动捕捉技术研究。Email:wwyang@zjgsu.edu.cn

set COCO, and then used the large-scale data set after style transfer to train the deep neural network model, so as to improve the two-dimensional human detection performance of the model under the environment of dramatic lighting change. Finally, through a large number of experimental comparison and analysis, it is verified that the proposed method can effectively improve the robustness and detection accuracy of the deep neural network model under the environment of drastic lighting changes, and the effectiveness does not depend on the specific style transfer algorithm, but mainly depends on the diversity and integrity of the lighting change styles transferred.

Keywords: object detection; deep learning; style transfer; data augmentation

1 引言

二维人体目标检测(2D Human Object Detection)在诸多应用领域中扮演着关键角色,例如人体姿态识别和视频监控等^[1-2]。传统的二维人体目标检测算法主要基于手工设计特征,例如使用HOG、SURF等算法提取图像中的局部特征进行人体目标识别^[3-4]。传统方法的主要弊端是需要大量的专业知识和经验来设计特征,并且手工设计的特征其表示质量和泛化能力存在局限性,最终限制了目标检测算法在实际应用环境中的性能。与传统方法不同,基于深度学习技术

的二维人体目标检测方法自动从大量数据中学习特征表示,从而能够生成更加高效的目标检测特征^[5-6]。近年来,基于深度卷积神经网络模型的目标检测算法研究取得了突破性进展^[7-9],在富有挑战性的公开数据集上显著提升了二维人体目标检测性能^[10-12]。公开数据集中的图片主要对应着日常的室内外白炽光照环境,但在一些特殊环境中,例如现场的舞台演出环境,场景中的灯光五颜六色、忽明忽暗且变化剧烈,现有的二维人体目标检测算法在这些灯光剧烈变化场景中的检测性能和鲁棒性均会发生明显下降,如图1所示。

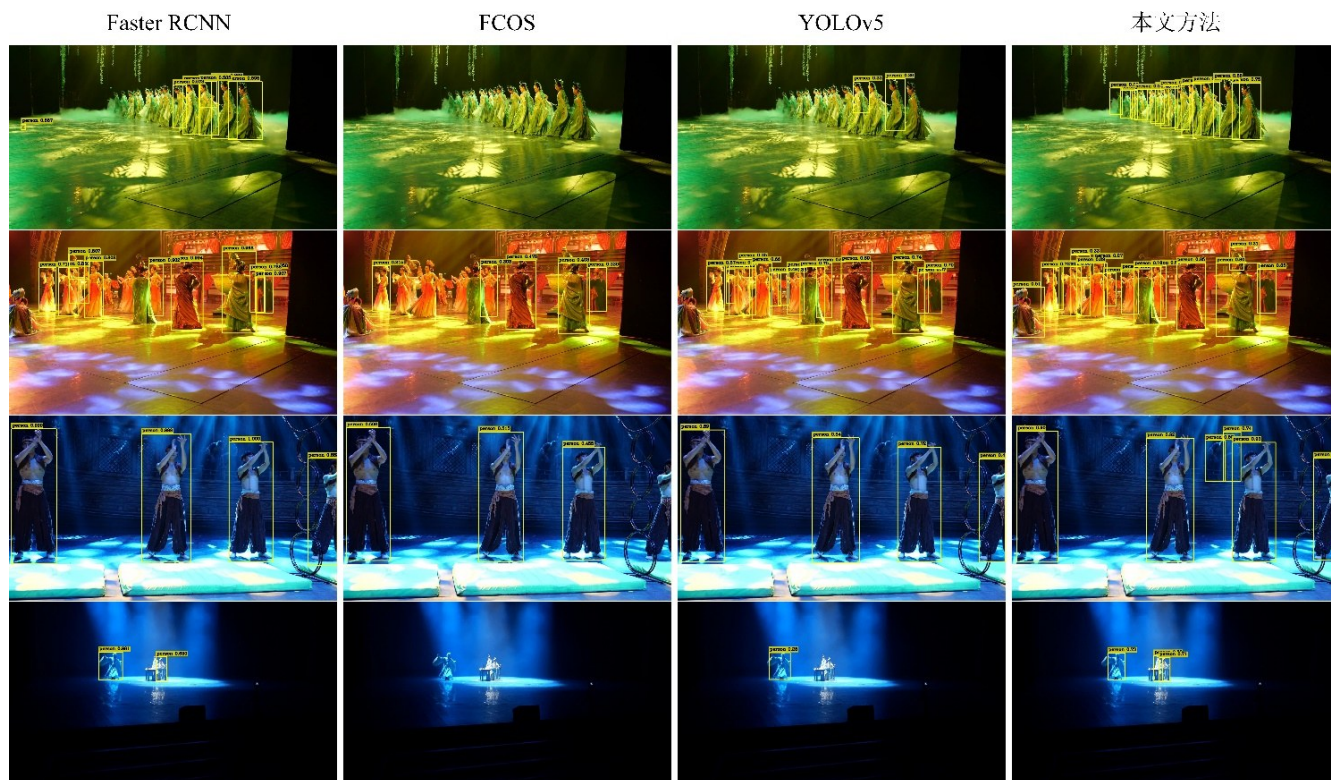


图1 经典算法(Faster RCNN^[7], FCOS^[8], YOLOv5^[9])在COCO公开数据集上训练得到的模型与本文方法(表1中的YOLO_F_MUL模型)在剧烈灯光变化环境下的二维人体目标检测效果

针对上述问题,本文提出了一种灯光剧烈变化环境自适应的二维人体目标检测方法。首先,为了验证当前二维人体目标检测算法在具有剧烈灯光变化环境中表现不足,同时为了设计和测试新算法,本文采集并构建了一个包含各种灯光颜色和剧烈灯光变化的人体图片基准数据集(命名为“StageHuman”)。StageHuman数据集中的图片来自于浙江横店影视城景区内的现场舞台演出环境,对应着多个风格迥异的舞台演出场景,场景中灯光复杂且变化剧烈,包含着各种颜色和明暗不均的舞台灯光,并且可编程电脑灯还会发生各种闪烁以及朝向变化。此外,舞台上的演员个数一直动态变化,少到仅1个人,多到二三十人,并且演员服装和姿态各异,此外演员之间以及演员与表演道具之间常常发生相互遮挡。因此,StageHuman数据集不仅具有剧烈灯光变化的特点,并且包含了丰富的人体与环境多样性。

基于StageHuman数据集,本文进一步提出一种基于风格迁移的数据增强策略,以提升深度神经网络模型在灯光剧烈变化环境下进行二维人体目标检测的鲁棒性和精度。该数据增强策略利用图片风格迁移算法,将StageHuman数据集图片中蕴含的剧烈灯光变化风格迁移到大规模公开数据集的图片中,例如COCO数据集,从而自动获得具有特定灯光变化风格的大规模标注数据。基于风格迁移后获得的具有特定灯光变化风格的大规模标注数据,本文训练深度神经网络模型使其学习到图片中蕴含的剧烈灯光变化风格,从而提升模型在剧烈灯光变化环境下的二维人体检测性能。以StageHuman数据集作为测试验证数据集,通过大量的实验对比与分析,验证了上述方法能够有效提升深度神经网络模型在剧烈灯光变化环境下的鲁棒性和检测精度,并且该有效性不依赖于具体的风格迁移算法,而主要取决于所迁移的灯光变化风格的多样性和完整性。

本文贡献可以归纳为以下三方面:

(1)采集并构建了一个名为StageHuman的二维人体目标检测基准数据集,该数据集中的图片蕴含了各种灯光颜色以及明暗不均的剧烈灯光变化。根据我们目前的调研结果,StageHuman数据集是国内外首个面向灯光剧烈变化环境的二维人体目标检测数据集。

(2)提出一种灯光剧烈变化环境自适应的二维人体目标检测方法,其核心是一个基于风格迁移的数据增强策略,它利用已有的大规模公开数据集,自动获

得具有特定灯光变化风格的大规模标注数据来训练深度神经网络模型,最终实现灯光剧烈变化环境自适应的二维人体目标检测。

(3)基于StageHuman数据集,通过大量实验,充分验证了本文所提出算法在剧烈灯光变化环境下的有效性和鲁棒性。

2 相关工作

2.1 目标检测算法

目标检测是计算机视觉领域中的一项基本任务,其目标是从图像或视频中识别并定位出多个物体的类别和位置。相较于图像分类任务只需输出物体的类别,目标检测则还需要准确地定位物体的空间位置信息。二维人体目标检测是目标检测的一个特定领域,与一般的目标检测相比,二维人体目标检测更加专注于人体目标的检测和定位。由于人体的形状、姿态和变化范围广泛,人体目标检测在许多应用中具有重要意义,如视频监控、人机交互、人脸识别、行为分析等。

Girshick等人^[6]提出了经典的R-CNN算法,它是一种基于锚框的两阶段检测算法。相比于传统的滑动窗口(Sliding Window)方法,R-CNN将目标检测任务转化为候选框的分类和回归问题,大大减少了计算量,并提高了检测性能。在R-CNN算法中,重叠框特征的冗余计算使得整个网络的检测效率较低。为了减少大量重叠框带来的冗余计算,K. He等人^[13]提出了一种空间金字塔池化层(Spatial Pyramid Pooling Layer, SPP),主要思路是对于一副图像分成若干尺度的图像块,然后对每一块提取的特征融合在一起,从而兼顾多个尺度的特征。进一步,R. Girshick等人^[7]提出了Fast RCNN,该网络使得我们可以在相同的网络配置下同时训练一个检测器和边框回归器,其改进版本Faster RCNN则是第一个端到端的深度学习检测算法,极大提升了检测框的生成速度。

两阶段目标检测虽然在目标检测领域取得了很好的成果,但是它的计算复杂度比较高,且需要多个阶段的计算。为了解决这些问题,研究人员开始探索单阶段目标检测算法。在最新的单阶段目标检测算法中,一些算法的性能已经超过了两阶段目标检测算法,同时计算复杂度也得到了很大的优化。YOLO^[9]作为一种较为流行的单阶段目标检测算法,其核心思想是将目标检测问题转化为一个回归问题。YOLO

结合了两阶段算法的优点,使用单个神经网络对检测到的对象执行边界框的分类和预测。此外,它针对检测器进行了大量优化,从而大幅提升了目标检测任务的检测速度,实现了实时的目标检测。最终,YOLO算法在准确性和速度上取得了很好的平衡。YOLO算法自提出以来一直在更新迭代,目前已经产生了从YOLOv1到YOLOv8的一系列目标检测算法。

随着大规模人体图片标注数据集的可公开获得,上述基于深度学习的二维人体目标检测方法均取得了明显的性能提升,已经成为该方向的主流技术。但是,不同于人类视觉能够适应各种灯光变化环境,现有的二维人体目标检测算法在剧烈灯光变化场景中其检测性能会明显下降,使得这些算法不能直接应用于具有剧烈灯光变化的复杂场景环境中。

2.2 风格迁移算法

风格迁移(Style Transfer)是指将一种图像的艺术风格应用到另一幅图像上,从而生成一幅新的图像,使得新图像既保留原图像中的主体内容,同时又带有另一幅图像的艺术风格。利用深度学习方法进行风格迁移是图像领域研究的热点之一。当前主流的基于深度学习的图像风格迁移方法可以划分为两类:基于图像迭代的图像风格迁移方法以及基于模型迭代的图像风格迁移方法。

Leon A. Gatys 等人^[14]首次提出一种基于图像迭代的图像风格迁移方法。论文作者发现了卷积神经网络(CNN)中内容和样式的表示是可分离的,因此可以独立地操纵这两种表示以产生新的、具有感知意义的图像。其主要思路是使用卷积神经网络进行特征提取,然后纹理合成,计算内容损失和风格损失,使用梯度下降优化总损失,然后不断迭代图像以得到艺术图像。然而,由于迭代次数过多、迭代时间较长,这种基于图像迭代的方法效率较低,随之出现了快速风格迁移方法。Justin Johnson 等人^[15]中提出了一种基于模型迭代的实时图像风格迁移方法。该方法使用前馈卷积神经网络对感知损失进行优化并生成图片,图像不是由随机噪声产生,而是通过输入图像的变换得到。因此,该方法被称为快速风格迁移方法,而图像风格转换网络也被称为快速风格化网络。

在最新的研究中,Domain^[16]提出了一种基于域感知的风格传输网络,不仅能够传输模版图片的样式,还能同时传递域的属性,同时针对不同的图像领域,自适应的调整风格转换参数,从而获得更好的结果。

FastPhoto^[17]网络主要由样式化以及平滑步骤两部分组成,样式化会将模版图片中的样式转移到待迁移图片中,平滑化则确保两者空间上的样式一致性。由于使用了封闭的函数来实现风格化,该方法迁移速度远高于传统算法。StyleNas^[18]是一种基于神经网络搜索架构的方法,它不需要进行后处理就能得到迁移后的图片,并且网络还使用了自动剪枝框架,能在保持风格化不变的情况下,提升风格迁移转换效率。

3 灯光剧烈变化环境自适应的二维人体目标检测方法

3.1 StageHuman:面向灯光变化剧烈舞台演出环境的二维人体目标检测数据集

为了验证并提升基于深度网络模型的二维人体目标检测算法在剧烈灯光变化场景中的性能,在真实的复杂演出环境中采集并构建了一个专门的二维人体目标检测基准数据集,该数据集中的图片蕴含了各种灯光颜色以及明暗不均的剧烈灯光变化。把该数据集命名为StageHuman,并且根据目前的调研结果,StageHuman数据集是国内外首个面向灯光剧烈变化环境的二维人体目标检测数据集。

StageHuman数据集的创建分为三个阶段:数据获取、数据预处理、数据集标注。在数据获取阶段,为了保证数据的多样性和真实性,选取了6个不同的真实舞台演出场景,各个场景都对应不同的舞台灯光变化情况,并且演出人物、服饰和动作都各有特色。在数据采集过程中,使用松下GH5S单反相机,以1080p的分辨率和50fps的拍摄帧率,从舞台前方的三个不同视角对正在进行的舞台表演进行拍摄,从而获取到覆盖了整个舞台表演的清晰视频图像。在数据预处理阶段,为了消除冗余数据,以20帧为步长对视频数据进行均匀采样,得到单张图像。在采样得到的所有图像中,以前70%作为训练集图像,后30%作为测试集图像。最后,在数据集标注阶段,采用开源标注软件Labelme,对训练集和测试集图像标注其中的二维人体检测框。采用交叉标注的方式,保证每张图像都至少经过两名标注人员的标注和检查。在两名标注人员标注情况下,整个标注过程由两名标注人员共耗时约1个月完成。最终,StageHuman数据集共包含30,346张单人检测数据,其中的21,242张为训练集,另外的9,104张为测试集。图2给出了数据集中的部分二维人体标注结果。



图2 StageHuman数据集中的部分二维人体标注结果

3.2 基于数据增强的复杂灯光环境下二维人体目标检测

为了提升现有的基于深度网络模型的二维人体目标检测算法在灯光剧烈变化环境中的检测性能,一种直接的方法是使用StageHuman数据集对已有的二维人体目标检测深度网络模型进行微调,以使其适应具有剧烈灯光变化的复杂环境。但是由于StageHuman数据集主要采集于舞台表演场景,并且数据集规模有限,容易造成深度网络模型在训练过程中过拟合,并且难以泛化到更多的场景和环境。针对这一问题,我们提出了一种基于风格迁移的数据增强策略,

将StageHuman数据集图片所对应场景中的剧烈灯光变化风格迁移到大规模公开数据集COCO的人体图片中,再利用风格迁移后的大规模数据集来训练深度神经网络模型,从而提升模型在剧烈灯光变化环境下的二维人体检测性能,并保持模型在各种场景和环境中的泛化能力。

为了实现StageHuman数据集场景中剧烈灯光变化风格的迁移,我们考虑了三种主流的风格迁移算法:Domain^[16]、FastPhoto^[17]、StyleNas^[18]。这三种风格迁移算法的迁移效果如图3所示,其中Reference指待迁移风格图片,也称为“迁移模板”。可以看出,这三



图3 三种不同风格迁移算法对剧烈灯光变化风格的迁移效果

种迁移算法的迁移效果各有不同,但是它们均较好地
把 StageHuman 数据集图片中所包含的灯光变化风格
迁移到了 COCO 数据集的人体图片中。在后面的实
验部分,我们将通过对比实验,来验证不同迁移算法
对最终的二维人体目标检测结果的影响。

在本文提出的基于风格迁移的数据增强策略中,
利用上述三种风格迁移算法,将 StageHuman 数据
集图片中所蕴含的灯光变化风格迁移到大规模公开
数据集 COCO 的人体图片中,从而实现 COCO 训练
数据集的增强。为了迁移 StageHuman 数据集图
片中所蕴含的灯光变化风格,需要从该数据集中选
取一组待迁移风格的图片,这些图片被称为“风格
迁移模板”。在实验中,根据所选用的风格迁移算
法以及风格迁移模块,本文制定了不同的数据增强
策略。为了验证不同迁移算法对二维人体目标检
测结果的影响,可以使用单个迁移算法、两个迁移
算法或三个迁移算法的不同组合来进行数据增强。
同时,为了验证风格迁移模板的数量对二维人体
目标检测结果的影响,可以使用单张迁移模板或多
张迁移模板来进行数据增强。

基于迁移算法和迁移模板数量的不同组合,最终可

以得到 11 种不同的数据增强策略,如表 1 所示。在
这些增强策略中,“单张模版”表示的是人工挑选
了 StageHuman 训练集中灯光颜色变化具有代
表性的某张图片,而“多张模版”表示的是人工
挑选了 StageHuman 训练集中灯光颜色变化具
有代表性的一组图片。在实验中,“多张模版”
包含了 10 张代表性图片。根据所选用的迁移算
法以及模版数量,为每种增强策略进行了命名。其
中,YOLO 表示将选用 YOLOv5 训练二维人体
目标检测的深度神经网络模型;D、F 和 S 分别
表示使用 Domain^[16],FastPhoto^[17]或 StyleNas^[18]
算法进行风格迁移;SGL 和 MUL 分别表示使用
了“单张模版”或“多张模版”。

在 YOLOv5 模型的训练过程中,对于每种数据
增强策略,本文采用如下方法进行数据增强:对于
COCO 数据集中的每个训练样本,首先根据该策
略确定迁移算法,如果该策略包含两个及以上的
迁移算法,随机选择其中的一个迁移算法;然后
根据该策略确定迁移模板,如果该策略包含多个
迁移模板,随机选择其中的一个作为迁移模板;最
后,根据确定的迁移算法和迁移模板,将该模板
中的风格迁移到上述的 COCO 的数据集中的训
练样本中。

表 1 迁移算法与迁移模板的不同组合所形成的数据增强策略

Domain	FastPhoto	StyleNas	单张模版	多张模版	命名
√			√		YOLO_D_SGL
	√		√		YOLO_F_SGL
		√	√		YOLO_S_SGL
√				√	YOLO_D_MUL
	√			√	YOLO_F_MUL
		√		√	YOLO_S_MUL
√	√		√		YOLO_D_F_SGL
√		√	√		YOLO_D_S_SGL
	√	√	√		YOLO_F_S_SGL
√	√	√	√		YOLO_D_F_S_SGL
√	√	√		√	YOLO_D_F_S_MUL

4 实验结果

根据上节介绍的数据增强策略,本文利用大规模
公开数据集 MS COCO 中的人体图片,来训练二
维人体检测深度神经网络模型 YOLOv5。根据不
同数据增强策略所训练得到的二维人体检测深
度神经网络模型,我们分别在 COCO 数据集的
验证集以及 StageHuman 数据集的测试集上对
其进行性能评估。我们分别使用了 AP、AR 和
F1 Score 三个指标进行了二维

人体检测性能的评估。

AP 指标:是一种用来衡量物体检测模型的精
度评价指标,反映了检测器对每个类别的识别准
确程度。AP 值越高,说明模型的检测性能越好。

AR 指标:是一种用来衡量物体检测模型在不
同召回率下的准确度评价指标,AR 指的是平均
召回率(Average Recall, AR),也就是在不
同的召回率阈值下,模型检测出的正确目标数
的平均值。AR 值越高,说明模型的检测性能越
好。

F1 Score 指标:是综合考虑 Precision 和 Recall 两个指标的评价指标,其计算方法是精确率与召回率的调和均值,即公式(1)所示:

$$F1\ Score = 2 * \frac{(Precision * Recall)}{(Precision + Recall)} \quad (1)$$

F1 Score 值越高,说明模型的分类性能越好。

表 2 给出了单一迁移算法和单张模板组合下的数据增强策略所训练得到的二维人体检测深度神经网络模型的性能结果。其中,YOLOv5s 表示在 COCO 数据集上训练而没有使用数据增强得到的深度模型。该实验主要用于验证不同迁移算法对最终检测结果的影响。由表可见,与未进行数据增强的 YOLOv5s 模型相

比,三种数据增强策略训练后得到的模型在 StageHuman 数据集上的 AP 和 AR 精度不仅没有提升反而都出现了下降,这表明了使用单张模板的策略未能提升模型在剧烈灯光变化环境下的二维人体检测性能。此外,我们注意到,三种数据增强策略训练后得到的模型在 COCO 数据集的 AP 精度仅出现了较小幅度的下降(-1% ~ -1.5%)以及 AR 指标的略微上升(+0.6% ~ +1.4%),这表明三种不同的风格迁移算法并未对 COCO 数据集中的图片本质内容产生较大影响,只是改变了数据集中图片的风格,并能够保持在 COCO 数据集所对应的各种不同环境中的泛化能力,进而验证了使用风格迁移算法进行数据增强策略的可行性。

表 2 单张模版+单一算法策略下进行数据增强的二维人体检测结果

方法	数据集	AP(%)	AR(%)	F1 Score
YOLOv5s	StageHuman	75.4	64	69.2
	MS COCO	80.9	64.8	72.0
YOLO_D_SGL	StageHuman	64	47.9	54.8
	MS COCO	79.9	65.9	72.2
YOLO_F_SGL	StageHuman	74.4	54.4	62.8
	MS COCO	79.9	66.2	72.4
YOLO_S_SGL	StageHuman	67.6	59	63.0
	MS COCO	79.4	65.4	71.7

表 3 给出了在单张模板下,不同迁移算法组合的数据增强策略所训练得到的二维人体检测深度神经网络模型的性能结果。其中,YOLOv5s 表示在 COCO 数据集上训练而没有使用数据增强得到的深度模型。该实验主要用于验证通过组合风格迁移算法是否可以提升灯光变化风格的迁移能力。由表可见,与未进行数据增强的 YOLOv5s 模型相比,四种数据增强策略训练后得

到的模型在 StageHuman 数据集上的 AP 和 AR 精度不仅没有提升反而都出现了下降。并且,从表 2 和表 3 可以看出,在相同的单张模板策略下,使用单一迁移算法与使用不同迁移算法的组合在最终的二维人体检测结果上性能较接近。这进一步验证了通过表 2 结果得出的结论:使用单张模板的策略不能提升模型在剧烈灯光变化环境下的二维人体检测性能。

表 3 单张模版+不同组合算法策略下进行数据增强的二维人体检测结果

方法	数据集	AP(%)	AR(%)	F1 Score
YOLOv5s	StageHuman	75.4	64	69.2
	MS COCO	80.9	64.8	72.0
YOLO_D_F_SGL	StageHuman	68.7	53.8	60.3
	MS COCO	80.9	65.9	72.6
YOLO_D_S_SGL	StageHuman	67.4	53.4	59.6
	MS COCO	80.1	65.4	72.0
YOLO_F_S_SGL	StageHuman	71.3	52.2	60.3
	MS COCO	79.4	66.5	72.4
YOLO_D_F_S_SGL	StageHuman	68.9	55.7	61.6
	MS COCO	80.8	66.1	72.7

表 4 给出了单一迁移算法和多张模板组合下的数据增强策略所训练得到的二维人体检测深度神经网络模型的性能结果。其中,YOLOv5s 表示在 COCO 数据集上训练而没有使用数据增强得到的深度模型。该实验主要用于验证模版图片样本多样性对最终检测结果的影响。由表可见,与未进行数据增强的 YOLOv5s 模型相比,三种数据增强策略训练后得到的模型在 StageHuman 数据集上的 AP 精度均有提升,提升幅度分别为 0.7%、4.6%、0.4%,以及在 AR 指标下精度也有着不同程度的提升。该结果表明,在使用多张模

版策略进行数据增强时,三种不同的风格迁移算法都能将 StageHuman 数据集图片中的灯光变化风格迁移到 COCO 数据集中,从而提升训练模型在剧烈灯光变化环境下的二维人体检测性能。其中,FastPhoto 风格迁移算法取得了 4.6% 的 AP 精度提升以及 6.1% 的 AR 精度提升,明显更优于 Domain 与 StyleNas 算法。综合上述分析,可以得出结论:1) 相比于单张模版策略,多张模版策略能够充分提取出复杂环境中的剧烈灯光变化风格;2) 不同的迁移算法具有不同的灯光变化风格迁移能力。

表 4 多张模版+单一算法策略下进行数据增强的二维人体检测结果

方法	数据集	AP(%)	AR(%)	F1 Score
YOLOv5s	StageHuman	75.4	64	69.2
	MS COCO	80.9	64.8	72.0
YOLO_D_MUL	StageHuman	76.1	62.3	68.5
	MS COCO	80.3	66.6	72.8
YOLO_F_MUL	StageHuman	80.0	70.1	74.7
	MS COCO	79.9	68.5	73.8
YOLO_S_MUL	StageHuman	75.8	64.8	69.9
	MS COCO	79	68.6	73.4

表 5 给出了三个迁移算法和多张模板组合下的数据增强策略所训练得到的二维人体检测深度神经网络模型的性能结果。其中,YOLOv5s 表示在 COCO 数据集上训练而没有使用数据增强得到的深度模型。该实验主要用于验证模版多样性与迁移算法多样性对最终检测结果的影响。由表可见,与未进行数据增强的 YOLOv5s 模型相比,该数据增强策略训练后得

到的模型在 StageHuman 数据集上的 AP 精度提升了 1.9% 以及 AR 精度提升了 5.5%。进一步对比表 3 和表 4,可以看出,在多张模板情况下,采用三个迁移算法的组合与单一迁移算法取得的检测精度提升相差较小。我们可以得出结论:相比于不同迁移算法对灯光变化风格迁移能力的影响,多张模版策略对灯光变化风格迁移能力具有更大的影响。

表 5 多张模版+混合三种算法策略下进行数据增强的二维人体检测结果

方法	数据集	AP(%)	AR(%)	F1 Score
YOLOv5s	StageHuman	75.4	64	69.2
	MS COCO	80.9	64.8	72.0
YOLO_D_F_S_MUL	StageHuman	77.3	69.5	73.2
	MS COCO	80.9	68.4	74.1

5 结论

本文提出了一种灯光剧烈变化环境自适应的二维人体目标检测方法。首先,为了验证并提升基于深度神经网络模型的二维人体目标检测算法在剧烈灯光变化场景中的性能,本文在真实的复杂演出环境中采集并构建了一个专门的二维人体目标检测基准数据集

StageHuman。然后,通过组合不同的风格迁移算法以及迁移模板数量,制定了多种的数据增强策略。最后,根据每种数据增强策略,将 StageHuman 数据集图片中的剧烈灯光变化风格迁移到大规模公开数据集 COCO 的人体图片中,再利用风格迁移后的大规模数据集训练得到二维人体检测深度神经网络模型。通过对每种数据增强策略下训练得到的模型进行对比

分析,发现本文提出的基于数据增强的复杂灯光环境下二维人体目标检测方法能够有效提升深度神经网络模型在剧烈灯光变化环境下的检测精度,并且该有效性不依赖于具体的风格迁移算法,而主要取决于所迁移的灯光变化风格的多样性和完整性。

参考文献(References):

- [1] Li D, Chen X, Zhang Z, et al. Pose guided deep model for pedestrian attribute recognition in surveillance scenarios [C]// 2018 IEEE International Conference on Multimedia and Expo (ICME), 2018: 1-6.
- [2] Luvizon D C, Picard D, Tabia H. 2D/3D pose estimation and action recognition using multitask deep learning [C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018:5137-5146.
- [3] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C]// 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005: 886-893.
- [4] Bay H, Ess A, Tuytelaars T, et al. Speeded-up robust features (SURF)[J].Computer Vision and Image Understanding, 2008, 110(3):346-359.
- [5] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multiBox detector [C]// European Conference on Computer Vision, 2016:21-37.
- [6] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]// 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [7] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149.
- [8] Tian Z, Shen C, Chen H, et al. FCOS: fully convolutional one-stage object detection [C]// 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 9626-9635.
- [9] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection [C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 779-788.
- [10] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: common objects in context [C]// European Conference on Computer Vision, 2014:740-755.
- [11] Everingham M, Eslami S M A, Van Gool L, et al. The pascal visual object classes challenge: a retrospective [J].International Journal of Computer Vision, 2015, 111: 98-136.
- [12] Deng J, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database [C]// 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009: 248-255.
- [13] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [14] Gatys L A, Ecker A S, Bethge M. A neural algorithm of artistic style [DB/OL]. arXiv:1508.06576, 2015.
- [15] Johnson J, Alahi A, Li F F. Perceptual losses for real-time style transfer and super-resolution [C]// European Conference on Computer Vision, 2016:694-711.
- [16] Hong K, Jeon S, Yang H, et al. Domain-aware universal style transfer [C]// 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 14589-14597.
- [17] An J, Xiong H, Huan J, et al. Ultrafast photorealistic style transfer via neural architecture search [DB/OL]. arXiv: 1912.02398, 2019.
- [18] Li Y, Liu M Y, Li X, et al. A closed-form solution to photorealistic image stylization [DB/OL]. arXiv :1802.06474, 2018.

编辑:赵志军,龙学锋