

融合注意力机制的人脸识别算法研究

张晋婧¹ 刘双峰² 丰雷² 张瑜²

(1. 中北大学电气与控制工程学院 太原 030051; 2. 中北大学仪器科学与动态测试教育部重点实验室 太原 030051)

摘要: FaceNet 人脸识别算法是目前较为主流的人脸识别算法,其运行速度快被广泛应用于工业界。针对 FaceNet 人脸识别网络存在对面部遮挡人脸识别精度低的问题,提出了一种融合注意力机制的 FaceNet 人脸识别算法。该算法在 FaceNet 的基础上引入 GhostNet 特征提取网络对人脸更好的提取人脸特征,并融合注意力机制与特征金字塔(feature pyramid networks, FPN)加强特征提取网络实现对 3 种尺度特征图中局部信息的放大,加强不同感受野下的特征提取,增强较为重要的特征信息。实验结果表明,提出的人脸识别算法取得了良好的识别效果,在人脸数据集(LWF)下准确率达到 99.62%。对有遮挡的面部识别也取得了较好的检测结果,可准确识别存在遮挡的人脸目标。

关键词: FaceNet; GhostNet; 注意力机制; 人脸识别; 面部遮挡

中图分类号: TP242.6 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Research on face recognition algorithm integrating attention mechanism

Zhang Jinjing¹ Liu Shuangfeng² Feng Lei² Zhang Yu²

(1. College of Electrical and Control Engineering, North University of China, Taiyuan 030051, China; 2. Key Laboratory of Instrument Science and Dynamic Testing, Ministry of Education, North University of China, Taiyuan 030051, China)

Abstract: FaceNet face recognition algorithm is currently the mainstream face recognition algorithm, its fast running speed is widely used in the industry. FaceNet face recognition network has the problem of low accuracy in face occlusion face recognition. This paper presents a FaceNet face recognition algorithm that combines attention mechanism. Based on FaceNet, GhostNet feature extraction network is introduced to extract face features better. Feature pyramid networks (FPN) is combined to enhance feature extraction network to enlarge local information in three scale feature maps, enhance feature extraction under different perception fields and enhance more important feature information. The experimental results show that the face recognition algorithm in this paper achieves good recognition results, and the accuracy reaches 99.62% in the face dataset(LWF). Better detection results are also obtained for face recognition with occlusion, which can accurately identify face targets with occlusion.

Keywords: FaceNet; GhostNet; attention mechanism; face recognition; facial occlusion

0 引言

人脸识别是基于人脸部特征信息进行身份识别的一种生物识别技术^[1],在安防、金融、交通、教育、警务等诸多场景应用广泛,呈现出显著的价值^[2]。人脸识别技术具有的非接触性、安全可靠性和便捷性等优势,使其在新冠疫情肆虐的当下被广泛应用于各行各业,其中作为火车站等交通公共场所对无接触式人物识别应用尤为广泛^[2-3]。如何更加精准的识别目标对象成为了近年来该领域较为火

热的研究方向^[4]。

国内外专家学者提出了大量人脸识别算法,大致可分为基于可见光图像人脸识别方法和基于三维重建人脸识别方法。三维图像人脸识别算法将人脸分为使用相机和激光雷达两种方式^[5],两种方式均需使用三维数据;使用相机拍摄图像,将图像转换为稠密点云信息对目标重建得到人脸面部模型后识别,文献[6]从二维图像构建三维样本,提取图像中人脸层次化特征,构建人脸曲面,在三维曲面中完成人脸识别。使用激光雷达扫描目标获取稀疏点

云信息,对目标重建得到面部模型后识别,文献[7]使用三维激光雷达得到三维数据,通过主成分分析(principal component analysis,PCA)识别目标,文献[8]使用三维数据创建模板匹配方法结合图像进行特征点识别,综合分析特征点的性质完成人脸识别。基于三维重建的方法可精准的还原被检测目标的面部信息,尤其对面部纹理信息有着较好的表达便于后期特征提取及学习,但该方法存在一定的缺陷,三维激光雷达数据量庞大,使用时存在大量非目标点无法精准定位目标不利于重建工作,图像转稠密点云同样在重建工作中存在该问题。基于可见光图像人脸识别方法相较于三维重建方法避免了图像与点云之间的转换,降低了算法的复杂度,在实时性上明显优于三维重建方法。图像人脸识别方法一般使用基于深度学习的方法,主要流程包括人脸预处理(检测、对齐、标准化、数据增强等)、特征学习、特征对比等步骤^[9]。文献[10]首次将深度卷积网络引入人脸识别网络取得了较好的识别效果,文献[11]利用多个深度网络提取特征后拼接特征使用PCA得到更有效的特征进行人脸识别,这种方法之后被VGG-Net^[12]、GoogLeNet^[13]、ResNet^[14]、FaceNet^[15]等深度网络所取代。

本文基于FaceNet和GhostNet^[16-17]模型设计了一种人脸识别算法。该算法在FaceNet中利用GhostNet模型对人脸图片特征提取,引入注意力机制^[18]聚焦于局部信息,使网络学习图像中更为重要的特征,达到抑制其他不重要特征的目的,同时增加了特征金字塔(FPN)加强特征提取结构,融合低层细节信息和高层语义信息,增大低层感受野更加有效的提取细节信息。经实验验证,本文算法在满足实时性的同时提高了人物识别的准确率,具有较强的鲁棒性,在LFW数据集上准确率达到99.62%。

1 相关工作

1.1 FaceNet 算法

FaceNet网络采用深度卷积神经网络ResNet作为骨干网络,将图像特征映射到多维空间,计算欧氏距离判断人脸的相似程度。欧氏距离直接和图片相似度相关,同一个人的不同图像在空间距离很小,不同人的图像在空间中有较大的距离,通过设置特征向量间欧氏距离阈值可区分不同人脸。FaceNet网络结构可分为3部分如图1所示,第1部分将图片送入深度网络使用残差网络(ResNet)结构加深网络增加特征能力,通过池化得到128维特征长条;第2部分使用L2标准化;第3部分设置阈值利用欧氏距离进行对比识别人脸。

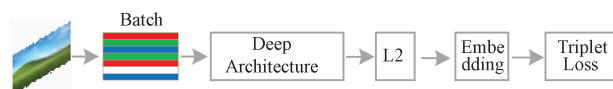


图1 FaceNet 结构

FaceNet使用了三元损失 Triplet Loss 通过设置阈值

控制正负样本间的距离,Triplet Loss 可以很好的对细节进行区分,当两个输入很相似的时候,Triplet Loss 对这两个差异性较小的输入向量可以学习到更好的表示同时也可以学习到更加细微的特征。如图2所示,每次选择3个样本计算 Loss,一个是目标样本记做 Anchor,一个是和目标为同一个人的正样本记做 Positive,另一个是其他人的负样本记做 Negative,选择 Anchor 到 Positive 的距离大于 Anchor 到 Negative 的,通过模型的不断训练,将他们调节为 Anchor 到 Positive 的距离小于 Anchor 到 Negative。

Triplet Loss 损失函数如式(1)。

$$\sum_i^N [\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha] \quad (1)$$

式中: x_i^a 表示 Anchor 图像; x_i^p 表示 Positive 图像; x_i^n 表示 Negative 图像; α 表示间隔值。

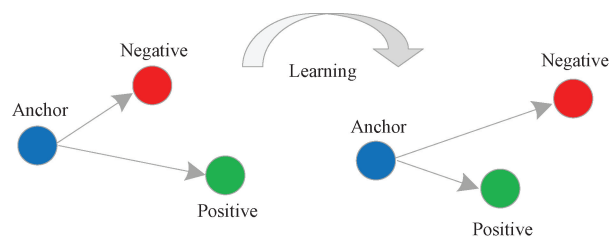


图2 Triplet Loss 损失函数

1.2 FaceNet 算法优化

1) GhostNet 模块

在卷积神经网络(CNN)模型中,特征图的冗余非常重要,一组特征图中存在冗余特征图,可保证在良好检测效果情况下减少模型的参数量,提高模型的执行速度。GhostNet 通过使用低计算量的操作生成冗余特征图,来增强网络的学习能力。引入 GhostNet 模块将极大改善 ResNet 模块相同 stage 之间不同层的输出特征相似性高,不同 stage 之间特征相似性较低引起的学习能力差的问题,同时提高了运行速度。

本文使用的模块可分为两部分,第1部分是 Ghost Module 模块,在网络中将使用该模块替代普通 3×3 卷积操作。Ghost Module 模块如图3所示,首先进行 1×1 卷积降低通道数整合特征,生成浓缩特征,再使用深度可分离逐层卷积,生成特征图。第2部分 Ghost Bottlenecks 模块如图4所示,这部分同样也是用了残差结构,残差结构将 Ghost Bottlenecks 分为主干部分和残差部分,主干部分本质是使用 Ghost Module 代替 3×3 卷积,同时增加一个 2×2 的逐层卷积压缩特征层的宽高,残差边部分增加 2×2 的逐层卷积和 1×1 卷积。

2) 注意力机制模块

注意力机制是一种简单而有效的模块可以在通道和空间维度上进行 Attention,其核心是让网络关注它更需

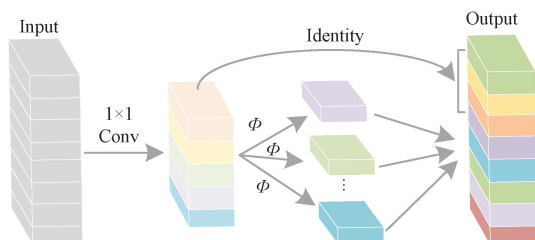


图3 Ghost Module 模块

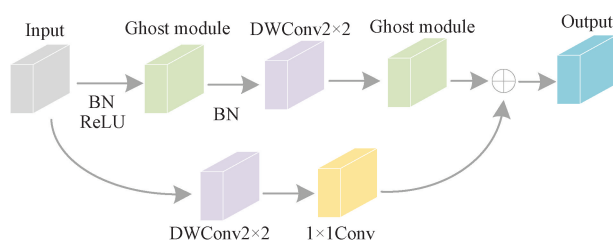


图4 Ghost Bottlenecks 模块

要关注的地方。本文借鉴基于卷积块的注意力机制(convolutional block attention module,CBAM),通过空间和通道两种注意力模型,使网络聚焦于局部信息,学习图像中更为重要的特征从而达到抑制其他不重要的特征。

特征图的一个维度是图像的尺度空间长和宽,另外一个维度是通道,从通道位置关注更高语义特征就是通道注意力机制。对输入进来的单个特征层,分别进行全局平均池化和全局最大池化。之后对平均池化 AvgPool 和最大池化 MaxPool 的结果利用共享的全连接层 MLP 进行处理,处理后的两个结果进行相加,然后取 sigmoid(σ),将得到的结果与原特征相乘,如图5所示。

$$M_c(F) = \sigma \{ \text{MLP} [\text{AvgPool}(F) + \text{MaxPool}(F)] \} = \sigma \{ W_1 [W_0 (F_{\text{avg}}^c)] + W_1 [W_0 (F_{\text{max}}^c)] \} \quad (2)$$

式中: W_0 、 W_1 均为多层感知机的可学习参数; $W_0 \in \mathbf{R}^{(C/r) \times C}$, $W_1 \in \mathbf{R}^{C \times (C/r)}$ 。

特征图中所有区域对于任务的重要性是同样重要的,

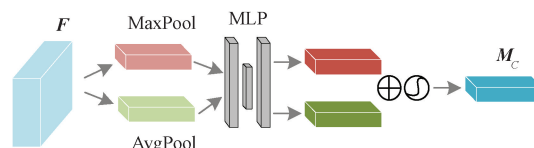


图5 通道注意力机制

只是任务相关区域的才是最需要关心的,空间注意力机制就是寻找网络中最重要的位置进行处理。对输入进来的特征层,在每一个特征点的通道上取最大值和平均值。之后将这两个结果进行堆叠,利用一次通道数为1的卷积调整通道数,然后取 sigmoid,将得到的结果与原特征相乘,如图6所示。

$$M_s(F) = \sigma \{ f^{7 \times 7} [\text{AvgPool}(F); \text{MaxPool}(F)] \} = \sigma \{ f^{7 \times 7} [F_{\text{avg}}^s; F_{\text{max}}^s] \} \quad (3)$$

式中: $f^{7 \times 7}$ 为 7×7 的卷积核卷积操作。

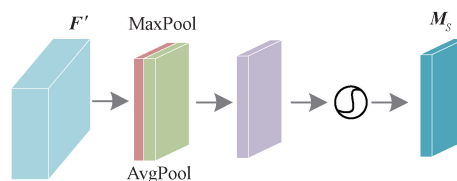


图6 空间注意力机制

并联注意力机制同时通过通道注意力机制和通过空间注意力机制,使其形成并联过程进行特征提取和学习。假设输入特征图 $F \in \mathbf{R}^{C \times H \times W}$,将 F 分为通道注意力 $M_c \in \mathbf{R}^{C \times 1 \times 1}$ 和空间注意力 $M_s \in \mathbf{R}^{1 \times H \times W}$ 。整个注意力处理过程如下:

$$F'_c = M_c(F) \otimes F \quad (6)$$

$$F'_s = M_s(F) \otimes F \quad (7)$$

$$F'' = F'_c + F'_s \quad (8)$$

式中: F'_c 为经过通道注意力处理后的数据; F'_s 为经过空间注意力处理后的数据; F'' 为经过通道和空间注意力处理后的数据。

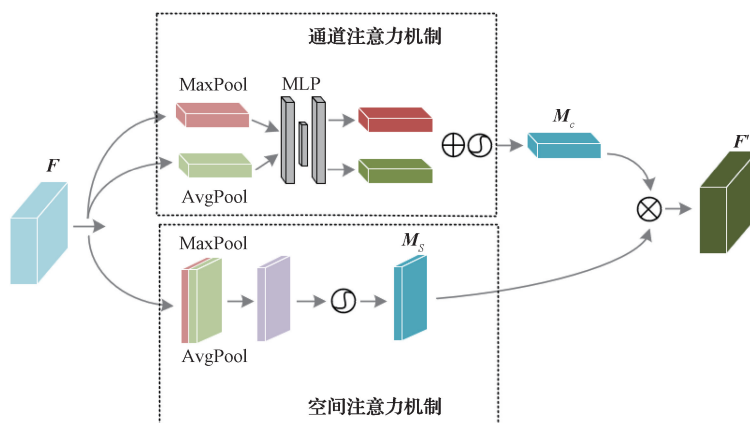


图7 并联结构

3) FPN 加强特征提取网络

FPN 结构是一种自上而下的网络结构如 8 所示,本文引入 FPN 模块构建不同尺寸具有高级语义的特征图,通过特征图金字塔结构融合特征提取网络的输出层获得

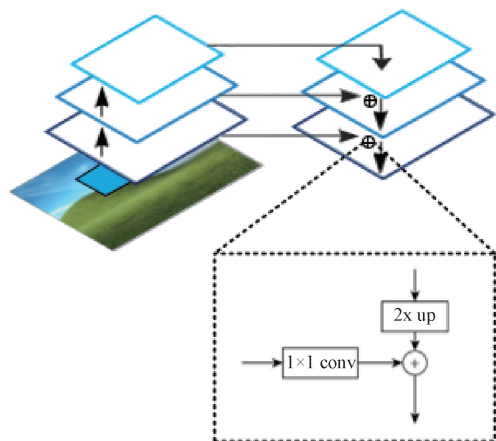


图 8 FPN 结构

多尺度信息,在不增加计算量的情况下,增强特征提取,增加网络深度提升网络的准确性。

2 网络总体模型设计

本文使用模型如图 9 所示,在 FcaeNet 的基础上使用 GhostNet 特征提取网络替代原模型中 ResNet 模块。GhostNet 模型中自带 SENet 注意力机制,该注意力机制仅为通道注意力,在人脸识别应用中更加注重对空间特征的提取,提取到更有价值的纹理信息、边缘信息将大大增加网络的性能,因为在此处将原有的注意力改进为注意力机制,保持通道注意力不变,增加空间注意力,提升 GhostNet 性能。同时受到 FPN 结构启发,设计了类似于 FPN 加强特征提取模块,FPN 网络一般应用于目标检测网络中,多尺度特征输出后直接送入检测头进行多尺度检测。本文修改 FPN 网络,将其多尺度输出的特征进行融合,在自下而上的特征学习增强后,增加一条自下而上的特征融合,这样可在不同感受野下不同的重要特征在最终的特征图中都得到增强。

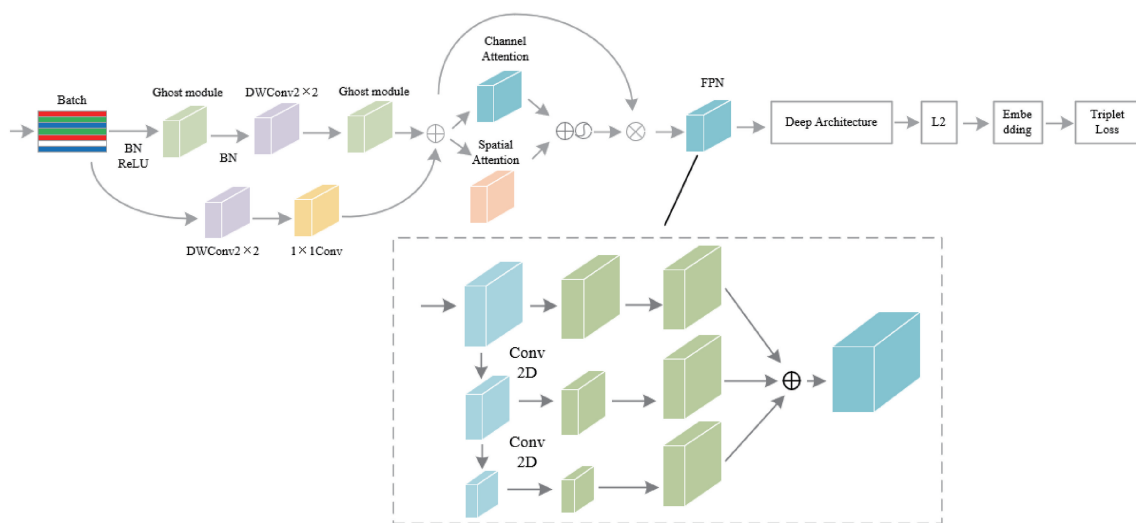


图 9 总体模型

3 实验结果分析

实验采用 CASIA-WebFace 数据集, LFW、CPLFW、CFP_FF 作为测试集,测试识别人脸的准确度。通过实验测试证明本文算法精度良好,通过与 ResNet50、MobileFaceNetV1 和 GhostFaceNet 网络对比表明本文算法优势较为明显。本文还进行了检测速度测试,其中 MobileFaceNetV1 识别速度为 23.8 ms, GhostFaceNet 识别速度为 15.5 ms, 本文识别速度为 20.2 ms, 略低于 GhostFaceNet。

如表 1 所示,与各网络相比, GhostFaceNet 的性能有所下降,尤其是在包含姿态变换的测试集中。 GhostFaceNet 的性能相比于表现最好的轻量级模型在 LFW、

CPLFW、CFP_FF 数据集分别下降了 0.1%、1.44%、和 0.45%,而本文方法在 3 种数据集表现优异,比 GhostFaceNet 性能分别上升了 0.09%、5.53%、0.46%。 GhostFaceNet 性能在 LFW 和 CFP_FF 测试集结果与其他方法相近,对于 CPLFW 测试集 GhostFaceNet 性能下降较为明显,主要原因是姿态变换需要更多的通道去表达丰富的特征信息,而本文方法中通过增加空间注意力和通道注意力,相比于其他模型能够在姿态变换中变换空间特征提取所需的特征信息,从而本文算法可弥补 GhostFaceNet 缺陷优化识别性能。本文方法是在 GhostFaceNet 改进,属轻量级网络,在不同数据集中均可以满足人脸识别技术对性能的要求。

为了进一步证明算法的鲁棒性,实验采用 CASIA-

表1 不同数据集图像训练结果 (%)

方法	LFW	CPLFW	CFP_FF
ResNet50	99.63	90.57	99.54
MobileFaceNetV1	99.52	87.03	99.13
GhostFaceNet	99.53	89.13	99.09
本文方法	99.62	94.66	99.55

WebFace 数据集,在 LFW、自制戴口罩 LFW 数据集 (MaskLFW)、戴口罩侧脸数据集 (side face mask, SFM)、戴眼镜人脸数据集 (face in sunglasses, FIG) 作为测试集,分别测试了正常人脸、佩戴口罩人脸、戴口罩侧脸的准确度。实验结果表明加入并联 CBAM 注意力机制相较于 GhostFaceNet 算法有了明显提升,且鲁棒性较强。图 10 所示为 MaskLFW 部分数据集图片,图 11 所示为不同角度及遮挡情况的可视化结果。



图10 MaskLFW 部分数据集图片

不同角度及遮挡情况图像训练结果如表 2 所示。由

表2 不同角度及遮挡情况图像训练结果

方法	LFW/%	MaskLFW/%	SFM/%	FIG/%	识别速度/ms
ResNet50	99.63	78.72	74.79	81.31	—
MobileFaceNetV1	99.52	88.52	86.90	92.13	23.8
GhostFaceNet	99.53	94.02	89.64	90.22	15.5
本文方法	99.62	98.71	94.66	90.85	20.2

3.1 注意力对照实验

为进一步验证注意力机制的性能,本文进行串联注意力机制与并联注意力机制对照实验。

串联注意力机制如图 12 所示,先通过通道注意力机



图11 可视化结果

表 2 可见,并联结构的注意力机制在 LFW 测试集下准确率达到 99.62%。表 2 中, GhostFaceNet 的性能相比于表现最好的轻量级模型在 LFW、MaskLFW、SFM、FIG 数据集分别上升了 -0.02%、4.69%、5.02% 和 -1.28%,但本文方法在 4 种不同角度及遮挡情况的数据集中整体表现优秀。 GhostFaceNet 性能在 MaskLFW 和 SFM 测试集中涨幅明显。进一步的,在 GhostFaceNet 效果不理想情况下,本文方法加入并联结构的注意力机制后针对墨镜遮挡目标人脸有了更好的检测效果。本文提出的 GhostFaceNet+CBAM 的目标检测算法主要用于人脸遮挡的识别和预测,其结果与 GhostFaceNet 算法相比都有了一定提升。通过对两种方法的比较表明,引入注意力机制将增强 GhostFaceNet 识别遮挡人脸的性能,提升对人脸识别及遮挡人脸识别的准确率。

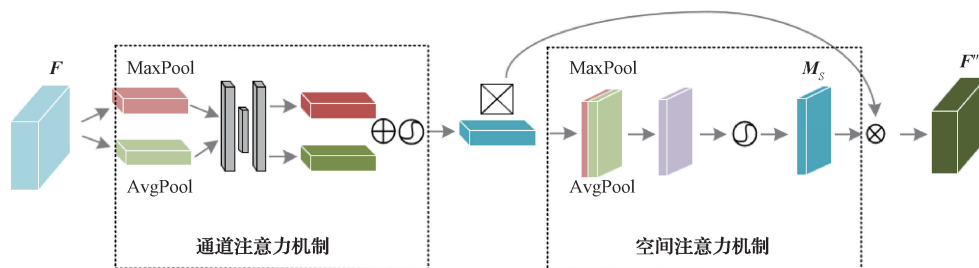


图12 串联结构

$$F' = M_c(F) \otimes F \quad (4)$$

$$F'' = M_s(F') \otimes F' \quad (5)$$

式中： F' 为经过通道注意力处理后的数据； F'' 为经过空间注意力处理后的数据。

图13所示为串联结构 ROC 测试结果曲线，图14所示为并联结构 ROC 测试结果曲线。ROC 曲线代表着对同一信号刺激的不同的感受性。比较图13和14结果表明，在 GhostFaceNet 加入并联结构的 ROC 测试结果比串联结构高7%，GhostFaceNet 加入并联结构的 ROC 测试曲线更加靠近左上角，则其试验灵敏度越高，误判率越低，整体性能越好。

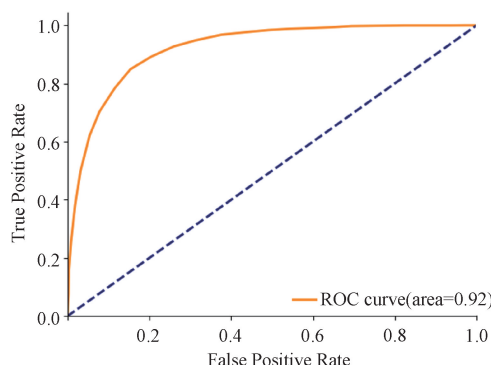


图13 串联结构 ROC 测试结果曲线

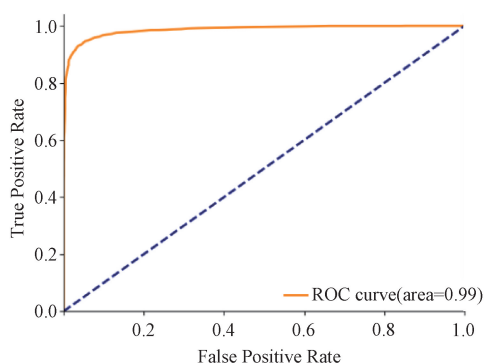


图14 并联结构 ROC 测试结果曲线

鉴于并联结构误判率低，对人脸识别准确率效果提升显著更适用于本文网络，本文采用并联结构算法。

3.2 子模块消融实验

本文通过消融实验研究各子模块，验证了不同模块对模型的提升效果，同时对比各子模块的性能。实验的设计与结果如表3所示，具体验证 GhostNet 模块、并联注意力机制模块、FPN 模块对模型的有效性。

表3 消融实验结果 (%)

子模块	LFW	MaskLFW
GhostNet	96.77	84.05
FPN	97.11	93.85
注意力	98.47	98.08

表3中各模块都表现良好，其中加入 GhostNet 模块后在 LFW 数据集中准确率达到 96.77% 相较于原始网络稍有下降，但在 MaskLFW 数据集中表现明显优于 FaceNet。引入 FPN 结构加强网络特征提取，在 MaskLFW 数据集存在遮挡的图片上识别上效果显著，将网络准确率提升近 10%。并联注意力机制的使用在 MaskLFW 中将准确率提升到了 98.08%，优化效果明显。综合各子模块的实验结果，注意力机制模块效果最好，大幅度提高了网络对存在遮挡情况的人脸识别精度。

4 结论

本文提出了改进的 FaceNet 人脸识别算法，通过 GhostNet 网络提取人脸面部特征，GhostNet 中的大小残差部分将加深网络，提取更精确的特征。同时又优化了 GhostNet 中的注意力机制模块，将通道注意力机制改进为通道加空间注意力机制，更注重对空间特征的学习。引入 FPN 加强特征提取结构，增加不同感受野下各特征的权重，使网络加强对更细微的特征学习。通过实验表明，本文算法可准确识别人脸，同时对存在遮挡的人脸也具有较好的识别效果，但因网络复杂度提升导致识别速度降低，后续还需优化。

参考文献

- [1] RANJAN R, SWAMI S, CHEN J C, et al. Deep learning for understanding faces: Machines may be just as good, or better, than humans [J]. IEEE Signal Processing Magazine, 2018, 35(1): 66-83.
- [2] 冯婧, 顾梅花, 刘晓龙, 等. 课堂场景下改进 MobileNetv3 人脸识别算法[J]. 国外电子测量技术, 2022, 41(10): 47-55.
- [3] 马博宇, 尉寅玮. 基于 AdaBoost 算法的人脸识别系统的研究与实现[J]. 仪器仪表学报, 2016, 37(S1): 162-167.
- [4] 李成勇, 王莎, 陈成瑞. 基于 OpenCV 的人脸识别系统设计与实现[J]. 国外电子测量技术, 2021, 40(11): 168-172.
- [5] 高工, 杨红雨, 刘洪. 基于深度学习的三维点云人脸识别[J]. 计算机应用, 2021, 41(9): 2736-2740.
- [6] 张旭. 基于二维照片进行人脸识别的算法研究[J]. 计算机仿真, 2012, 29(3): 330-333.
- [7] 王一炜, 孙楠. 基于激光扫描的面部智能识别人机交互界面[J]. 激光杂志, 2019, 40(4): 161-166.
- [8] 高工, 杨红雨, 刘洪. 基于深度学习的三维点云人脸识别[J]. 计算机应用, 2021, 41(9): 2736-2740.
- [9] 余璀璨, 李慧斌. 基于深度学习的人脸识别方法综述[J]. 工程数学学报, 2021, 38(4): 451-469.
- [10] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural

- networks [C]. 26th Annual Conference on Neural Information Processing Systems, 2017, 60(6):84-90.
- [11] TAIGMAN Y, YANG M, RANZATO M, et al. DeepFace: Closing the gap to human-level performance in face verification[C]. 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014: 1701-1708.
- [12] 焦亚萌, 周成智, 李文萍, 等. 融合多头注意力的 VGGNet 语音情感识别研究[J]. 国外电子测量技术, 2022, 41(1):63-69.
- [13] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions [C]. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015:1-9.
- [14] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016:770-778.
- [15] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: A unified embedding for face recognition and clustering [C]. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015:815-823.
- [16] HAN K, WANG Y, TIAN Q, et al. Ghostnet: More features from cheap operations[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020:1580-1589.
- [17] 陶志勇, 闫明豪, 刘影, 等. 基于 AG-CNN 的轻量级调制识别方法[J]. 电子测量与仪器学报, 2022, 36(4): 241-249.
- [18] 程换新, 成凯, 蒋泽芹. 基于注意力机制的 CNN 人脸表情识别[J]. 电子测量技术, 2021, 44(10):128-132.

作者简介

张晋婧, 硕士, 主要研究方向为图像处理与目标检测。
E-mail: zhangjinjingemail@qq.com

刘双峰(通信作者), 博士, 副教授, 主要研究方向为动态测试与传感器。
E-mail: liusf@nuc.edu.cn