



基于改进的Faster RCNN面部表情检测算法

伍锡如, 凌星雨

引用本文:

伍锡如, 凌星雨. 基于改进的Faster RCNN面部表情检测算法[J]. 智能系统学报, 2021, 16(2): 210–217.

WU Xiru, LING Xingyu. Facial expression recognition based on improved Faster RCNN[J]. *CAAI Transactions on Intelligent Systems*, 2021, 16(2): 210–217.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.201910020>

您可能感兴趣的其他文章

生成对抗网络辅助学习的舰船目标精细识别

Fine-grained inshore ship recognition assisted by deep-learning generative adversarial networks
智能系统学报. 2020, 15(2): 296–301 <https://dx.doi.org/10.11992/tis.201901004>

注意力机制和Faster RCNN相结合的绝缘子识别

Insulator recognition based on attention mechanism and Faster RCNN
智能系统学报. 2020, 15(1): 92–98 <https://dx.doi.org/10.11992/tis.201907023>

基于跳跃连接金字塔模型的小目标检测

Skip feature pyramid network with a global receptive field for small object detection
智能系统学报. 2019, 14(6): 1144–1151 <https://dx.doi.org/10.11992/tis.201905041>

基于改进的Faster R-CNN高压线缆目标检测方法

Object detection of high-voltage cable based on improved Faster R-CNN
智能系统学报. 2019, 14(4): 627–634 <https://dx.doi.org/10.11992/tis.201905026>

多层卷积特征的真实场景下行人检测研究

Research on pedestrian detection based on multi-layer convolution feature in real scene
智能系统学报. 2019, 14(2): 306–315 <https://dx.doi.org/10.11992/tis.201710019>

多特征的光学遥感图像多目标识别算法

Research on multi-feature based multi-target recognition algorithm for optical remote sensing image
智能系统学报. 2016, 11(5): 655–662 <https://dx.doi.org/10.11992/tis.201511011>

微信公众平台



关注微信公众号, 获取更多资讯信息

DOI: 10.11992/tis.201910020

网络出版地址: <https://kns.cnki.net/kcms/detail/23.1538.TP.20200717.1718.006.html>

基于改进的 Faster RCNN 面部表情检测算法

伍锡如, 凌星雨

(桂林电子科技大学 电子工程与自动化学院, 广西 桂林 541004)

摘要: 针对真实环境下多目标表情分类识别算法准确率低的问题, 提出一种基于改进的快速区域卷积神经网络(Faster RCNN)面部表情检测算法。该算法利用二阶检测网络实现表情识别中的多目标识别与定位, 使用密集连接模块替代原始的特征提取模块, 该模块能够融合多层次特征信息, 增加网络深度并避免网络梯度消失。采用柔性非极大抑制(soft-NMS)改进候选框合并策略, 设计衰减函数替换传统非极大抑制(NMS)贪心算法, 避免相邻或重叠目标漏检, 提高网络在多目标情况下的检测准确率。通过构建真实环境下的表情数据集, 基于改进的 Faster RCNN 进行实验测试, 在不同场景中能够检测出目标的面部表情, 检测准确率相比原始检测模型提高 5%, 取得较好的检测精度。

关键词: 目标检测; 深度学习; 表情识别; 快速区域卷积神经网络; 特征提取; 分类识别; 多目标识别; 多目标定位
中图分类号: TP391.4 **文献标志码:** A **文章编号:** 1673-4785(2021)02-0210-08

中文引用格式: 伍锡如, 凌星雨. 基于改进的 Faster RCNN 面部表情检测算法[J]. 智能系统学报, 2021, 16(2): 210-217.

英文引用格式: WU Xiru, LING Xingyu. Facial expression recognition based on improved Faster RCNN[J]. CAAI transactions on intelligent systems, 2021, 16(2): 210-217.

Facial expression recognition based on improved Faster RCNN

WU Xiru, LING Xingyu

(College of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin 541004, China)

Abstract: To address the problem of the low accuracy rate of the multi-target facial expression classification and recognition algorithm in real environments, in this paper we propose a facial expression detection algorithm based on an improved faster region-based convolutional neural network (RCNN). The proposed algorithm uses a two-stage detection network to accomplish multi-target recognition and location in facial expression recognition. Instead of the original feature extraction module, densely connected convolutional networks are used, which can fuse multi-level feature information, increase network depth, and prevent network gradient disappearance. Soft non-maximum suppression (NMS) is used to improve the candidate-box merging strategy, and the attenuation function is designed to replace the traditional NMS greedy algorithm, thereby preventing the missed detection of adjacent or overlapping targets and improving the detection accuracy of the network under multi-target conditions. Through the construction of an expression data set in a real environment and an experiment based on the improved Faster RCNN, the facial expression of the target was detected in different scenes with a detection accuracy rate 5% higher than that of the original detection model. Therefore, good accuracy is achieved by the proposed algorithm.

Keywords: target detection; deep learning; expression recognition; Faster RCNN; feature extraction; classification and recognition; multi-target recognition; multi-target location

面部表情包含了丰富的信息, 直接反映了人

们的心理特征, 是表达情感的重要途径之一。面部表情检测可应用于人机交互、安防监视、医疗及认知科学等多个领域, 是计算机视觉研究热点之一。面部表情的表达通常分为愤怒、厌恶、恐惧、开心、悲伤和惊讶 6 类^[1-2]。面部表情检测任务的重点是从面部图像中提取面部表情特征, 并

收稿日期: 2019-10-07. 网络出版日期: 2020-07-17.

基金项目: 国家自然科学基金项目(61863007); 广西自然科学基金项目(2020GXNSFDA238029); 广西研究生教育创新计划项目(YCSW2020159); 桂林电子科技大学研究生教育创新计划项目(C20YJM00BX0M, 2021YCXSI22).

通信作者: 凌星雨. E-mail: lingxychina@163.com.

使用经过训练的分类器识别不同的面部表情。传统的表情识别依赖手工提取特征,特征提取方法主要分为3类:基于外观的特征提取、基于几何的特征提取和基于运动的特征提取。常用的外观特征提取包括像素强度^[3]、Gabor滤波^[4]、局部二值模式LBP(local binary patterns)^[5]及方向梯度直方图^[6],其中Gabor特征提取计算成本高昂,而LBP具有良好性能,被广泛用于面部表情识别^[7-8]。在基于几何的特征提取方法中,提取人眼、眉毛、嘴角等面部器官的位置和形状,形成能够代表人脸几何的特征向量^[9-10]。基于运动的特征提取方法提取动态图像序列为运动特征,根据特征部位的运动变化对面面部表情进行识别^[11-12]。由于光照变化、遮挡等多种因素,表情识别仍具有挑战性^[13],这些因素会影响识别精度,手工提取特征不适用于具有干扰的面部表情检测任务,深度学习的提出为这些问题提供了解决方案。

深度学习概念由Hinton在2006年提出^[14-15],比传统的网络具有更强的特征表达能力和泛化能力,近几年在面部表情检测中得到大量应用。如Yang等^[16]使用VGG16网络及DNN(deep neural networks)形成双通道对不同特征进行提取来完成表情识别。Wu等^[17]使用遗传算法优化神经网络来进行表情识别任务。Salmam等^[18]使用CNN(convolutional neural networks)提取外观特征,使用DNN提取几何特征点,合并为CNN-DNN模型进行表情识别任务。

在目标检测任务中,Girshick^[19]提出的R-CNN(Region-CNN)方法是目标检测中重要的参考方法,目标检测系列算法很多都借鉴了R-CNN的思路。R-CNN模型是将目标区域建议与CNN分类相结合,使用Selective Search算法在输入图像中提取2000个候选区域,经过CNN网络进行特征提取,通过训练好的分类器来判断候选区域中是否含有目标,再使用回归器对候选框进行调整。Girshick等^[20]结合SPP-net网络的思想对R-CNN进行了改进,提出Fast R-CNN模型。相比R-CNN,Fast R-CNN对整幅图仅进行一次特征提取,再与候选框映射,避免候选框重复提取特征而浪费时间。Fast RCNN采用Softmax分类与边框回归一起进行训练,省去特征存储,提高空间和时间利用率,同时分类和回归任务也可以共享卷积特征。Ren等^[21]提出用深度学习方法来进行区域建议即区域建议网络RPN(region proposal network),把RPN与Fast R-CNN结合,形成新的网络模型Faster RCNN,提高整体检测性能。

针对多目标复杂场景下的面部表情检测问

题,本文创新性地引入Faster RCNN网络对面面部表情进行识别及定位。根据表情检测特点在Faster RCNN网络框架基础上进行改进,使用密集连接网络代替原有特征提取模块,提取融合目标多层次特征,使特征更具表达力。采用Soft-NMS替换原有候选框合并策略,设计衰减函数提高目标框定位精度。采集制作真实环境下的表情数据集,通过训练最终实现野外环境下各表情的识别与定位,在精度上取得了很好的效果。

1 Faster RCNN 网络结构

Faster RCNN是目前主流的二阶段检测网络,是由RPN和Fast RCNN合并而来,每一阶段的网络都可以输出检测类别与边框定位,以网络结构分析,Faster RCNN网络可以分为3个部分,基础特征提取网络、区域建议网络RPN和Fast RCNN检测网络,算法的具体步骤如下,算法框架如图1所示。

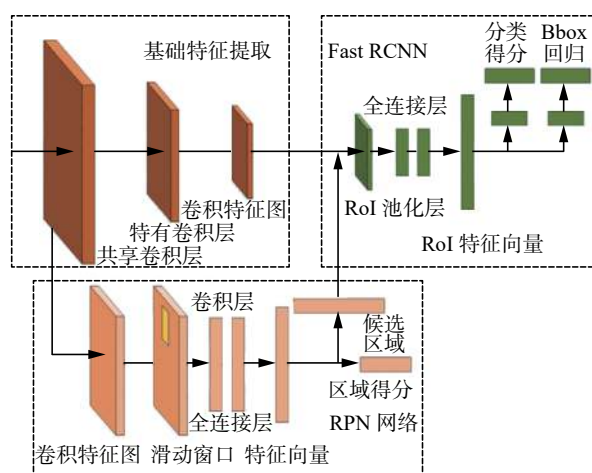


图1 Faster RCNN 网络结构

Fig. 1 Faster RCNN architectures

1.1 特征提取网络

特征提取网络部分由卷积神经网络CNN构成,CNN基本结构包括卷积层、池化层、全连接层及softmax分类层。使用不同的CNN会对检测精度、检测时间等造成不同的影响。

Faster RCNN常采用的特征提取网络有3个,分别是ZFNet、VGG-16、ResNet,其中1)ZFNet^[22]是在AlexNet的基础上进行细节改动,减少卷积核数量及步长大小,保留更多的特征,从中也可推理出网络深度增加,网络特征提取性能越好,特征提取效果也越优秀;2)VGG-16^[23]验证了卷积神经网络深度与性能之间的关系,通过反复堆叠3×3的卷积核与2×2的最大池化层而来。VGG-16网络结构简单,特征提取效果好,但

是参数大,训练的特征数量多,对硬件要求高;
3)ResNet^[24] 又称为残差网络,设计一种残差模块,解决网络深度增加时带来的梯度消失问题,实现单位映射之间的连接路线,能够提取目标更深层次的特征,实现很好的识别效果。

1.2 区域建议网络 RPN

RPN 用来提取候选区域,结构如图 2。接收来自基础特征提取网络传入的卷积特征图,通过卷积核将每一个 3×3 的滑动窗口 (sliding window) 卷积成为 256 维的特征向量。对每一个滑动窗口通过 1×1 的卷积输出为两个全连接层,即边框分类层 cls layer(box-classification layer) 和边框回归层 reg layer(box-regression layer)。cls layer 输出属于前景和背景的概率,reg layer 输出预测区域的中心点坐标: x, y 和长宽: w, h 4 个参数。滑动窗口中心对应的感受野来判断是否存在目标,由于目标长宽大小不一,以 16 为基准窗口大小,通过 (8,16,32)3 种窗口尺度和 (1:2,1:1,2:1)3 种长宽比生成 k 个 anchor 对特征图进行多尺度多点位采样。

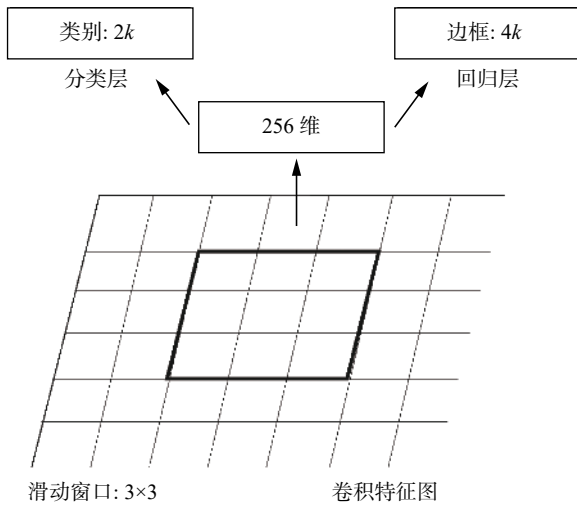


图 2 RPN 结构

Fig. 2 Region proposal networks structure

RPN 的损失函数定义为

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{\text{cls}}} \sum_i L_{\text{cls}}(p_i, p_i^*) + \lambda \frac{1}{N_{\text{reg}}} \sum_i p_i^* L_{\text{reg}}(t_i, t_i^*) \quad (1)$$

式中:角标 i 表示 anchor 的索引; p_i 表示每一个 anchor 中对应 $k+1$ 类 (k 个类别+1 个背景) 的概率分布; p_i^* 表示是否含有目标 (有目标则为 1, 反之为 0); N_{cls} 为 mini-batch 大小 (一般为 256); N_{reg} 为 anchor 数量; λ 为平衡权重,取值为 1; t_i 是建议框坐标 $\{t_x, t_y, t_w, t_h\}$; t_i^* 是标记框的坐标,具体参数值如下:

$$\begin{aligned} t_x &= \frac{x - x_a}{w_a}, t_y = \frac{(y - y_a)}{h_a} \\ t_w &= \log\left(\frac{w}{w_a}\right), t_h = \log\left(\frac{h}{h_a}\right) \\ t_x^* &= \frac{x^* - x_a}{w_a}, t_y^* = \frac{y^* - y_a}{h_a} \\ t_w^* &= \log\left(\frac{w^*}{w_a}\right), t_h^* = \log\left(\frac{h^*}{h_a}\right) \end{aligned} \quad (2)$$

$x, x_a, x^*(y, w, h$ 同理) 分别表示建议框、anchor 框和标定框的位置参数。

分类损失 L_{cls} 是目标和非目标的对数损失:

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

回归损失 $L_{\text{reg}}(t_i, t_i^*) = R(t_i - t_i^*)$, 其中

$$R(t_i - t_i^*) = \begin{cases} 0.5(t_i - t_i^*)^2, & |t_i - t_i^*| < 1 \\ |t_i - t_i^*| = 0.5, & \text{其他} \end{cases} \quad (4)$$

RPN 通过损失函数对边框进行回归,并对检测器输出的预测框进行非极大值抑制方法合并,作为输入连接到 Fast RCNN 中。RPN 产生的候选区域与特征提取网络输出的特征图相映射,ROI 池化层对于不同大小的候选区域输入都能得到固定维度的输出,再通过 cls layer 和 reg layer 得到最终的结果。

2 改进的 Faster RCNN

2.1 密集连接网络

采用更深的特征提取网络能够提取更深层次的语义信息,但是随着网络加深,参数不可避免地加大,这给网络优化和实验硬件带来一系列问题,在本文的面部表情检测算法中,单独制作出来的数据集样本数量少,网络训练容易造成过拟合,采用 DenseNet 密集连接网络作为特征提取网络可以解决上述问题。

DenseNet 借鉴了 ResNet 的思想,与 ResNet 网络不同,是全新的网络结构。两种网络结构最直观的区别在于每一个网络模块的传递函数不同。

$$x_l = H_l(x_{l-1}) + x_{l-1} \quad (5)$$

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]) \quad (6)$$

式 (6) 为 ResNet 网络传递函数,可以看出该网络第 l 层的输出是 $l-1$ 层输出的非线性变化加 $l-1$ 层的输出。而 DenseNet 一个网络模块第 l 层的输出是前面所有层输出的非线性变换集合,网络模块 (Dense Block) 如图 3 所示。

每一个 Dense Block 内的卷积都互相连接, H 表示对每个输入使用 Batch Norm、ReLU,用 k 维的 3×3 卷积核进行卷积,保证每个节点输出同样维度的特征图。 k 表示每一层卷积输出特征图的厚度,相比其他网络输出特征图厚度能够达

到几百甚至上千, DenseNet 整体厚度仅为 32。因为网络中每个模块的密集连接能够有效利用浅层与深层特征, 能够使网络高效而狭窄, 并且大幅度减少网络复杂程度与计算量, 连接节点 H_i 的参数如图 4。

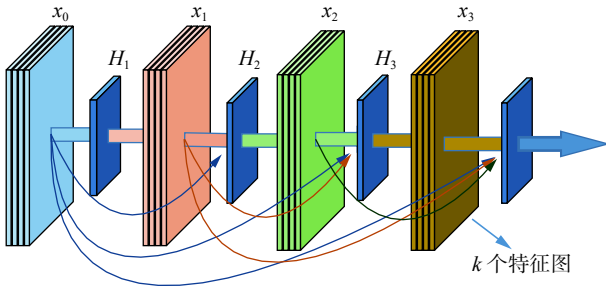
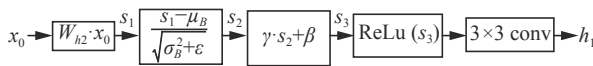


图3 Dense Block 结构

Fig. 3 Dense Block architectures

图4 节点 H_i 参数Fig. 4 Node H_i parameter

本文采用 4 个 Dense Block 网络 121 层作为特征提取网络, 去除全连接层和分类层, 再连接 RPN 及 RoI 池化层, 完成目标识别与定位。4 层 Dense Block 结构参数如表 1 所示。

表1 DenseNet 结构参数

Table 1 DenseNet structure parameters

结构	参数
卷积层	7×7 conv, stride 2
池化层	3×3 max pool, stride 2
Dense block(1)	$\begin{bmatrix} 1 \times 1 \text{conv} \\ 3 \times 3 \text{conv} \end{bmatrix} \times 6$
连接层1	1×1 conv
	2×2 average pool, stride 2
Dense block(2)	$\begin{bmatrix} 1 \times 1 \text{conv} \\ 3 \times 3 \text{conv} \end{bmatrix} \times 12$
连接层2	1×1 conv
	2×2 average pool, stride 2
Dense block(3)	$\begin{bmatrix} 1 \times 1 \text{conv} \\ 3 \times 3 \text{conv} \end{bmatrix} \times 24$
连接层4	1×1 conv
	2×2 average pool, stride 2
Dense block(4)	$\begin{bmatrix} 1 \times 1 \text{conv} \\ 3 \times 3 \text{conv} \end{bmatrix} \times 16$

2.2 非极大值抑制

非极大值抑制 NMS(non-maximum suppres-

sion) 是检测流程中重要的组成部分, 本质是搜索局部极大值, 抑制非极大值元素。Faster RCNN 会在图片中生成一系列检测框 $B=\{b_1, b_2, \dots, b_N\}$ 和对应的检测框得分集合 S_i , NMS 算法将选出最大得分前的物体检测流程中的检测框 M , 与其余的检测框进行重叠度 IoU(Intersection over Union) 计算, 如果计算结果大于设定阈值 N_t 则此检测框将被抑制。NMS 算法公式如下:

$$S_i = \begin{cases} S_i & \text{IoU}(M, b_i) < N_t \\ 0 & \text{IoU}(M, b_i) \geq N_t \end{cases} \quad (7)$$

式中 IoU 计算公式如下:

$$\text{IoU} = (A \cap B) / (A \cup B) \quad (8)$$

其中 A, B 为两个重叠的检测框:

从式 (7) 中可以看出 NMS 算法会将与检测框 M 相邻并大于阈值的检测框归零, 如果一个待检测目标在重叠区域出现, NMS 算法则会导致该目标检测失败, 降低检测模型的准确率。

针对这个问题, 本文使用 Soft-NMS 算法替代传统的 NMS 算法。在该算法中, 相邻检测框基于重叠部分的大小设置一个衰减函数而非将其分数置为零, 保证相邻目标能够准确识别。Soft-NMS 公示表示如下:

$$S_i = \begin{cases} S_i & \text{IoU}(M, b_i) < N_t \\ S_i(1 - \text{IoU}(M, b_i)) & \text{IoU}(M, b_i) \geq N_t \end{cases} \quad (9)$$

本文改进 Faster RCNN 检测算法的前端特征提取网络及末端回归器, 用于完成真实环境下面部表情检测, 算法流程如下所示:

算法 改进 Faster RCNN 流程

1) 输入图像 A , 调整图像尺寸, 输出为规定尺寸 $M \times N$ 的图 B ;

2) B 作为特征提取模块的输入, 通过 DenseNet 得到多层次融合特征图 C ;

3) C 作为区域建议 (RPN) 的输入, 采用滑动窗口的方法得到 300 个 proposals: D 。RPN 使用边框回归改变生成的 anchors, 使之更加接近标记框;

4) C 与 D 作为感兴趣区域 (RoI) 的输入, 得到建议框与特征图之间的映射图 E 。

5) 把 E 分别输出到分类器与回归器两个分支中。分类器采用 Softmax 对 E 进行分类识别, 回归器采用边框回归 Soft-NMS 进一步纠正边框, 最终分类目标并定位。

3 实验

3.1 数据集制作及处理

为验证所提出 Faster RCNN 面部表情检测算

法的有效性,本文独自采集愤怒、厌恶、恐惧、开心、悲伤和惊讶6类表情数据集,共4152张图片。

为保证检测模型能够完成日常生活情况下的面部表情检测,数据具有不同的光照强度、不同的人物位姿、复杂的背景及多个目标,包含不同肤色、年龄、种族等,并对数据中50%的数据集采取镜像扩充,50%数据集采取平移扩充,通过LabelImg软件对数据进行标注,如图5所示。扩充后的数据集为8304张图片,其中90%作为训练集,10%作为测试集,数据集数量分布如表2所示。

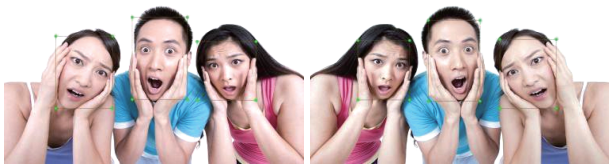


图5 数据扩充及标注

Fig. 5 Data expansion and labeling

表2 数据参数

Table 2 Data parameters

数据集	愤怒	厌恶	恐惧	开心	悲伤	惊讶
总数	1506	1034	1398	1430	1456	1480
训练集	1204	827	1118	1144	1164	1184
测试集	151	103	140	143	147	148

从测试数据集中挑选出困难样本用于对比改进算法在复杂背景下的准确率。其中困难样本的选取范围为图片中检测目标多于4个,面部有光照影响,面部遮挡及侧面情况。困难样本测试数据如表3所示。在困难样本中,部分图像具备多个困难属性,下文介绍。图像存在多个人物表情且存在属于黑夜拍摄,该图像既属于多目标类别样本也属于光照影响样本。

表3 不同场景的数据

Table 3 Test sets of different scenarios

困难样本	多目标	光照影响	遮挡	侧面
数量	143	73	61	39

考虑算法在不同环境下的有效性,本文添加日本女性面部表情JAFPE(Japanese Female Facial Expressions)数据集^[25]进行对照实验。JAFPE数据由10名女性的7种表情构成,包括6种基本情绪和一种中性情绪,总共213副图像,原始图像为256像素×256像素大小,数据都已经进行过裁剪和调整,人物面部居中,仅有少量光照差别,是一个质量较高的面部表情数据集。试验选取数据

集中6类相关表情进行验证。

3.2 实验参数及评价指标

由于DenseNet在传输过程中需要融合当前阶段所有特征图,对显存要求巨大,因此采用密集连接网络的高效内存实现方法。提出两个预先分配的共享内存存储位置,存放用来连接的共享特征图。在正向传递期间,将所有中间输出分配给这些存储器块;在反向传递期间,根据需要即时重新计算更新传递函数。采用这种策略使得DenseNet在增加较少的计算开销下能够在单块显卡中工作。

实验基于Tensorflow框架,采用I76700处理器,内存为32G,显卡GeForce RTX2080Ti,显存为11G进行训练。实验数据由个人采集,有生活照、剧照等不同场景下的人物表情,并由LabelImg软件进行人工标注。

总数据训练迭代10万次,批大小Batchs为64,初始学习率设置为0.001,并且在训练时期总数的75%后设置为0.0001。

评价指标为平均精度AP(Average-Precision),是Precision-recall(P-R)曲线所围成的面积。在P-R曲线中,P表示精确率,R表示召回率,计算如下式:

$$P = TP / (TP + FP) \quad (10)$$

$$R = TP / (TP + FN) \quad (11)$$

式中:TP(True positives)为正确样本被识别为正样本的数量;FP(False positives)为负样本被错误识别为正样本的数量;FN为正样本被错误识别为负样本的数量。AP值表示单个类别的识别准确率,越高表示网络模型性能越好。mAP(mean Average-Precision)表示所有类别总体识别准确率,与AP值之间的关系如式(12)所示。

$$mAP = \frac{\sum_{i=1}^C AP_i}{C} \quad (12)$$

3.3 结果比较及分析

分别用ResNet和DenseNet作为特征提取网络进行训练,使用测试集对网络模型进行测试,得到每类表情的AP值如表4所示,困难样本的检测对比结果如表5所示。

表4 不同网络模型的测试结果

Table 4 Test results of different network models

网络	愤怒	厌恶	恐惧	开心	悲伤	惊讶
ResNet	76.2	72.8	79.3	81.8	79.8	83.1
DenseNet	80.1	74.8	83.6	87.4	84.3	86.9
DenseNet+SoftNMS	80.1	74.8	83.6	88.8	84.6	88.2

表5 困难样本测试结果
Table 5 Test results of difficult sample

网络	多目标	光照影响	遮挡	侧面
原始算法	75.6	75.3	67.2	68.3
改进算法	79.7	76.7	72.4	71.8

从检测结果可以看出,采用ResNet的Faster RCNN在各类表情检测中mAP达到78%以上,部分检测结果如图6所示。图6(a)中目标特征明显,光照充足,模型能够达到很好的检测结果,图6(b)中脸部特征有部分遮挡,且含有不同表情类型,检测效果令人满意,而图6(c)中存在漏检情况,可以看出使用ResNet具有一定的检测能力,但依旧存在一些漏检和误检情况。这是因为数据量过小,ResNet无法充分训练,在复杂情况下鲁棒性不高。



(a) 无遮挡



(b) 有遮挡



(c) 漏检

图6 ResNet-Faster RCNN 检测效果
Fig. 6 ResNet-Faster RCNN detection result

从表4可以看出采用DenseNet-121作为特征提取网络mAP能够达到83%,相比ResNet提高5%。其中在开心、悲伤、惊讶这3类表情中,模型

检测结果提高较多,因为这3类表情的测试集存在多目标、有遮挡及复杂背景的样本,DenseNet能够提取目标更多的特征,达到更好的效果。采用Soft-NMS对检测框进行改进,准确率分别在开心、悲伤、惊讶3类存在多目标样本的数据集中提高了一个百分点,说明Soft-NMS在多目标及目标重叠情况下能够避免检测框重复度高于阈值导致的候选框归零的错误,达到更好的检测效果。通过表5则可以看出,改进的检测网络在困难样本中相比原版具有更高的鲁棒性,其中在多目标、遮挡及侧面3类样本中提高较为明显。采用改进Faster RCNN算法与原始Faster RCNN在多目标复杂背景下的效果对比如图7。



(a) 复杂环境下漏检



(b) 错检



(c) 漏检

图7 原始网络与改进网络对比

Fig. 7 Comparison of accuracy with different backbone

图7中选取多目标、多位姿及黑夜情况下的样本进行检测,第1行为原始网络检测结果,第2行为改进网络的检测结果。结果表明,原始网络在复杂情况下普遍存在漏检情况。图7(b)中原始网络出现错误检测,表情类型应为悲伤,而检测结果为恐惧。图7(c)中出现漏检,并且检测框并未准确包含面部目标。对比可以看出采用DenseNet及Soft-NMS改进的Faster RCNN在检测结果上优于原始的Faster RCNN,能够检测到更多的目标表情,并在相邻目标距离过近的情况下准确地框选出独立个体。在背景复杂及多目标情况下改进的Faster RCNN性能提高更为明显。

考虑制作的数据集具有相似的环境特性,试

验加入 JAFFE 数据集来验证网络模型在不同环境下的检测性能。试验将已训练好的网络模型在 JAFFE 数据中测试,用来比较原版网络与改进网络在新数据集中的泛化性。将 JAFFE 数据的 50% 纳入训练集作为第 3 组对照实验,对比数据集对实验结果造成的影响,对比结果如图 8 所示。

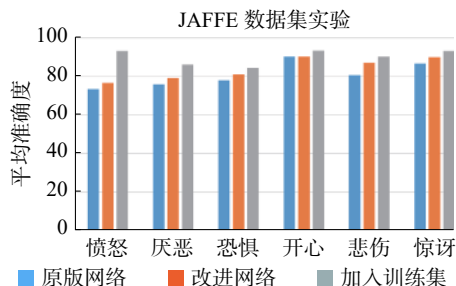


图 8 原始网络与改进网络对比

Fig. 8 Comparison of test results

由于 JAFFE 数据集为单目标高质量数据集,光照影响小、背景单一,与原数据集差异较大,因此算法的提高不如在原数据集中明显,但依旧可以看出改进算法较原始算法具有一定优势。将 50% 数据集加入训练集后,算法模型在 JAFFE 数据集中性能有明显提高,部分检测结果如图 9 所示,可以看出数据集对算法具有较大影响。如果进一步扩充数据集的多样性,算法的鲁棒性将进一步提高。

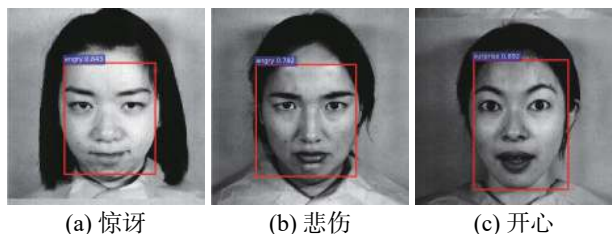


图 9 JAFFE 数据集检测结果

Fig. 9 Test results of JAFFE dataset

通过一系列实验对比,可以看出改进的 Faster RCNN 算法在面部表情检测任务中具有较高的检测精度,同时在不同环境中具备良好的鲁棒性,提高了算法的应用范围,更有实际应用价值。

4 结束语

针对传统表情检测算法对于环境光线不同、背景多样及位姿角度变化等情况下无法发挥有效作用的问题,提出深度学习表情检测算法,以 Faster RCNN 为基础改进,使用密集连接网络作为特征提取模块,每个模块的密集连接能够有效利用浅层与深层特征,提高网络对面部表情检测的准确率,采用 Soft-NMS 替换原有的 NMS 算法,优

化候选框合并策略,使候选框更加精确。制作真实环境下的表情数据集,并进行扩充,提高训练模型的鲁棒性。本文提出的检测算法能够实现日常生活中的多目标面部表情检测,在黑夜、部分遮挡、佩戴饰品等复杂情况下取得较好的精度,达到了良好的检测效果。

参考文献:

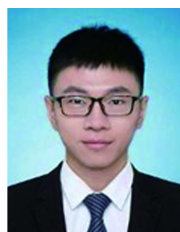
- [1] CALVO R A, D'MELLO S. Affect detection: an interdisciplinary review of models, methods, and their applications[J]. *IEEE transactions on affective computing*, 2010, 1(1): 18–37.
- [2] IZARD C E. Innate and universal facial expressions: evidence from developmental and cross-cultural research[J]. *Psychological bulletin*, 1994, 115(2): 288–299.
- [3] SUN Zhe, HU Zhengping, ZHAO Mengyao. Automatically query active features based on pixel-level for facial expression recognition[J]. *IEEE access*, 2019, 7: 104630–104641.
- [4] ALPHONSE A S, STARVIN M S. A novel maximum and minimum response-based Gabor (MMRG) feature extraction method for facial expression recognition[J]. *Multimedia tools and applications*, 2019, 78(16): 23369–23397.
- [5] 童莹. 一种方向性的局部二值模式在人脸表情识别中的应用[J]. *智能系统学报*, 2015, 10(3): 422–428.
- TONG Ying. Local binary pattern based on the directions and its application in facial expression recognition[J]. *CAAI transactions on intelligent systems*, 2015, 10(3): 422–428.
- [6] SADEGHI H, RAIE A A. Histogram distance metric learning for facial expression recognition[J]. *Journal of visual communication and image representation*, 2019, 62: 152–165.
- [7] HAPPY S L, ROUTRAY A. Automatic facial expression recognition using features of salient facial patches[J]. *IEEE transactions on affective computing*, 2015, 6(1): 1–12.
- [8] BEJAOU H, GHAZOUANI H, BARHOUMI W. Sparse coding-based representation of LBP difference for 3D/4D facial expression recognition[J]. *Multimedia tools and applications*, 2019, 78(16): 22773–22796.
- [9] LI Ruiqi, TIAN Jing, CHUA M C H. Facial expression classification using salient pattern driven integrated geometric and textual features[J]. *Multimedia tools and applications*, 2019, 78(20): 28971–28983.
- [10] ZANGENEH E, MORADI A. Facial expression recognition by using differential geometric features[J]. *The imaging science journal*, 2018, 66(8): 463–470.
- [11] HAPPY S L, ROUTRAY A. Fuzzy histogram of optical

- flow orientations for micro-expression recognition[J]. *IEEE transactions on affective computing*, 2019, 10(3): 394–406.
- [12] LU Hua, KPALMA K, RONSIN J. Motion descriptors for micro-expression recognition[J]. *Signal processing: image communication*, 2018, 67: 108–117.
- [13] 刘帅师, 郭文燕, 张言, 等. 鲁棒的正则化编码随机遮挡表情识别 [J]. *智能系统学报*, 2018, 13(2): 261–268.
LIU Shuaishi, GUO Wenyang, ZHANG Yan, et al. Recognition of facial expression in case of random shielding based on ro-bust regularized coding[J]. *CAAI transactions on intelligent systems*, 2018, 13(2): 261–268.
- [14] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. *Science*, 2006, 313(5786): 504–507.
- [15] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets[J]. *Neural computation*, 2006, 18(7): 1527–1554.
- [16] YANG Biao, CAO Jinmeng, NI Rongrong, et al. Facial expression recognition using weighted mixture deep neural network based on double-channel facial images[J]. *IEEE access*, 2017, 6: 4630–4640.
- [17] WU Min, SU Wanjuan, CHEN Luefeng, et al. Weight-adapted convolution neural network for facial expression recognition in Human-robot interaction[J]. *IEEE transactions on systems, man, and cybernetics: systems*, 2019, 51(3): 1473–1484.
- [18] SALMAM F Z, MADANI A, KISSI M. Fusing multi-stream deep neural networks for facial expression recognition[J]. *Signal, image and video processing*, 2019, 13(3): 609–616.
- [19] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus, Ohio, USA, 2014: 580–587
- [20] GIRSHICK R. Fast R-CNN[C]//Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile, 2015: 1440–1448.
- [21] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2017, 39(6): 1137–1149.
- [22] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks[C]//Proceedings of the 13th European Conference on Computer Vision (ECCV). Zurich, Switzerland: Springer, 2014: 818–833.
- [23] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2019-12-12]. <http://arXiv:1409.1556>, 2014.
- [24] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA, 2016: 770–778.
- [25] LYONS M, AKAMATSU S, KAMACHI M, et al. Coding facial expressions with Gabor wavelets[C/OL]. Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition. Nara, Japan: IEEE, 1998.[2012-12-05]. <http://www.doc88.com/p-6921152816768.html>.

作者简介:



伍锡如, 教授, 博士, 主要研究方向为深度学习、神经网络、机器人控制。主持国家自然科学基金项目 2 项, 主持广西省自然科学基金项目 3 项, 获国家发明专利 10 余项。出版专著 1 部、教材 1 部, 发表学术论文 40 篇。



凌星雨, 硕士研究生, 主要研究方向为深度学习、计算机视觉。