ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 37,No. 1,Jan. 2022,pp. 73－81
DOI:10. 16337/j. 1004-9037. 2022. 01. 006

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

基于层级注意力增进网络的多尺寸遮挡人脸检测

王麟阁[1]，蒋宝军[2]，潘铁军[1]

(1. 宁波财经学院数字技术与工程学院,宁波 315175;2. 吉林建筑大学市政与环境工程学院,长春 130118)

摘　要：在SSD(Single shot multibox detector)单阶段人脸检测模型的基础上,针对复杂局部遮挡下人脸检测精确性差的问题,提出了一种基于层级注意力增进网络的多尺寸遮挡人脸检测方法。首先,在SSD基础网络的多层初始特征图上,通过引入注意力增进机制提升人脸可见区域的响应值。然后为不同增强特征层设计不同尺寸的锚框,以提高对多尺寸遮挡人脸的分层识别效果。最后在训练时将注意力损失函数、分类损失函数和回归损失函数融合为多任务损失函数,共同优化网络参数。在WIDER FACE人脸数据集和MAFA遮挡人脸数据集上的实验表明,本文方法的检测精确性和时效性均优于目前主流遮挡人脸检测方法。

关键词：遮挡人脸检测;特征图;注意力增进;锚框;损失函数

中图分类号：TP183　　**文献标志码**:A

Multi-size Occlusion Face Detection Based on Hierarchical Attention Enhancement Network

WANG Linge[1], JIANG Baojun[2], PAN Tiejun[1]

(1. College of Digital Technology and Engineering, Ningbo University of Finance & Economics, Ningbo 315175, China;2. School of Civil and Environmental Engineering, Jilin Jianzhu University, Changchun 130118, China)

Abstract: Based on the single shot multibox detector (SSD) single-stage face detection model, this paper proposes a multi-size occlusion face detection method based on a hierarchical attention enhancement network to solve the problem of poor accuracy of face detection under complex partial occlusion. Firstly, on the multi-layer original feature map of SSD basic network, the attention enhancement mechanism is introduced to improve the response value of the visible region of the face. Then, different anchor sizes are designed for different enhancement feature layers to improve the hierarchical recognition effect of multi-scale occluded face. In training, the attention loss function, the classification loss function and the regression loss function are fused into a multi-task loss function to jointly optimize the network parameters. Experiments on the WIDER FACE dataset and the MAFA occlusion face dataset show that the detection accuracy and timeliness of the method are better than those of the current mainstream occlusion face detection methods.

Key words: occlusion face detection; feature map; attention enhancement; anchor box; loss function

---

基金项目:2020年度宁波市“科技创新 2025”重大专项暨“246”产业集群发展支撑引领计划 (2020Z008); 浙江省高等教育“十三五”第二批教学改革研究项目 (jg20190514)。
收稿日期:2021-03-05;**修订日期**:2021-05-10

引 言

近年来,人脸表情识别[1]、人脸对齐[2]、人脸重建[3]和人脸聚类[4]等众多人脸检测问题得到了广泛的关注。在自然场景中,人脸图像会受到光照变化、姿态变化和局部遮挡等外部因素影响,当人脸区域出现墨镜、口罩、围巾和人体自遮挡时,会造成人脸检测区域的特征缺失,严重降低人脸检测的准确率。因此,如何有效减小遮挡区域的干扰,成为目前人脸检测领域亟待解决的问题。

传统遮挡人脸检测方法可分为两类。第1类方法通过对遮挡区域进行稀疏表示分类(Sparse representation classification, SRC)[5],在人脸检测时恢复缺失的人脸特征。第2类方法通过分块识别遮挡区域,在特征提取时避开受遮挡而损坏的部分[6-7]。由于传统方法提取的浅层特征表现能力较弱,上述方法已被基于深度学习的方法取代[8-9]。

目前基于深度学习的遮挡人脸检测方法可分为3类。第1类方法通过扩充训练集中遮挡人脸的数量,取得了检测性能的提高。例如,文献[10]对训练人脸样本进行分块,并对人脸分块区域采用构造解析方法随机生成合成遮挡人脸,提高了训练样本数量和遮挡人脸检测精度。但是训练样本的扩充只能保证更加均匀的特征提取,因此该方法的检测性能提升有限。第2类方法通过重建遮挡区域的特征描述,消除了由遮挡造成的人脸特征元素损坏。文献[11]通过融合多个人脸回归网络和去遮挡自编码网络,逐步恢复受遮挡损坏的人脸特征描述。文献[12]通过对遮挡区域执行鲁棒编码和循环遮挡去除,提高了遮挡人脸的检测精度。该类方法可在一定程度上弱化遮挡对识别的影响,但是随着遮挡程度的提高,网络模型编码和去遮挡的计算量急剧提升,且重建出的遮挡区域也会加快偏离真实人脸的特征描述。第3类方法通过抑制人脸遮挡区域的特征响应来减小遮挡造成的影响。文献[13]采用最大化剔除场景标签和多尺度弥补锚框匹配方法,增强了对尺寸较小遮挡人脸的检测精度。文献[14]在网络模型的中间层设置了掩膜生成机制,以降低遮挡区域权重的方式,尽量弱化遮挡区域的特征响应干扰。受限于掩膜生成过程的稳定性和监督性较低,其对遮挡区域的区分性较差。

近年来,SSD(Single shot multibox detector)[15-16]单阶段人脸检测模型以其高效可扩展的优点得到广泛关注和应用,本文在SSD单阶段人脸检测模型的基础上,针对复杂局部遮挡下人脸检测精确性差的问题,进行了以下3个方面的改进和创新:(1)在SSD的多个原始特征层通过引入注意力增进机制提升人脸可见区域的响应值,进而提出了人脸注意力增进网络;(2)为不同增强特征层设计不同大小的锚框,提高了对多尺寸遮挡人脸的分层识别效果;(3)通过融合注意力损失函数、分类损失函数和回归损失函数,提出了一种新的多任务损失函数,提升了测试阶段遮挡人脸检测的精确性。

1 SSD单阶段人脸检测模型

作为经典的单阶段目标检测算法之一,SSD算法不需要区域建议即可得到一系列候选框集合和框内目标的类别得分,经非极大值抑制后输出最终的检测结果。该算法由一个提取基础特征的主网络和一系列预测目标类别及位置的检测网络构成,检测网络连接在主网络生成的多尺度特征图后实现对多尺度目标的检测。受SSD算法的启发,基于SSD算法的人脸检测方法得到了快速发展,一种通用的单阶段人脸检测模型如图1所示。

由图1可知,该模型的输入是包含单个人脸或者多个人脸的图像,尺寸为300×300×3。在特征提取阶段,主网络采用VGG16的前5个卷积层,并将VGG16的2个全连接层转换为卷积层Conv6和Conv7,然后又连接了从Conv8_2到Conv10_2的3个卷积层,最后经过全局平均池化输出1×1×256的特征图。在目标预测阶段,检测网络分别从卷积层Conv4_3、Conv7、Conv8_2、Conv9_2、Conv10_2和Conv11_2提取特征图进行人脸/背景分类和人脸边界框回归,经过非极大值抑制(Non-maximum suppression, NMS)后,输出最终的多尺度检测人脸。

SSD单阶段人脸检测模型的损失函数包含人脸/背景分类损失L_{cls}和人脸边界框回归损失L_{reg},表

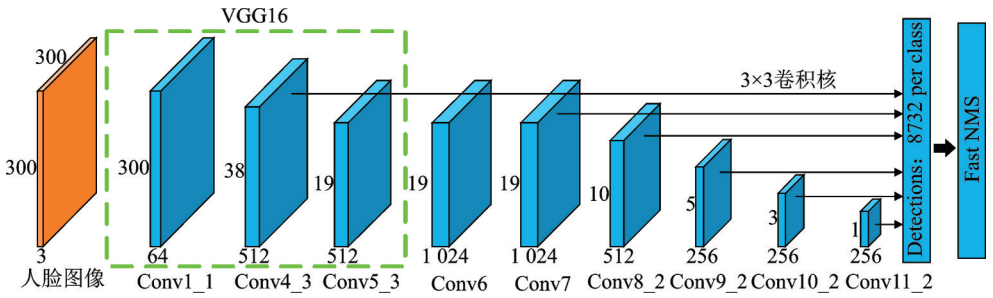

图1 SSD单阶段人脸检测模型

Fig.1 SSD single-stage face detection model

达式为

$$L(x,c,t,t^*)=\frac{1}{N}(L_{\text{cls}}(x,c)+\lambda L_{\text{reg}}(x,t,t^*)) \tag{1}$$

式中:N表示匹配到人脸真实框的锚框数量,在训练时选择与人脸真实框交并比大于0.5的锚框为正样本,选择与人脸真实框交并比小于0.5的锚框为负样本;x表示锚框匹配人脸真实框的标签,是人脸取1,是背景取0;c表示锚框区域属于人脸的预测概率;t和t^*分别表示人脸预测框和真实框,λ表示平衡2种损失的权重系数。

2 本文模型

2.1 模型框架

检测局部遮挡人脸时,受遮挡的影响,人脸遮挡区域的特征元素会损坏,这时如果均匀地提取特征元素会出现偏差,造成人脸检测精度下降。此时,可以通过充分参考人脸未被遮挡的检测区域来辅助推断遮挡区域属于真实人脸的概率,即增强人脸可见区域的响应值。此外,卷积深度神经网络在不同的特征层含有层次性的结构分辨率和差异性的语义信息。其中,浅层特征图的空间分辨率较高,利于检测小尺寸遮挡人脸,深层特征图的语义信息丰富,利于检测大尺寸遮挡人脸,所以还需要合理地设置锚框尺寸以充分学习不同特征层的人脸特征。基于上述考虑,本文提出了一种基于层级注意力增进网络的多尺寸遮挡人脸检测方法,对应的网络模型如图2所示。

由图2可知,基于层级注意力增进网络的多尺寸遮挡人脸检测模型由基础网络、注意力增进网络和并行检测网络3部分组成。基础网络在SSD单阶段人脸检测模型的Conv3_3、Conv4_3、Conv5_3、Conv8_2、Conv9_2和Conv10_2卷积层提取多个尺寸的初始特征图F_1、F_2、F_3、F_4、F_5和F_6。由图2可知,输入一幅尺寸为640×640的人脸图像,经过注意力增进网络后,F_1~F_6的尺寸分别为:160×160、80×80、40×40、20×20、10×10和5×5。通过反复迭代优化注意力增进网络的损失函数,逐步提升初始特征图中人脸未被遮挡部分的特征响应,得到F_1~F_6的增强特征图SF_1~SF_6。并行检测网络包含并连的人脸分类网络与人脸回归网络,人脸分类网络和人脸回归网络分别在增强特征图SF_k(k=1,2,…,6)上进行人脸/背景分类和人脸边界框回归。最终对检测出的众多人脸边界框执行非极大值抑制处理,输出精确的遮挡人脸检测结果。

2.2 注意力增进网络

在现实场景中,出现在面部区域的遮挡均会造成人脸区域特征元素损坏,降低人脸检测精度。因此,在检测遮挡人脸时,有必要增强人脸可见区域的响应值来辅助确认该检测区域是否包含人脸。本文提出了可提高人脸可见区域特征响应的注意力增进网络,其网络结构如图3所示。之所以把注意力增进网络的关注点放在人脸未被遮挡区域而不是人脸遮挡区域,是因为未被遮挡区域一般都是真实人

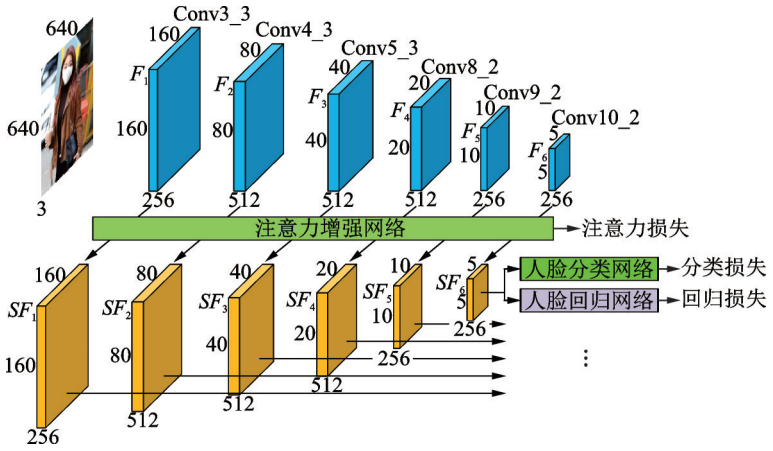

图2 基于层级注意力增进网络的多尺寸遮挡人脸检测模型

Fig.2 Multi-size occlusion face detection model based on hierarchical attention enhancement network

脸，虽然千人千面，但是不同人脸的深层特征是趋于相近的。而遮挡区域可能是帽子、口罩、围巾、衣服和水杯等物品，它们的深层特征是趋于不同的，因此选择人脸可见区域进行特征增强。

由图3可知，注意力增进网络的输入是初始特征图F_k，尺寸为$l_k\times l_k\times w_k$，F_k经过4个级联的卷积层后，输出尺寸为$l_k\times l_k\times 1$的特征图C_k。特征图C_k中的每个元素都是一个二分类结果，表示该元素属于人脸的概率，将每个元素的概率与阈值T_F比较，大于T_F的元素置1，小于T_F的元素置0，得到得分图C'_k。然后对得分图C'_k进行e指数提升，将置0的元素转换为$e^0=1$，将置1的元素转换为$e^1=e$。接着将转换后的得分图与初始特征图F_k的各个通道进行逐元素点乘，得到初始特征图F_k的增强特征图SF_k。

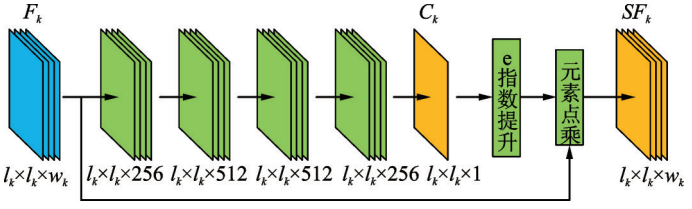

图3 注意力增进网络

Fig.3 Attention enhancement network

在注意力增进网络的训练阶段，需要为人脸区域设置标记，通常人脸数据集中仅标记了矩形框，因此在网络训练时要将矩形框内的人脸全部置1，矩形框外的背景全部置0。此时标记为1的区域仍然存在遮挡，但是在大量数据集训练下，人脸可见部分（即人脸未被遮挡部分）的特征响应是趋于统计集中的，这样通过大量训练样本学习出的网络参数，可以增强深层特征图中人脸可见部分的响应值。在测试阶段，经过训练的注意力增进网络会提升人脸可见部分的响应值。

2.3 锚框设置

文献[17]对WIDER FACE数据集中的人脸真实标注进行统计后发现，80%的人脸尺寸处于区间[16, 406]。由于小尺寸人脸的特征描述在浅层特征图较为丰富，大尺寸人脸的特征描述在深层特征图较为丰富，因此在不同特征图上层次性的设计多尺寸锚框，首先计算每个特征层的锚框尺寸与输入图像尺寸之比为

$$\begin{aligned} s_{\min}^{k} &= s_{\max}\left(\frac{k+1}{8}\right)^{\frac{5}{2}} \quad k\in[1,\ 6] \\ s_{\max}^{k} &= s_{\max}\left(\frac{k+2}{8}\right)^{\frac{5}{2}} \quad k\in[1,\ 6] \end{aligned} \tag{2}$$

式中:k表示特征图的层数;$s_{\max}$表示锚框尺寸与输入图像尺寸的最大比例;$s_{\min}^{k}$表示第k层特征图对应的最小锚框比例;$s_{\max}^{k}$表示第k层特征图对应的最大锚框比例,本文令$s_{\max}=0.72$。

接着,利用第k层特征图对应的锚框比例,为第k层特征图设计4种尺寸的锚框为

$$\begin{aligned} &\text{Anchor}_1^k:\ 640s_{\min}^{k}\times 640s_{\min}^{k} \\ &\text{Anchor}_2^k:\ 640s_{\min}^{k}{}^{\frac{3}{4}}s_{\max}^{k}{}^{\frac{1}{4}}\times 640s_{\min}^{k}{}^{\frac{3}{4}}s_{\max}^{k}{}^{\frac{1}{4}} \\ &\text{Anchor}_3^k:\ 640s_{\min}^{k}{}^{\frac{1}{2}}s_{\max}^{k}{}^{\frac{1}{2}}\times 640s_{\min}^{k}{}^{\frac{1}{2}}s_{\max}^{k}{}^{\frac{1}{2}} \\ &\text{Anchor}_4^k:\ 640s_{\min}^{k}{}^{\frac{1}{4}}s_{\max}^{k}{}^{\frac{3}{4}}\times 640s_{\min}^{k}{}^{\frac{1}{4}}s_{\max}^{k}{}^{\frac{3}{4}} \end{aligned} \tag{3}$$

通过式(3)的锚框设置方法可以确保每个人脸真实标注都能匹配到并交比(Intersection-over-union, IoU)≥0.6的锚框。在训练时,如果锚框与人脸真实标注的IoU≥0.55,则定义该锚框为正样本,如果锚框与人脸真实标注的IoU≤0.4,则定义该锚框为背景。

2.4 损失函数设置

本文为注意力增进网络、人脸分类网络和人脸回归网络设置了不同的损失函数。在模型训练时通过联合3种损失函数共同优化网络参数。

(1) 在注意力增进网络中,通过逐元素sigmoid交叉熵损失函数来增强人脸可见区域,表达式为

$$L_{\text{atten}}(p,\ p^*)=\frac{1}{N_{\text{A}}}\sum_{i=1}^{N_{\text{A}}}p_i^*\log p_i+(1-p_i^*)\log(1-p_i) \tag{4}$$

式中:p表示输入锚框内的特征元素属于人脸的概率;p^*表示人脸区域的真实标记,它是通过对人脸标注框内区域全部置1,框外区域全部置0得到的;p_i表示锚框内第i个特征元素属于人脸的概率;p_i^*表示锚框内第i个特征元素的真实标记,属于人脸为1,属于背景为0;N_{A}表示输入锚框的特征元素数值。

(2) 在人脸分类网络中,通过softmax交叉熵损失函数区分人脸和背景,函数表达式为

$$L_{\text{cls}}(y,\ y^*)=y^*\log y+(1-y^*)\log(1-y) \tag{5}$$

式中:y表示输入锚框预测为人脸概率;y^*表示人脸真实标记,当输入锚框是人脸时y^*为1,否则y^*为0。

(3) 在人脸回归网络中,通过smooth L_1损失函数进行边框回归,函数表达式为

$$L_{\text{reg}}(\boldsymbol{t},\boldsymbol{t}^*)=\sum_{j\in\{x,\ y,\ w,\ h\}}\begin{cases}0.5(t_j-t_j^*)^2 & |t_j-t_j^*|<1\\ |t_j-t_j^*|-0.5 & \text{其他}\end{cases} \tag{6}$$

式中:$\boldsymbol{t}=(t_x,\ t_y,\ t_w,\ t_h)$和$\boldsymbol{t}^*=(t_x^*,\ t_y^*,\ t_w^*,\ t_h^*)$均为1×4的向量,$\boldsymbol{t}$表示预测人脸框坐标,$\boldsymbol{t}^*$表示真实人脸框坐标。

在模型训练时,通过将注意力损失函数L_{atten}、人脸分类损失函数L_{cls}和人脸回归损失函数L_{reg}加权求和,得到多任务损失函数L_{acr},共同优化网络参数,表达式为

$$\begin{aligned} L_{\text{acr}}(p,y,\boldsymbol{t},p^*,y^*,\boldsymbol{t}^*)=&\frac{1}{N_{p\wedge n}}\sum_{k=1}^{N_{p\wedge n}}L_{\text{atten}}(p,\ p^*)+\\ &\frac{\alpha}{N_{p\wedge n}}\sum_{k=1}^{N_{p\wedge n}}L_{\text{cls}}(y,\ y^*)+\frac{\beta}{N_p}\sum_{l=1}^{N_p}L_{\text{reg}}(\boldsymbol{t},\boldsymbol{t}^*) \end{aligned} \tag{7}$$

式中：$N_{p\wedge n}$表示正负样本锚框数量；N_p表示正样本锚框数量；α和β为平衡L_{atten}、L_{cls}和L_{reg}的权重系数。

3 实验验证

为验证本文方法的检测性能，在WIDER FACE[18]人脸数据集和MAFA遮挡人脸数据集[19]上选择Mask Net[14]，SFD[13]，NMR[12]和RetinaFace[16]方法进行对比实验。所有模型均采用随机梯度下降(Stochastic gradient descent, SGD)在4个GPU上进行训练，批大小设置为16，权重衰减设置为0.000 5，动量设置为0.9，前8×10^4次迭代，学习率为10^{-3}，后2×10^4次迭代，学习率减小为10^{-4}。为了平衡本文模型在速度和精度指标上的表现，将式(7)损失函数的权重系数设置为$\alpha=1$，$\beta=1.5$。另外，为保证正负训练样本的平衡性，通过控制负样本锚框数量使正样本锚框数量达到负样本锚框数量的1/2。

3.1 WIDER FACE人脸数据集对比实验

WIDER FACE人脸数据集具有32 203幅包含遮挡、模糊、多尺度、多光照和多姿态变化的人脸图像，其中标记了393 703个人脸矩形框。该数据集按照4∶1∶5的比例分为训练集、验证集和测试集。此外，根据61种场景分类中人脸检测的难易不同，数据集还可分为简单、中等和困难3个等级。为验证本文方法在现实场景中的人脸检测效果，在WIDER FACE通用评价标准下进行对比测试。

通过统计本文方法与Mask Net，SFD，NMR和RetinaFace方法在WIDER FACE简单、中等和困难3个难度等级上的实验结果，得到平均精度均值(Mean average precision，MAP)和检测速度对比结果如表1所示，PR(Precision-recall)曲线对比如图4所示。由表1可知，本文方法在WIDER FACE简单、中等和困难3个子测试集上的平均精度均高于其他4种方法，并且本文方法的检测速度也仅比NMR方法慢2帧/s，具有较好的实时性。观察图4中不同方法的PR曲线可得，在简单、中等和困难3个等级的测试中，本文方法的PR曲线都高于其他4种方法，证明了本文方法的检测性能优于其他4种人脸检测方法。由于WIDER FACE困难子测试集中具有大量的遮挡人脸，而本文方法在该测试集中取得了最优的检测效果，证明了本文所提出遮挡人脸检测方法的有效性。

表1 平均精度和检测速度对比结果

Table 1 Comparison results of MAP and detection speed

方法	MAP			速度/(帧·s^{-1})
	简单	中等	困难	
Mask Net[14]	0.904	0.891	0.813	18
SFD[13]	0.934	0.923	0.858	23
NMR[12]	0.947	0.932	0.863	34
RetinaFace[16]	0.962	0.947	0.896	21
本文方法	0.969	0.954	0.908	32

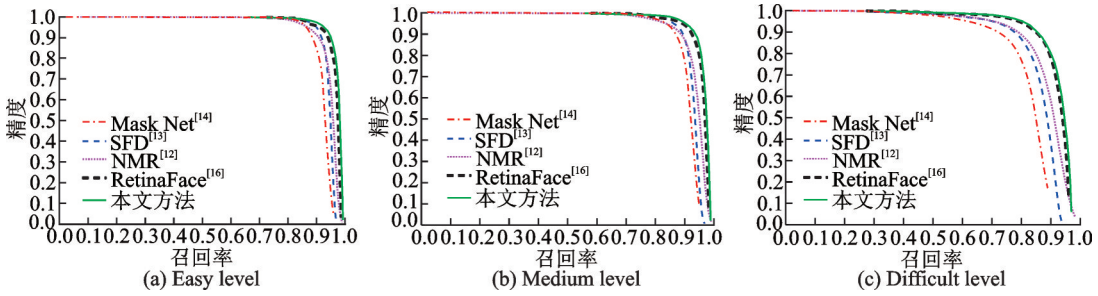

图4 PR曲线对比

Fig.4 Precision-recall curve comparison

3.2 MAFA遮挡人脸数据集对比实验

MAFA遮挡人脸数据集在30 811幅包含多种遮挡场景和遮罩类型的人脸图像中标记了35 806个人脸矩形框。其中25 876幅图像(标记29 452个遮挡人脸)为训练集,4 935幅图像(标记6 354个遮挡人脸)为测试集。该数据集为每个遮挡人脸赋予了6种属性,分别是人脸位置、眼睛位置、遮罩位置、人脸方向、遮挡级别和遮罩类型。为验证本文方法对遮挡人脸的检测效果,在MAFA遮挡评价指标下进行对比测试。

本文方法与Mask Net,SFD,NMR和RetinaFace方法在MAFA测试集的平均精度(Average precision, AP)对比结果如表2所示。在表2中,前5种属性对应人脸的5个方向,分别是左边脸、左前脸、正脸、右前脸和右边脸,随着人脸偏转角度增加,5种方法的平均精度均下降明显,但本文方法取得了最高的检测精度;第6~8种属性对应3个遮挡级别,分别是较弱、中等和较强,随着遮挡级别提高,5种方法的平均精度均有下降,而本文方法的下降范围相对较小,说明本文方法具有更强的抗遮挡能力;第9~12种属性对应4种遮罩类型,分别是简单遮罩、复杂遮罩、人体自遮罩和混合遮罩,本文方法在这4种类型的遮挡下均取得最优的人脸检测结果。综合所有属性下的遮挡人脸检测结果,本文方法的平均精度达到了81.8%,比次优的RetinaFace方法精度提升约8%,表明本文方法的检测精度高于目前主流的遮挡人脸检测方法。本文方法与Mask Net,SFD,NMR和RetinaFace等方法在MAFA测试集的部分对比结果如图5所示。

表2 平均精度对比结果

Table 2 Average precision comparison %

编号	属性	Mask Net[14]	SFD[13]	NMR[12]	Retina Face[16]	本文方法
1	Left	12.60	14.10	14.90	19.5	21.7
2	Left-front	35.80	37.20	38.20	48.0	65.0
3	Front	72.20	72.40	74.30	76.9	83.7
4	Right-front	26.10	28.50	28.50	33.3	59.9
5	Right	8.15	9.78	10.25	18.4	19.5
6	Weak	65.40	64.80	68.00	69.5	80.6
7	Medium	42.10	53.30	44.40	58.4	74.1
8	Heavy	12.50	14.40	14.90	17.8	27.7
9	Simple	61.80	59.40	63.90	62.8	78.8
10	Complex	54.00	56.60	56.40	60.7	76.2
11	Body	30.80	37.50	32.90	43.1	68.8
12	Hybrid	13.10	16.20	15.70	19.1	29.1
综合		**67.60**	**68.10**	**70.10**	**73.8**	**81.8**

3.3 自对比实验

为证明本文所提出方法的精确性和有效性。在MAFA遮挡人脸数据集上进行自对比实验。令SSD为基准线方法,SSD+AEN表示在SSD的基础上增加了注意力增进网络,SSD+AEN+MAS表示在SSD的基础上增加了注意力增进网络和多锚框设置,SSD+AEN+MAS+ML表示在SSD的基础上增加了注意力增进网络、多锚框设置和多任务损失函数。上述方法在MAFA测试集的自对比实验结果如表3所示。由表3可以看出,通过引入注意力增进网络,可以显著提高对多尺寸遮挡人脸的检测

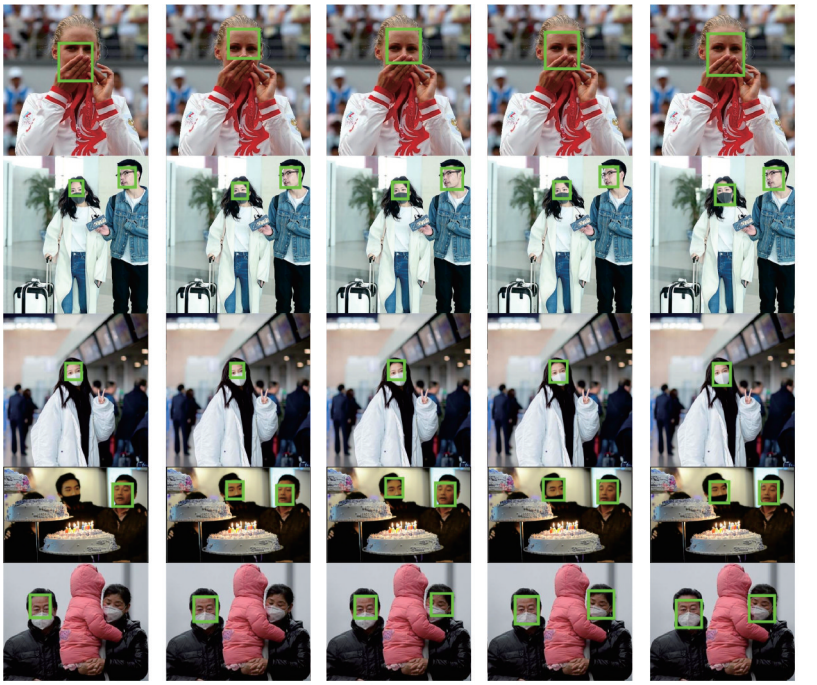

图5 MAFA测试集部分对比结果

Fig.5 Partial comparison results of MAFA test set

精度。多锚框设置和多任务损失函数充分参考了人脸周围的上下文信息,使得人脸可见区域特征学习更加全面,进一步提升了多尺寸遮挡人脸的检测精度。

表3 自对比实验结果

Table 3 Self comparison experiment results %

属性	平均精度
SSD	68.4
SSD+AEN	75.1
SSD+AEN+MAS	78.5
SSD+AEN+MAS+ML	80.7

4 结束语

本文提出一种基于层级注意力增进网络的多尺寸遮挡人脸检测方法,通过在SSD单阶段人脸检测网络的多个初始特征图上引入注意力增进机制,提升了人脸可见区域的响应值。同时,为不同增强特征层设计不同大小的锚框,提高了对多尺寸遮挡人脸的分层识别效果。此外,在训练时还将注意力损失函数、分类损失函数和回归损失函数融合为多任务损失函数,共同优化网络参数。在WIDER FACE和MAFA数据集的实验结果证明,本文方法的检测精度高于目前主流的遮挡人脸检测方法。

参考文献:

[1] 谭小慧, 李昭伟, 樊亚春. 基于多尺度细节增强的面部表情识别方法[J]. 电子与信息学报, 2019, 41(11): 2752-2759.
TAN Xiaohui, LI Zhaowei, FAN Yachun. Facial expression recognition method based on multi-scale detail enhancement[J]. Journal of Electronics and Information Technology, 2019, 41(11): 2752-2759.

[2] LOU J, CAI X, WANG Y, et al. Multi-subspace supervised descent method for robust face alignment[J]. Multimedia Tools and Applications, 2019, 78(24): 35455-35469.

[3] ROTGER M, MORENO-NOGUER F, LUMBRERAS F, et al. Detailed 3D face reconstruction from a single RGB image [J]. Journal of WSCG, 2019, 27(2): 103-112.

[4] DENG W, HU J, GUO J. Face recognition via collaborative representation: Its discriminant nature and superposed representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(10): 2513-2521.

[5] WRIGHT J, YANG A Y, GANESH A, et al. Robust face recognition via sparse representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(2): 210-227.

[6] OH H J, LEE K M, LEE S U. Occlusion invariant face recognition using selective local non-negative matrix factorization basis images[J]. Image and Vision Computing, 2008, 26(11): 1515-1523.

[7] PARK S, LEE H, YOO J H, et al. Partially occluded facial image retrieval based on a similarity measurement[J]. Mathematical Problems in Engineering, 2015, 2015(1): 1-11.

[8] LI X X, LIANG R H. A review for face recognition with occlusion: From subspace regression to deep learning[J]. Chinese Journal of Computers, 2018, 41(1): 177-207.

[9] XIAO Y, CAO D, GAO L. Face detection based on occlusion area detection and recovery[J]. Multimedia Tools and Applications, 2019, 1(2): 1-16.

[10] ZHAO F, FENG J, ZHAO J, et al. Robust LSTM-autoencoders for face de-occlusion in the wild[J]. IEEE Transactions on Image Processing, 2018, 27(2): 778-790.

[11] DENG J, TRIGEORGIS G, ZHOU Y, et al. Joint multi-view face alignment in the wild[J]. IEEE Transactions on Image Processing, 2019, 28(7): 3636-3648.

[12] TRIGUEROS D S, MENG L, HARTNETT M. Enhancing convolutional neural networks for face recognition with occlusion maps and batch triplet loss[J]. Image and Vision Computing, 2018, 79: 99-108.

[13] ZHANG S, ZHU X, LEI Z, et al. S^3FD: Single shot scale-invariant face detector[C]//Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE Computer Society, 2017: 192-201.

[14] WAN W, CHEN J. Occlusion robust face recognition based on mask learning[C]// Proceedings of IEEE International Conference on Image Processing. Beijing, China: IEEE Computer Society, 2017: 3795-3799.

[15] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]//Proceedings of the European Conference on Computer Vision. Amsterdam: Springer, 2016: 21-37.

[16] DENG J, GUO J, ZHOU Y, et al. RetinaFace: Single-stage dense face localisation in the wild[EB/OL]. [2020-2-10](2021-02-10). https://arxiv.org/abs/1905.00641.

[17] WANG J, YUAN Y, YU G. Face attention network: An effective face detector for the occluded faces[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE Computer Society, 2017: 5021-5033.

[18] YANG S, LUO P, CHEN C L, et al. WIDER FACE: A face detection benchmark[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE Computer Society, 2016: 5525-5533.

[19] GE S, LI J, YE Q, et al. Detecting masked faces in the wild with LLE-CNNs[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE Computer Society, 2017: 2682-2690.

作者简介:

王麟阁(1979-),**通信作者**,男,硕士,讲师,高级工程师,研究方向:人工智能、无线传感器网络,E-mail:wanglinge1979@163.com。

蒋宝军(1979-),男,博士,副教授,研究方向:人工智能、智慧城市。

潘铁军(1972-),男,博士,教授,研究方向:人工智能。

(编辑:张黄群)