

教室场景下人脸检测与识别

杨凯雯¹, 杨佳乐²

(1.沈阳工业大学信息学院, 辽宁 沈阳 110870;

2.沈阳工业大学人工智能学院, 辽宁 沈阳 110870)

✉527410159@qq.com; 2082873733@qq.com



摘要: 提出一种教室场景下人脸检测与识别的算法, 基于RetinaFace人脸检测框架进行改进, 在主干网络中引入可变形卷积以适应人脸遮挡以及人脸变形, 调整预设Anchor并在上下文敏感模块中引入残差结构以适应教室场景下尺度变化的特点。在公开数据集WIDER FACE上训练基础权重, 然后在教室场景下自标注的数据集中进行迁移学习以适应教室场景, 最后通过ArcFace人脸识别网络进行人脸识别。本算法在公开数据集WIDER FACE上batch size设置为16时, Easy、Medium和Hard的人脸检测精度分别为96.34%、95.12%和89.64%; 在自标注的数据集上batch size设置为4时, 人脸检测精度为94.72%, 人脸识别精度为92.11%。实验结果表明, 该算法可以有效提高教室场景下人脸检测与识别的效果。

关键词: 人脸检测; 人脸识别; RetinaFace; ArcFace

中图分类号: TP399 **文献标识码:** A

Face Detection and Recognition in Classroom Scenarios

YANG Kaiwen¹, YANG Jiale²

(1.School of Information, Shenyang University of Technology, Shenyang 110870, China;

2.School of Artificial Intelligence, Shenyang University of Technology, Shenyang 110870, China)

✉527410159@qq.com; 2082873733@qq.com

Abstract: This paper proposes an algorithm for face detection and recognition in classroom scenes, which is improved based on the RetinaFace face detection framework. Deformable convolution is introduced into the core network to adapt to face occlusion and face deformation. The preset Anchor is adjusted and the residual structure is introduced into the context-sensitive module to adapt to the characteristics of scale changes in classroom scenes. The basic weights are trained on the public dataset WIDER FACE, and then the self-labeled dataset in the classroom scene is used for transfer learning to adapt to the classroom scene. Finally, face recognition is carried out through ArcFace face recognition network. When the batch size of this algorithm is set to 16 on the public dataset WIDER FACE, the face recognition precision rates of Easy, Medium and Hard are 96.34%, 95.12% and 89.64% respectively. When the batch size on the self-labeled dataset is 4, the face detection precision is 94.72%, and the face recognition precision is 92.11%. Experimental results show that the algorithm can effectively improve the effect of face detection and recognition in classroom scenes.

Keywords: face detection; face recognition; RetinaFace; ArcFace

1 引言(Introduction)

如今校园教学中的考勤系统大部分通过传统点名签到的人工考勤方式来记录学生的出勤情况, 这样不仅会耗费师生大量的时间, 更有可能影响高校的教学进度和安排。人脸检测和人脸识别技术为教室考勤提供了新的方向, 从图像中快

速、准确地检测到人脸位置信息是人脸识别技术的关键, 如果人脸检测的性能不佳, 势必会导致人脸识别的效果大打折扣。

在深度学习广泛应用前, 著名的人脸检测器MTCNN^[1]使用图像金字塔的方法来检测不同分辨率的人脸目标, 后来提出了将不同层级的特征图融合起来以提升目标检测的性能,

即特征金字塔网络FPN^[2]。还有其他检测不同尺度大小人脸特征的方式如SSH^[3]、R-FCN和PyramidBox^[4]，以及对Anchor采样和匹配策略改进的方式如FaceBoxes、S3FD和半监督学习方法maskedFaceNet^[5]等。

RetinaFace^[6]是2019年提出的人脸检测模型，原模型添加了SSH网络的检测模块，标注了WIDER FACE^[7]的人脸五个关键点标注数据，融合了多任务损失，同时提出了一个自监督的人脸编码器用于人脸的检测。

2 人脸检测(Face detection)

2.1 RetinaFace网络

本文对比了多个人脸检测网络的性能，最终选用RetinaFace人脸检测框架进行实验。RetinaFace是一个单阶段的人脸检测网络，网络结构图如图1所示，在多特征层级的基础上使用了特征金字塔结构，更充分地利用了多层级信息，并通过上下文融合模块进行不同层级间信息的特征融合，采用多任务损失函数获得更好的人脸检测效果^[8]。相对于自然场景，教室场景下人脸有固定的尺度特征范围，且存在学生姿态轻微扭曲以及前后排遮挡问题，本文针对以上问题将骨干网络替换为ResNet50，并从检测框策略、信息融合模块和迁移学习等方面对人脸检测网络进行改进。

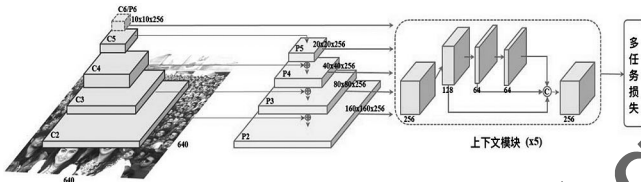


图1 RetinaFace网络结构图

Fig.1 RetinaFace network structure

2.2 可变形卷积

卷积神经网络在目标检测与识别任务上解决了很多问题，激活函数、网络结构和损失函数也在不断更迭创新，而卷积层和池化层的操作一般都只能在正方形的区域内对应地进行映射运算，如图2(a)所示，这种大小比例固定的方法对于教室场景下学生人脸姿态轻微扭曲的问题检测效果不够好。

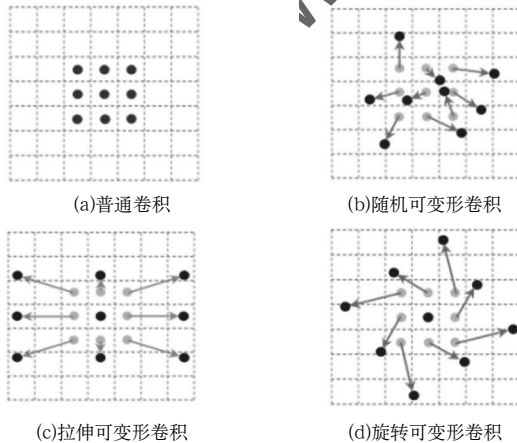


图2 传统卷积与可变形卷积比较

Fig.2 Comparison of traditional convolution and deformable convolution

本文在骨干网络ResNet50中引入了可变形卷积结构^[9]，

对每一层卷积核的对应位置增加偏移，加上偏移量的学习之后，可变形卷积核的大小和位置可以根据当前需要识别的图像内容进行动态调整，其直观效果就是不同位置的卷积核采样点位置会根据图像内容发生自适应的变化，以适应不同物体的形状、大小等几何形变，从而提高对教室中每个学生脸部的大小、姿态和扭曲情况下的检测效果。

2.3 Anchor缩减策略

经过计算，RetinaFace一共需要在特征金字塔的五层特征层上生成102,300个Anchor预测框，如表1所示，并提取出每个预测框内的特征信息进行判断以及筛选是否存在人脸和置信度等信息进行人脸检测。虽然这种方式可检测到的人脸尺度范围非常大，但是在训练和测试过程中耗费的算力和显存也是非常巨大的。考虑到教室场景下人脸尺度大小范围较自然场景下范围较小，人脸图像约为20—100像素，所以为了避免显存的浪费，本设计去掉了P5和P6两层特征层，将五层特征金字塔缩减为可以检测16—101.59像素大小人脸图像的三层特征层结构，同时上下文模块也由原始的五个缩减为三个融合模块。

表1 FPN结构中的Anchor策略

Tab.1 Anchor strategy in FPN structure

特征层	步长	先验框
P2(160×160×256)	4	16, 20, 16, 25, 40
P3(80×80×256)	8	32, 40, 32, 50, 80
P4(40×40×256)	16	64, 82, 63, 101, 59
P5(20×20×256)	32	128, 161, 26, 203, 19
P6(10×10×256)	64	256, 322, 54, 406, 37

2.4 上下文预测模块

RetinaFace网络结构中的信息融合模块使用的是SSH上下文模块，SSH通过在不同分支上不同的卷积核层堆叠来扩展感受野的大小，学习更多的上下文信息；DSSD网络在单阶段检测网络SSD^[10]中引入残差网络，从而得到更深度的预测分支。而本设计使用的CPM上下文预测模块借鉴了以上两种思路。CPM模块结构如图3所示，在SSH上下文模块中的前面添加残差预测模块，在网络结构深度增加的同时增加了网络宽度以提取更深层次的特征信息，既能增加特征信息提取能力，又防止由于网络深度增加而带来的梯度消失和梯度爆炸问题，使得预测模块在分类和定位上能得到更好的效果。

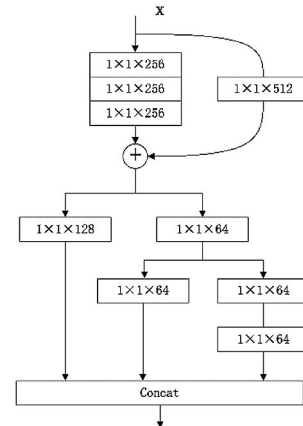


图3 CPM模块结构

Fig.3 CPM module structure

2.5 迁移学习

目前很多骨干网络通过预训练提取特征的效果较好,但是这些网络基本都是在ImageNet数据集上进行预训练的,而实际中的各种任务目标各有不同,如果只依靠ImageNet数据集进行预训练可能无法涵盖各个领域目标检测的需求。由于教室场景下数据量有限,因此本文使用迁移学习中fine-tuning的方式,在大型公开数据集上预训练得到的网络模型基础上,使用教室场景下自标注的数据集再次训练人脸检测网络,从而使这个网络能够学习到教室场景下的特征特点,在数据集较小的情况下训练出效果更好、更适合教室场景的人脸检测模型。

3 人脸识别(Face recognition)

3.1 ArcFace模型

在人脸识别算法的改进上,国内外研究者首先将精力放在提出一个识别、验证、聚类等问题的一解决框架,考虑如何将人脸更好地映射到特征空间;后来考虑到深度学习的网络复杂程度已经很高,研究者们便开始了对损失函数优化的研究。ArcFace^[11]是目前人脸识别性能最好的开源模型之一,该模型提出了新的损失函数以更好地缩小预测与实际数据的差距,从而得到性能更好的人脸识别模型。

3.2 ArcFace损失函数

无论是SphereFace、CosFace、CurricularFace还是ArcFace的损失函数,都是基于Softmax函数进行改进的。ArcFace在SphereFace的基础上改进了对特征归一化和加性角度间隔,提高了类间可分性的同时增大了类内紧度和类间差异。ArcFace损失函数如下式所示,自然数的指数部分使用的是 $\|x\|_2 \cos(\theta + m)$,其中 m 是间距, $\|x\|_2$ 是样本大小, s 指scale参数。

$$L_{\text{arc}} = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s \cos(\theta_{yi} + m)}}{e^{s \cos(\theta_{yi} + m)} + \sum_{j \neq y_i} e^{s \cos(\theta_{ji})}}$$

以二分类为例,通过上述各人脸识别模型的损失函数可以求得SphereFace、CosFace和ArcFace损失函数的分类边界函数。将分类边界公式用二维坐标表示,如图4所示,由图中可以直观地看出ArcFace是直接在角度空间最大化分类边界。

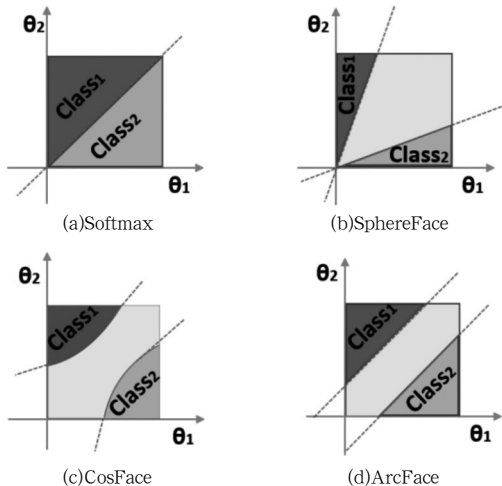


图4 各个边界函数的二维坐标图

Fig.4 2D coordinate plot of each boundary function

4 实验(Experiment)

4.1 实验参数

本文运行实验的平台环境为1080Ti GPU、CUDA 10.2、Cudnn v7和PyTorch组成的框架。本文人脸检测模型使用在WIDER FACE数据集上预训练的ResNet50模型为基础网络,预训练权重采用SGD优化方法,batch size设置为4,初始化学习率设置为0.001,动量为0.9,权重衰减为0.0005。输入图像大小为1920×1080,通过主干网络提取特征前改变尺寸为840×840。人脸识别网络在CASIA-WebFace数据集中3,000个名人以及教室场景下的学生合并得到的数据集下进行训练,主干网络为ResNet50,损失函数为ArcFace Loss,采用SGD优化方法,共训练130个轮次得到最优训练结果。

4.2 消融实验

本文分别以batch size为4和16在WIDER FACE数据集和自标注数据集上进行了对比试验,验证人脸检测算法是否有效。

实验一:如表2所示为自标注数据集下显存使用情况及精度,第一行为原始五层特征金字塔,第二行为缩减后的三层特征金字塔。通过第二列的比较可以得出将特征层进行缩减后,训练时耗费的显存明显减少,检测精度轻微下降。如表3前两行数据所示,第一行是batch size设置为4时,原始RetinaFace网络在公开数据集WIDER FACE上的实验结果;第二行是对原始RetinaFace网络进行缩减Anchor策略后的实验结果。由表2、表3综合对比得知,Easy、Medium和Hard的精度都有0.58%—1.62%的下降。由于Anchor缩减策略本质上是特征金字塔去掉了两层特征层,并调整每层特征层上的预设Anchor以减少算力和显存的浪费,导致网络可以检测到的人脸框大小范围缩小了,因此在公开数据集上检测精度下降,但对自标注的数据集检测效果几乎没有下降,表明Anchor缩减策略符合且适应本文场景数据集尺度特征。

实验二:通过在主干网络ResNet50中引入可变形卷积以适应教室场景人脸姿态轻微扭曲问题。表3第三行是在Anchor缩减策略上加入了可变形卷积后的实验结果,与第二行未引入可变形卷积时相比精度Easy提高了2.75%,Medium提高了2.81%,Hard提高了1.87%。由此可以得出,可变形卷积的引入提高了公开数据集WIDER FACE的精度,增强了对变形人脸目标的检测效果。

实验三:通过替换为CPM模块,即在SSH上下文模块中的前面添加残差预测模块提高模型精度。表3第四行是替换为CPM模块后模型的精度实验结果,虽然与第三行实验结果差距较小,但是与原始RetinaFace人脸检测模型相比,公开数据集精度Easy提高了2.26%,Medium提高了3.1%,Hard提高了0.95%。可以得出替换为CPM模块后,在增加网络结构深度的同时增加了网络宽度以提取更深层次的特征信息,既能

增加特征信息提取能力，又防止由于网络深度增加而带来的梯度消失和梯度爆炸问题，使得预测模块在分类和定位上取得了更好的效果。

实验四：表3后四行为batch size设置为16时各方面改进后的对比实验结果，虽然最后一行全部改进后在公开数据集上的精度没有达到原始网络的实验结果，但是经过后三行数据的对比可以看出，无论是可变形卷积的加入还是CPM模块的替换，都有效地提高了人脸检测精度。

表2 自标注数据集Anchor缩减前后性能对比

Tab.2 Performance comparison of self-labeled dataset before and after Anchor reduction

特征金字塔层	显存/MiB	精度/%
P2—P6	11,032	90.35
P2—P4	8,666	90.23

表3 WIDER FACE数据集上对比实验结果

Tab.3 Comparative test results on the WIDER FACE dataset

骨干网络	batch size	先验框	可变形卷积	上下文模块	精度(WIDER FACE)/%		
					Easy	Medium	Hard
ResNet50	4	×	×	×	93.56	92.44	89.28
ResNet50	4	✓	×	×	92.98	91.68	87.66
ResNet50	4	✓	✓	×	95.73	94.49	89.53
ResNet50	4	✓	✓	✓	95.82	95.54	90.23
ResNet50	16	×	×	×	96.50	95.60	90.40
ResNet50	16	✓	×	×	95.07	94.02	89.64
ResNet50	16	✓	✓	×	95.98	94.82	90.77
ResNet50	16	✓	✓	✓	96.44	95.12	89.64

实验五：表4展示了在自标注的教室场景数据集上的消融实验。第一行是原始网络在自标注数据集上的实验结果。第二行是进行迁移学习fine-tuning的方法，由于自标注数据集较小且与公开数据集相似，检测目标相同而检测场景略有区别，通过fine-tuning方式能够提高人脸检测模型针对教室场景的检测效果，使得精度提高了1.66%，模型更加适应教室场景。第三行是缩减Anchor后的实验结果，由表2也可以得出，自标注数据集精度变化不大，但在训练时明显更加节省显存。第四行是加入可变形卷积后的实验结果，自标注数据集检测精度提高了2.37%。第五行替换为CPM模块后精度提高了2.12%，对比原始未改进网络检测结果，自标注教室场景下数据集人脸检测精度一共提高了6.03%，证明一系列网络改进对教室场景人脸检测以及识别的效果提高显著。

实验六：本文使用ResNet50作为基础特征提取网络、ArcFace作为损失函数对教室场景下人脸识别网络进行建模，得到教室场景下人脸识别精度为92.11%。

表4 自标注数据集上对比实验结果

Tab.4 Comparative test results on the self-labeled dataset

骨干网络	batch size	迁移学习	先验框缩减	可变形卷积	上下文模块	精度/%
ResNet50	4	×	×	×	×	88.69
ResNet50	4	✓	×	×	×	90.35
ResNet50	4	✓	✓	×	×	90.23
ResNet50	4	✓	✓	✓	×	92.60
ResNet50	4	✓	✓	✓	✓	94.72

从教室场景下人脸检测以及人脸识别结果上看，本文算法改进后的实验效果有一定程度的提高，具体检测示例效果如图5所示。图5(a)为教室场景下采集到的原始图像；图5(b)为本算法人脸检测效果，可以看出对于小尺寸以及略微遮挡的人脸目标检测效果良好，人脸检测置信度也较高；图5(c)为本算法人脸识别实验效果。



(a)原图



(b)本算法检测实验效果



(c)本算法识别实验效果

图5 测试结果示例

Fig.5 Examples of test results

5 结论(Conclusion)

本文提出一种针对教室场景下人脸检测与识别的算法，以提高学生出勤管理的效率。首先基于RetinaFace人脸检测框架进行了改进，在主干网络中引入可变形卷积以适应人脸遮挡以及人脸变形；针对教室场景下的尺度特征，本文调整预设Anchor并在上下文敏感模块中引入残差结构；先在公开数据集WIDER FACE上进行基础权重的训练，然后通过迁移学习的方法在教室场景下训练得到适应场景的人脸检测模型；最后通过ArcFace人脸识别网络进行人脸识别，并通过界面显示将检测和识别结果显示出来。本算法在公开数据集WIDER FACE上batch size设置为16时，Easy、Medium和Hard的人

(下转第23页)