

文章编号 1004-924X(2023)21-3192-11

融合卷积块注意力模块和 Siamese 神经网络的人脸识别算法

孟祥周, 李映君*, 王桂从, 蒙天生
(济南大学 机械工程学院, 山东 济南 250022)

摘要:针对传统人脸识别方法识别性能较差,基于深度学习的方法在非限制条件下识别较为困难,人脸特征区分性弱,识别精度容易受到姿势、表情等方面影响的问题,提出了一种引入卷积块注意力模块的孪生神经网络模型结构。该结构是基于孪生神经网络(Siamese neural network)的基础框架进行改进的,在框架中引入改进的 VGG11_BN 模型进行特征提取。该模型是在 VGG11 结构的基础上引入批归一化(Batch Normalization, BN)技术,在原模型结构的基础上,提出引入 CBAM 混合注意力机制的特征提取网络;最后,针对目前亚洲人的人脸识别训练较少的情况,采用更加符合亚洲人脸特征的 CASIA-FaceV5 数据集进行识别训练。实验结果表明:本文算法在人脸识别方面的准确率达到 96.67%,并且在 CAS-PEAL-R1 人脸数据集上比 SRGES, VGG11+siamese 算法的准确率分别提升 6.05%, 6.7%。该算法可以在多因素影响下更好地进行人脸识别验证,具有良好的稳定性。

关键词:人脸识别;孪生神经网络;深度学习;注意力机制;稳定性

中图分类号: TP394.1; TH691.9 **文献标识码:** A **doi:** 10.37188/OPE.20233121.3192

Face recognition algorithm incorporating CBAM and Siamese neural network

MENG Xiangzhou, LI Yingjun*, WANG Guicong, MENG Tiansheng

(School of Mechanical Engineering, University of Jinan, Jinan 250022, China)

* Corresponding author, E-mail: me_liyj@ujn.edu.cn

Abstract: Traditional face recognition methods have poor recognition performance; deep learning-based methods face difficulty recognizing under unrestricted conditions, face features are weakly differentiated, and recognition accuracy is easily affected by pose and expression. To address this, a twin neural network model structure that introduces a Convolutional Block Attention Module (CBAM) hybrid attention mechanism is proposed. First, the algorithm structure based on the basic framework of the Siamese neural network was improved and the improved VGG11 into the framework is introduced. The BN model for feature extraction is used, which introduces batch normalization (BN) technology based on the VGG11 structure. Second, a feature extraction network incorporating a CBAM mixed attention mechanism was introduced based on the original model structure. Finally, in response to the lack of facial recognition training for Asians, the CASIA-FaceV5 dataset was employed, which is more aligned with Asian facial features,

收稿日期:2023-05-30;修订日期:2023-07-05.

基金项目:国家自然科学基金资助项目(No. 51875250);山东省自然科学基金资助项目(No. ZR2023ME109);山东省高等学校青创科技支持计划资助项目(No. 2019KJB018);济南市“新高校 20 条”项目(No. 202228116)

for recognition training. The experimental results show that the algorithm's accuracy reaches 96.67% in face recognition, and the accuracy on CAS-PEAL-R1 face dataset is 6.05% and 6.7% higher than that of SRGES and VGG11+siamese algorithms, respectively. The algorithm in this study can better verify facial recognition under multiple factors, has good robustness, and greater application value.

Key words: face recognition; Siamese neural network; deep learning; attention mechanism; robustness

1 引言

随着目标识别^[1]以及计算机技术的发展,利用计算机进行生物特征识别训练成为当前的研究热点,并得到了快速发展。人脸识别作为生物特征识别技术的重要组成部分,相对于指纹识别、DNA 识别等其他识别技术具有可识别性强、误识率低等优点,还具有非接触性、自然性等特点,具有广阔的应用前景。目前,人脸识别技术^[2-3]广泛应用于安防^[4]、金融^[5]和身份认证^[6]等领域,为信息安全和公共安全提供了重要保障。人脸识别技术作为生物特征识别的一部分,其特点在于人脸作为人类的内在属性存在个体的差异性和稳定性,即每个个体的人脸特征都是独一无二的。因此,利用人脸识别的方式进行人脸特征的提取以及身份检验是一个十分有效的手段。

传统的人脸识别方法包括基于几何特征法^[7]、基于特征脸^[8]和基于模板匹配^[9]等。这些方法通过对人脸浅层表面进行特征提取进而实现识别,但其特征提取能力较弱,只有在简单的环境中进行识别验证具有良好的效果,无法提取深层次的人脸特征,在出现光照、角度、表情变化等其他复杂环境情况时,识别精度及准确度会下降。近年来,随着深度学习的发展^[10],基于卷积神经网络(Convolutional Neural Network, CNN)的人脸识别算法成为新的主流,利用神经网络进行训练可以使算法具备更强的表征学习能力,并且从大量的数据中学习更多的特征信息,能够在不同光照、姿态和表情等因素下学习到人脸特征,因此对于不同场景下的人脸识别具有更强的稳定性。

在利用深度学习进行人脸识别训练时,主要通过隐藏层对图像特征进行提取与处理,自动学习到更复杂、更高级的图像特征,这些特征与传统方法所获取的特征相比具有更强的表达能力。Schroff 等^[11]使用 Facenet 模型来实现从人脸图像

到欧几里得空间的映射,并利用分布方法训练模型,使用三元组损失来提高模型性能,但该方法采用的数据集不具备普遍性,难以反映真实情况下人脸的多样性和复杂性。薛继伟、孙宇锐等^[12]提出了采用 ArcFace 算法进行人脸识别的方法,并针对照片尺寸、人脸位置等方面不一致的情况进行相关研究,改善了人脸位置等造成的影响,但同时对图像要求较高,需要进行精确的预处理,并且对噪声数据敏感,局限性较大。在针对孪生神经网络^[13]进行人脸识别方面,Mathi 等^[14]提出利用孪生神经网络和 MTCNN 算法结合的方式实现不同场景下的人脸识别,其准确率可以达到 90%,但由于训练时采用自制的数据集,多以印度地区的人脸为主。Koch 等^[15]提出利用孪生神经网络进行单次学习任务,通过使用全局仿射变换处理图像对来训练验证识别任务,并指出一次面部识别可以评估学习的特征在数据集和任务中的能力。Song 等^[16]设计了一个基于局部二进制模式(Local Binary Pattern, LBP)和频率特征感知的孪生神经网络,利用 LBP 算法来消除光照对图像的影响,适用于非限制条件下的小规模数据集的人脸识别,该算法在单一场景中的应用效果较好,但是不适合多场景的情况。

综上所述,基于深度学习的人脸识别训练多采用欧美人脸数据集,较少使用亚洲人脸数据集进行训练与测试。不同人种的人脸特征存在较大差异,这些面部特征差异对最终的训练结果会造成巨大的影响。为了针对亚洲人的人脸特征进行有效的训练与识别,本文采用中国科学院发布的 CASIA-FaceV5 亚洲人脸数据集。为了解决当存在光照、表情、姿态等干扰时,人脸识别过程较为困难,识别精度低等问题,通过融合孪生神经网络与 VGG-BN 网络结构,实现对样本的特征提取,加快模型的训练速度。针对特征提取能力有限、泛化能力较弱的情况,在 VGG-BN 网络结构的基础上引入卷积块注意力模块(Conv-

lutional Block Attention Module, CBAM), 通过强调通道和空间信息的方式, 增强对目标区域的注意力和对图像中细节信息的提取, 并通过调整特征图中的权重信息, 增强特征的稳定性和区分度, 提高模型的性能。

2 网络模型结构

2.1 Siamese 神经网络

Siamese 神经网络又叫孪生神经网络, 是一种用于比较相似度或距离的神经网络模型, 由一对相同结构、共享权重 w 和偏置值 b 的网络模型组成。其具体工作方式是输入两张不同的样本图片, 输出高维度空间表征, 并比较两张输出特征的相似度。该网络最早由 Bromley 等^[17]在 1993 年提出用于签名识别任务, 验证原签名和现签名是否相同, 但是受限于当时硬件技术的限制, 计算成本高, 无法获得广泛应用, 因此在很长一段时间内的发展是停滞的。直到深度学习技术的诞生, 计算机性能逐渐增强及神经网络快速发展, 孪生神经网络在人脸识别、文本匹配和目标跟踪等领域得到了广泛的应用。

孪生神经网络突出的特点就是可以根据不同的使用场景选择使用权重共享或不共享的两个子网络, 以及不同的距离度量和损失函数。该网络还可以进行小样本学习, 适用于样本量较少的情况, 具备较高的容错识别能力, 因此, 可以用于对容错流程要求严格的人脸识别问题。作为一种特殊结构的网络, 孪生神经网络能够与其他神经网络相结合, 以适应不同的任务场景, 如在结合卷积神经网络实现两个图像相似度的比较中, 可以利用卷积神经网络的高效性、可扩展性等进一步提高识别的准确性和稳定性。

如图 1 所示, 孪生神经网络中输入图片为 X_1, X_2 , 两幅图分别经过一个特征提取网络形成特征向量 $G_w(X_1), G_w(X_2)$, 该特征向量通过共享权重的方式进行特征提取, 可以有效地增加模型的可解释性, 并保证特征向量 $G_w(X_1), G_w(X_2)$ 具有相同的特征提取能力, 更加准确地表示两个输入样本之间的差异或相似度。采用欧氏距离这种距离度量的方式获取两个特征向量之间的差异, 得到的映射函数如下:

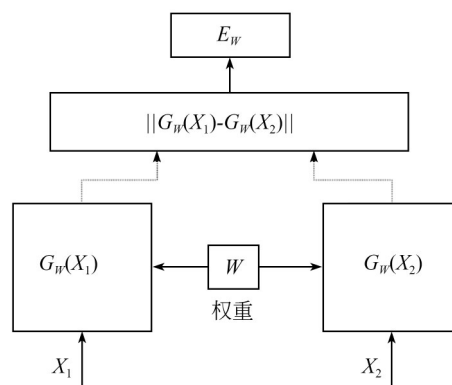


图 1 孪生神经网络结构示意图

Fig. 1 Structure diagram of Siamese network

$$E_w = \|G_w(X_1) - G_w(X_2)\|. \quad (1)$$

映射函数 E_w 主要是计算两个特征向量的差异, 可以通过两向量之差的模长判断。 E_w 越接近 1, 两者相似度越高, 越符合要求。

2.2 VGG-BN 网络模型

VGG 网络^[18]作为卷积神经网络在目标检测等计算机视觉中的重要模型, 是在 AlexNet 网络结构的基础上改进的, 具体是采用 3×3 的卷积核代替大卷积核, 如 $5 \times 5, 7 \times 7$ 的卷积核。这种采用小卷积核代替的方式会使网络深度增加, 通过重复的使用卷积核和池化层能够消除影响, 有效地提取图像特征, 并在 ImageNet 数据集上实验证明了优越性。

VGG11 网络结构就是在 VGG16 网络的基础上, 通过改变模型的深度和宽度提出的一种网络模型结构, 在具有较浅的深度同时也具有很强的表达能力和特征提取能力。本文提出的 VGG11_BN 网络结构是在 VGG11 网络模型结构的基础上添加了批量归一化 (Batch Normalization, BN) 层。BN 层是 Google^[19]提出来的一个结构, 主要用于解决梯度消失问题及加速模型的收敛速度等。BN 层将输入数据进行归一化的总公式如下:

$$y_{bn} = \gamma \hat{y} + \beta, \quad (2)$$

式中: \hat{y} 代表卷积层中输出 y 的估计值, γ, β 表示可学习的参数量。最终输出为:

$$y_m = \frac{\gamma}{\sqrt{\sigma_y^2 + \epsilon}} \left(\sum_i \omega_i x_i + b - \mu_y \right) + \beta, \quad (3)$$

式中: μ_y, σ_y 代表 BN 层的均值, ω 是权重参数, b 是偏置参数, ϵ 是一个正数。

通过这一系列的操作实现归一化的过程,使激活函数的输入分布更加稳定,加速网络的收敛。采用的激活函数为 Relu 激活函数,如下:

$$f(x) = \begin{cases} x & x > 0 \\ 0 & x \leq 0 \end{cases}, \quad (4)$$

式中 x 为来自于 BN 层的输出。

VGG11_BN 网络的结构示意图如图 2 所示。

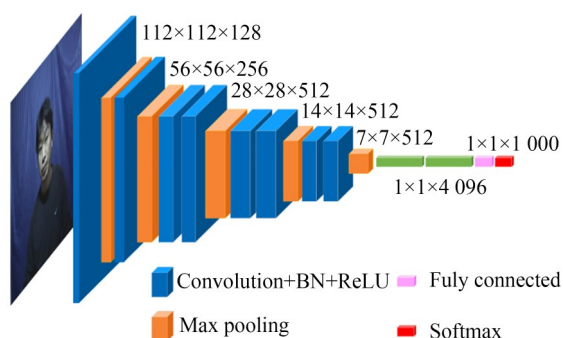


图2 VGG11_BN网络结构示意图

Fig. 2 VGG11_BN network structure diagram

2.3 混合注意力机制

计算机视觉中把注意力聚集在图像的重要区域并丢弃不相关的方法称为注意力机制,这种注意力机制能够更快捷和高效地分析复杂场景信息,快速提取有价值的信息,摒弃无用信息。应用到计算机领域可以有效地降低计算复杂度,减少干扰,提高计算效率。

注意力机制作为一种对图像输入重要信息的动态选择过程,是基于特征自适应权重实现的,在目标检测、人脸识别、小样本检测等任务中发挥着重要作用。在人脸识别方面,注意力机制可以更好地处理人脸的不同部位和特征。相较于传统的人脸识别算法使用固定的特征表示方法,注意力机制可以根据人脸的不同部位和特征自适应地计算不同输入数据的权重,更加有效地利用输入数据的信息,提高模型的性能和鲁棒性。张晋婧等^[20]提出了融合注意力机制的人脸识别算法,将注意力机制融合到 facenet 人脸识别算法中,用于加强不同感受野下的特征提取,增强较为重要的特征信息,并抑制其他不重要的特征。

注意力机制通常划分为通道注意力机制、空

间注意力机制、时间注意力机制和混合注意力机制等,具体的实现方式和效果不同,其联系如图3所示。

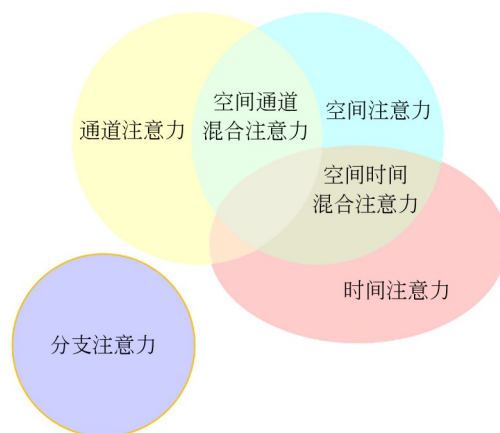


图3 注意力机制关联图

Fig. 3 Attention mechanism correlation chart

针对待识别图像中存在一些不相关的背景噪声等影响因素,这里选择CBAM进行处理,它属于一种混合注意力机制,有助于细化特征、提高性能。CBAM的示意图如图4所示。

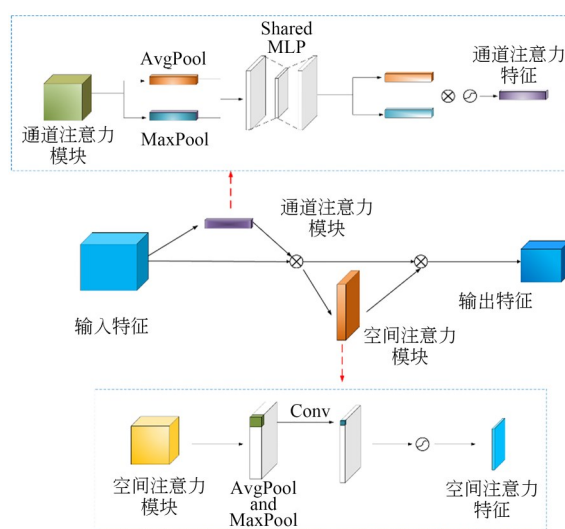


图4 CBAM整体示意图

Fig. 4 Overall diagram of CBAM

总体运算公式如下:

$$\begin{aligned} F' &= M_c(F) \otimes F \\ F'' &= M_s(F') \otimes F' \end{aligned} \quad (5)$$

式中: \otimes 表示元素级乘法, $M_c(F)$ 表示通道注意力特征图, $M_s(F)$ 表示空间注意力特征图。

3 基于注意力机制的 Siamese 神经网络算法

VGG 网络模型是当前比较优秀的图像特征提取算法之一,广泛应用在人脸识别领域,以及迁移学习方面。通过使用自定义全连接层和 VGG11_BN 模型相拼接来修改原模型并形成新的网络结构,具体方式是将原模型中的全连接层去掉,然后通过最大池化操作及 Flatten 标准降维方法来“压平”数据,最后利用两个全连接层实现特征的比较与融合。在训练过程中,还利用迁移学习的方式调用 VGG11_BN 网络的预训练模型,该模型在训练过程中采用 ImageNet 数据集,数据种类众多,只需要保留原来模型中的底层特征的有用信息即可,该算法在预训练模型的基础上进行训练,能够减少模型的训练时间,加快收敛速度。

在不同的注意力机制中,通道注意力机制可以自适应地调整输入张量的不同通道的权重,使得模型更加关注某些重要的通道,如 SENet^[21] 注意力算法包括一个全局平均池化层和两个全连接层。Almadan 等^[22] 采用 SENet 注意力进行人脸识别,采用两个全连接层可以有效地提升特征的维度,但在参数过多时会大幅度降低训练效率。空间注意力机制可以自适应地调整输入张量中不同位置的权重,使得模型更加关注某些重要的位置,在文本处理和图像处理任务中的应用较为广泛。特别在图像处理任务中,自注意力机制可以根据图像中每个像素点的不同特征自适应地调整每个像素点的权重,以提高模型的表现能力。但由于只能调整空间位置的权重而无法调整不同通道的权重,在处理通道特征时效果较差。因此,采用 CBAM 中的空间注意力可以在空间和通道两个维度上同时调整权重,可以适应不同的输入数据,提高模型的性能。CBAM 通过在隐藏层的前半部分和后半部分分别添加注意力机制的方式来提高模型的性能,在前面添加注意力机制有利于减少计算量,加强特征提取能力和提高模型的表现能力;在后面添加注意力机制可以增强模型的泛化能力,优化模型结构和增强模型的稳定性。

将 CBAM 混合注意力模块和提出的 VGG 网络相结合,实现对空间和通道两个维度上权重的同时调整和模型的优化。最终经过注意力机制后的输出为:

$$y'' = s \cdot y'_c, \quad (6)$$

$$s = \sigma(m), \quad (7)$$

$$m = f(\text{Avgpool}(y'_c) + \text{Maxpool}(y'_c)) \quad (8)$$

式中: m 是一个二维特征图, f 是一个卷积层, s 是一个空间注意力权重, σ 是 sigmoid 激活函数, y'_c 是通道注意力输出。 y'_c 的公式为:

$$y'_c = s_c \cdot y_c, \quad (9)$$

$$s_c = \sigma(f(\text{MLP}(\text{Avgpool}(\hat{y}_c)) + \text{MLP}(\text{Maxpool}(\hat{y}_c)))) \quad (10)$$

$$\hat{y}_{cij} = \frac{y_{cij} - \mu_c}{\sqrt{\sigma_c^2 + \epsilon}}, \quad (11)$$

式中: c 是通道索引, i 和 j 是空间索引, ϵ 是一个很小的常数, MLP 是一个多层感知器, AvgPool 和 MaxPool 分别代表平均池化和最大池化, σ 是 sigmoid 函数, S_c 是通道注意力权重, μ_c 和 σ_c^2 分别是均值与方差。其具体表达式为:

$$\mu_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W Y_{i,j}, \quad (12)$$

$$\sigma_c^2 = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W (y_{cij} - \mu_c^2), \quad (13)$$

式中: H, W, C 分别表示特征图的高度、宽度和通道数, Y 是来自于 BN 层的输出。 Y 的计算公式如下:

$$Y = \gamma \cdot \frac{X - \mu}{\sqrt{\sigma^2 + \epsilon}} + \beta, \quad (14)$$

$$u = \frac{1}{m} \sum_{i=1}^m x_i, \quad (15)$$

$$\sigma^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu)^2, \quad (16)$$

式中: x 是输入, μ 是均值, σ^2 是方差, ϵ 是一个很小的常数, γ 和 β 是可学习的参数, 用来调整数据的分布, 输出为特征图 $y \in \mathbf{R}^{H \times W \times C}$ 。

将修改后的 VGG11_BN 网络和孪生神经网络融合, 网络模型的结构示意图如图 5 所示。将该修改后的网络结构放到特征提取部分, 在这双分支的 VGG11_BN 网络中, 分别输入一张人脸图片, 形成正样本或负样本, 再分别进行特征提

取,将样本映射到低纬度的空间中。然后,将提取的特征向量压平成一维向量,通过全连接层提取高级特征以更好地度量相似性,并通过相似度判定是否为同一人。

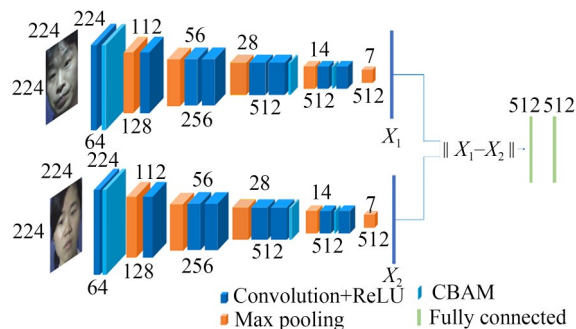


图5 VGG11_BN结合孪生网络的模型结构示意图

Fig. 5 Model structure diagram of VGG11_BN combined with Siamese neural network

4 实验及结果分析

4.1 实验设置及数据集预处理

实验的硬件搭建平台为 Windows11 操作系统,英特尔 Core i9-10900K 处理器,16 G 内存,显卡型号为 NVIDIA Quadro RTX 4000,8G 显存;软件平台为 Pytorch1.12.1 深度学习框架搭配 Cuda11.3 高性能并行运算平台。

采用的人脸数据集为中国科学院发布的 CASIA-FaceV5 亚洲人脸数据集,该数据集的类内变化包括照明、姿势、表情、眼镜和成像距离等,能够针对不同的姿态、表情变化等情况进行相关训练。作为一个开放的亚洲人脸数据集,该数据集一共对 500 个中国人进行采样拍照,每个人有 5 张不同类内变化的图片,共 2 500 张高质量的彩色图片,每个图片的像素都是高 480、宽 640,由于图片采集的是每个人的上半身图像,故图片需要进行预处理。

利用 PaddlePaddle(飞桨)^[23]预训练模型管理和迁移学习工具 PaddleHub 进行人脸检测以及 CASIA-FaceV5 数据集的预处理,具体方式是采用 PaddleHub 中的 ultra_light_fast_generic_face_detector_1mb_640 模型进行预处理,获取图片中的人脸位置,针对获取到的数据点的位置,将图片裁剪成 224×224 的人脸图片。该工具

具有检测速度快、精度高等优点,利用 MTCNN 检测需要 20 ms,而该模型只要 4 ms。

4.2 模型评价指标

模型的评价指标主要采用平均准确率(Accuracy, Acc),AUC(Area Under the Receiver Operating Characteristic Curve)和 ROC(Receiver Operating Characteristic)曲线。平均准确率是指在测试集的样本中正确分类样本所占的比例;ROC 曲线可以反映人脸识别的整体性能,又称为受试者工作特征曲线,主要用来评价人脸识别模型在不同阈值下的识别性能。通过 ROC 曲线线形可明确看出模型性能的优劣程度,其横轴表示假阳性率(FPR),纵轴表示真阳性率(TPR),将不同的阈值作为分界点,绘制出模型训练后的结果在不同阈值下的性能表现,帮助更好地了解模型性能。TPR 和 FPR 的计算公式如下:

$$TPR = \frac{TP}{TP + FN}, \quad (17)$$

$$FPR = \frac{FP}{FP + TN}, \quad (18)$$

式中:TP(True Positive)表示真正类;TN(True Negative)表示真负类;FP(False Positive)表示假负类;FN(False Negative)表示假正类。

平均准确率的计算公式如下:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%. \quad (19)$$

AUC 的具体意义是指 ROC 曲线图形的下面积,用来衡量人脸识别模型的整体性能。它作为衡量指标,面积越大,性能越好。具体计算公式如下:

$$AUC = \frac{\sum_{i \in \text{positive}} rank_i - \frac{M(1+M)}{2}}{M \times N}, \quad (20)$$

式中:rank_i代表第 i 条样本的序号,∑_{i ∈ positive} 代表把正样本的序号进行相加,M,N 分别是正样本和负样本的个数。

4.3 实验过程及结果分析

在进行人脸识别的训练过程中超参数的调整对于最终的训练结果具有举足轻重的作用,而超参数包括学习率(learning rate)、批量大小(batch size)和激活函数等。学习率参数的设置影响着模型的收敛速度和性能,从而最终影响模型训练的效果,因此确定学习率的范围十分重

要。根据经验以及查阅的资料,将学习率的初始值设定为 1×10^{-6} ,然后设定以 1×10^{-8} 进行线性增长,记录学习率和损失值的数据关系,绘制的曲线如图 6 所示。

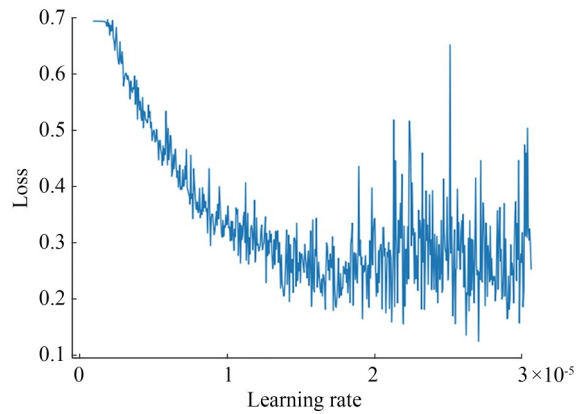


图 6 损失随学习率的变化曲线
Fig. 6 Loss variation curve with learning rate

分析图 6 中学习率与损失值之间的关系可知,损失值先降低然后出现振荡。学习率的选择范围为损失值快速下降的区间^[24],因此基于以上描述和曲线图可确定学习率在 $1 \times 10^{-6} \sim 1 \times 10^{-5}$ 之间。

在进行网络模型的训练过程中,多是采用批量(Batch)训练法,因此在训练过程中,模型每次输入输出一个 Batch,就要调整一次参数,其参数调整方法如下:

$$\theta_{t+1} = \theta_t - lr * g(\theta_t), \quad (21)$$

式中: θ_{t+1} 是第 $t+1$ 次迭代后的参数值, θ_t 是第 t 次迭代后的参数值, lr 是学习率。动态调整学习率十分重要,选用周期性学习率(CyclicLR)调整学习率的策略,这种调度器会在训练过程中周期性地调整学习率,从较小的值逐渐增加到较大的值,然后再逐渐降低回较小的值,从而有效地帮助模型跳出局部最优解,避免过早陷入局部最优解,提高模型的泛化能力。在 CyclicLR 学习率中有 3 种不同的学习率策略,包括 triangular, triangular2 和 exp_range 3 种学习率。将这 3 种学习率分别放到模型中进行训练,并在 CASIA-FaceV5 人脸数据集中进行测试,结果如图 7 所示。由图 7 可知, CyclicLR 学习率中 3 种不同学习率策略对于该模型的匹配程度不一,使用 triangular2 学

习率的 ROC 曲线的表现优于其他两种学习率策略。

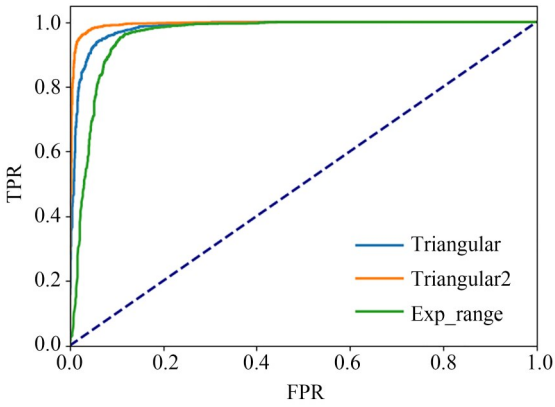


图 7 不同学习率下的 ROC 曲线
Fig. 7 ROC curves at different learning rates

比较在不同学习率下的平均准确率、AUC 和平均精度,结果如表 1 所示。由表 1 可知,相同的网络模型下,3 种不同的学习率策略中 triangular2 学习率策略的平均准确率, AUC 和平均精度最高,其平均准确率比 triangular 学习率策略高将近 4%,比 exp_range 学习率策略高将近 7%。

平均精度(mAP)能够全面地评估性能,因此 triangular2 学习率比 exp_range 学习率高 5.79%,具备更高的准确性。最终表明, CyclicLR 学习率下 triangular2 策略对于网络模型训练具有更好的收敛性。

表 1 本文算法中不同学习率的比较

Tab. 1 Comparison of different learning rates in proposed algorithm (%)

LR	AUC	平均准确率 (Acc)	平均精度 (mAP)
Triangular	98.22	92.34	98.11
Triangular2	99.36	96.67	99.37
Exp_range	95.76	89.54	93.58

为了进一步证明 CBAM_VGG 算法的准确性及泛化能力,更换数据集进行上述网络模型的测试,采用由中国科学院技术研究所 2003 采集创建的 CAS-PEAL-R1^[25]数据集。这个数据集包含姿态、表情、光照和距离等特征的变化,适应多种场景的应用。该数据库一共采集了 1 040 位志愿

者,共 99 594 个面部图像,这一千多位志愿者由 595 名男性和 445 位女性组成,选取其中一部分面部图像作为测试集,部分样例如图 8 所示。



图 8 数据集样例

Fig. 8 Sample from dataset

在该网络模型结构下,为了评价各种算法与本文算法的性能差异,采用 ROC 曲线评估模型性能,如图 9 所示。从图 9 可以看出,通过修改后的算法相对于传统的融合 VGG 和孪生神经网络模型的算法具备更优的性能,意味着该模型在识别人脸时具有更高的真阳性率和更低的假阳性率,在实际检验时能够更加准确地识别,减少误判。将本文算法也与其他经典算法如 LWSFLA-SVM^[26], SRGES^[27], KEHD^[28], KCCA-SVM^[29] 及 VGG+Siamese 进行比较,结果如表 2 所示。

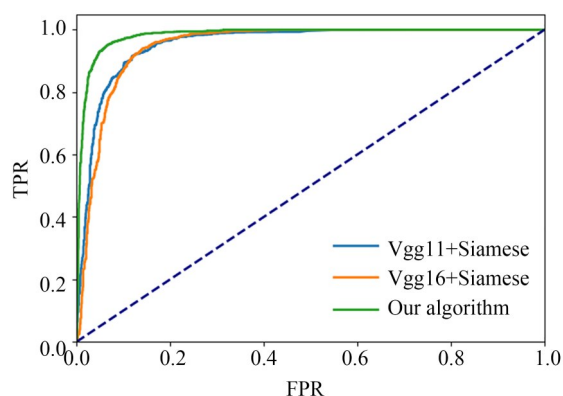


图 9 本文算法和传统 Siamese 神经网络的比较

Fig. 9 Comparison between proposed algorithm and traditional Siamese neural network

由表 2 可知,该算法在 CAS-PEAL-R1 数据集上与其他算法相比,识别率分别提高了 1.3%, 6.05%, 2.13%, 0.8%, 8.8% 和 6.7%,证明该算

表 2 不同算法的准确率比较

Tab. 2 Comparison of accuracy rates of different algorithms (%)

Method	Recognition rate
LWSFLA-SVM ^[26]	93.00
SRGES ^[27]	88.25
KEHD ^[28]	92.17
KCCA-SVM ^[29]	93.50
VGG16+siamese	85.50
VGG11+siamese	87.60
Proposed	94.30

法在存在姿态、表情等变化时仍然具有较好的人脸识别效果。相较于使用 SVM 的方法,虽然在识别效果上的提升不大,但是孪生神经网络更容易改进,能够及时融合先进算法,本文就是通过结合改进的 vgg 算法进行训练;同时在处理小规模样本上也具有较大的优势,能够提供更快的收敛速度,具备更高的性能和识别能力。与其他孪生神经网络相比,添加 CBAM 注意力机制后其平均准确率得到了较大的提升。具体的测试效果样例如图 10 所示。图 10 中,4 组测试分别是一

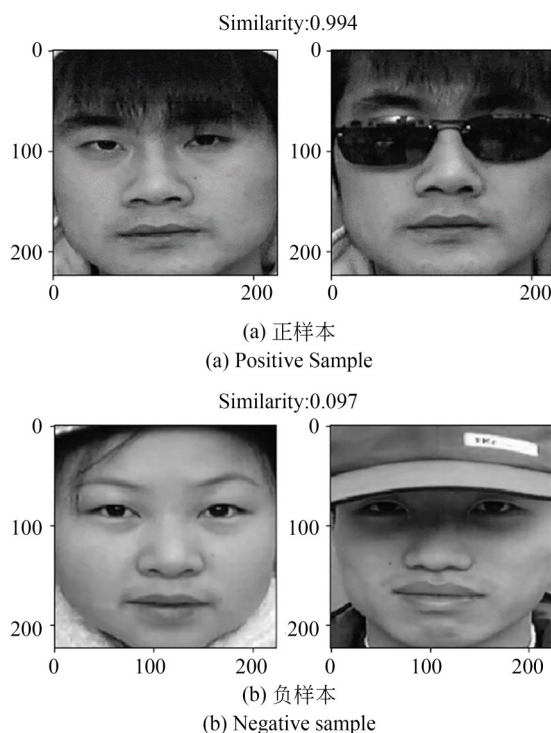


图 10 测试效果样例

Fig. 10 Sample test results

对正样本和负样本,可以看出正样本中相似度较高,负样本中图片的相似度就较低,实验结果也与期望一致。

4.4 消融实验

为了进一步探究本文方法对模型性能产生的积极效果,在 CASIA-FaceV5 和 CAS-PEAL-R1 数据集中设计消融实验进行对比分析。同时,为了证明融合 CBAM 注意力机制模块的有效性,分别单独加入 CBAM 模块进行消融实验来判断它对模型性能的影响。实验结果如表 3 所示。

表 3 消融实验结果

Tab. 3 Results of ablation experiment

模 型	计算量 /GFLOPs	平均准确率/%	
		CASIA- FaceV5	CAS- PEAL-R1
Base	15.06	81.40	87.6
CBAM+Base	12.60	85.90	87.5
Base+CBAM	14.80	90.30	92.2
本文算法	13.50	96.67	94.3

表 3 中,Base 模型表示基础模型,未添加 CBAM 注意力机制;CBAM+Base 表示在提取网络的前端添加 CBAM 注意力机制;Base+CBAM 表示在提取网络的后端添加 CBAM 注意力机制。通过分析可知,基础模型在两个数据集中的平均准确度最低,模型的计算量最高;当在提取网络的前端添加 CBAM 时,识别的平均正确率在 CASIA-FaceV5 数据集中只提高了 4.5%,但减少了整个模型的计算量;在提取网络

的后端添加 CBAM 时,模型整体的计算量并没有改变,但识别的平均准确率分别增加了 8.9%,4.6%,比在前端添加 CBAM 提高了将近五个百分点,提高了模型对于特征的提取能力,增强了模型的稳定性。最终,模型结构的计算量降至 13.5GFLOPs,平均准确率分别提高了 15.27%,6.7%。消融实验结果表明,本文方法能够快速有效地进行人脸识别。

5 结 论

本文提出了一种融合 CBAM 和 Siamese 神经网络的人脸识别算法,该算法在融合 VGG 和孪生神经网络的基础上修改了部分结构并引入 CBAM 混合注意力模块,提高了特征提取能力。相较于其他需要进行身份标记的算法,该算法通过直接测量两张人脸特征间的相似性进行识别,既简化了过程,又提高了效率。在此基础上,通过判断不同的学习率策略,选择最佳的 Radam 优化器和 CyclicalLR 循环学习率组合策略,其中 Radam 优化器可以在进行人脸识别训练时自适应动量机制和偏置校正,提高训练效果。最终,该算法测试的平均准确率达到 96.67%,在 CAS-PEAL-R1 数据集的平均准确率达到 94.3%,相比于修改前的算法提高的 6.7%,并且相较于其他经典算法的表现也十分出色,能适应在数据集中因姿态、表情、光照等变化产生的影响。综合可知,本文提出的算法具备良好的稳定性和泛化能力,在应对不同姿态、表情等变化时能够提高识别性能,对于人脸识别的应用与研究具有重要的参考价值。

参考文献:

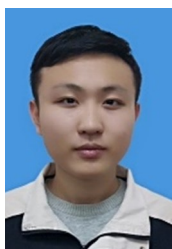
- [1] 刘可佳,马荣生,唐子木,等. 采用优化卷积神经网络的红外目标识别系统[J]. 光学精密工程, 2021, 29(4): 822-831.
LIU K J, MA R SH, TANG Z M, *et al.* Design of infrared target recognition system with optimized convolutional neural network [J]. *Opt. Precision Eng.*, 2021, 29(4): 822-831. (in Chinese)
- [2] 胡麟苗,张湧,楼晨凤. 基于内容特征提取的短波红外-可见光人脸识别[J]. 光学精密工程, 2021, 29(1): 160-171.

- HU L M, ZHANG Y, LOU CH F. Shortwave infrared visible-light face recognition based on content feature extraction[J]. *Opt. Precision Eng.*, 2021, 29(1): 160-171. (in Chinese)
- [3] LI M, HUANG B, TIAN G. A comprehensive survey on 3D face recognition methods [J]. *Engineering Applications of Artificial Intelligence*, 2022, 110: 104669.
- [4] GUO M H, XU T X, LIU J J, *et al.* Attention mechanisms in computer vision: a survey[J]. *Computational Visual Media*, 2022, 8(3): 331-368.

- [5] KHAN A A, SHAIKH A A, SHAIKH Z A, *et al.* IPM-Model: AI and metaheuristic-enabled face recognition using image partial matching for multimedia forensics investigation with genetic algorithm [J]. *Multimedia Tools and Applications*, 2022, 81 (17): 23533-23549.
- [6] CONWAY D, SIMON L, LECHERVY A, *et al.* *Training Face Verification Models from Generated Face Identity Data* [EB/OL]. 2021: *arXiv*: 2108.00800. <https://arxiv.org/abs/2108.00800>. pdf.
- [7] 曹海燕, 曹广. 基于几何特征的人脸识别方法研究 [J]. *电子技术*, 2021, 50(6): 27-29.
CAO H Y, CAO G. Research on face recognition method based on geometric features [J]. *Electronic Technology*, 2021, 50(6): 27-29. (in Chinese)
- [8] LIYAKAT K K S, MANE V A, PARADESHI K P, *et al.* Development of pose invariant face recognition method based on PCA and artificial neural network [J]. *Journal of Algebraic Statistics*, 2022, 13 (3): 3676-3684.
- [9] JAYA P A, KUMAR P K, SAUNAK S, *et al.* A deep learning technique for biometric authentication using ECG beat template matching [J]. *Information*, 2023, 14(2): 65.
- [10] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks [J]. *Science*, 2006, 313(5786): 504-507.
- [11] 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)*. June 7-12, 2015. Boston, MA, USA. IEEE, 2015: 815-823.
- [12] 薛继伟, 孙宇锐, 辛纪元. 基于ArcFace算法的人脸识别应用研究 [J]. *电子设计工程*, 2022, 30 (11): 168-172.
XUE J W, SUN Y R, XIN J Y. Research on the application of face recognition based on ArcFace algorithm [J]. *Electronic Design Engineering*, 2022, 30(11): 168-172. (in Chinese)
- [13] 孙运达, 万雪, 李盛阳. 基于孪生网络的航天器部件追踪 [J]. *光学精密工程*, 2021, 29(12): 2915-2923.
SUN Y D, WAN X, LI SH Y. Siamese network based satellite component tracking [J]. *Opt. Precision Eng.*, 2021, 29(12): 2915-2923. (in Chinese)
- [14] MATHI R, MOTHUKURI J V, PASUMARTHY V A, *et al.* Face recognition in different scenarios using Siamese network [C]. 2022 *IEEE 3rd Global Conference for Advancement in Technology (GCAT)*. October 7-9, 2022. Bangalore, India. IEEE, 2022: 1-6.
- [15] KOCH G R. Siamese neural networks for one-shot image recognition [C]. *ICML Deep Learning Workshop*, 2015.
- [16] SONG C L, JI S Y. Face recognition method based on Siamese networks under non-restricted conditions [J]. *IEEE Access*, 2022, 10: 40432-40444.
- [17] BROMLEY J, GUYON I, LECUN Y, *et al.* Signature verification using a "Siamese" time delay neural network [C]. *Proceedings of the 6th International Conference on Neural Information Processing Systems*. 29 November 1993, Denver, Colorado. New York: ACM, 1993: 737-744.
- [18] SIMONYAN K, ZISSERMAN A. *Very Deep Convolutional Networks for Large-scale Image Recognition* [EB/OL]. 2014: *arXiv*: 1409.1556. <https://arxiv.org/abs/1409.1556>. pdf.
- [19] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift [C]. *Proceedings of the 32nd International Conference on International Conference on Machine Learning-Volume 37*. New York: ACM, 2015: 448-456.
- [20] 张晋婧, 刘双峰, 丰雷, 等. 融合注意力机制的人脸识别算法研究 [J]. *国外电子测量技术*, 2023, 42(2): 107-113.
ZHANG J J, LIU SH F, FENG L, *et al.* Research on face recognition algorithm integrating attention mechanism [J]. *Foreign Electronic Measurement Technology*, 2023, 42(2): 107-113. (in Chinese)
- [21] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. June 18-23, 2018. Salt Lake City, UT. IEEE, 2018: 7132-7141.
- [22] ALMADAN A, KRISHNAN A, RATTANI A. BWCFace: open-set face recognition using body-worn camera [C]. 2020 *19th IEEE International Conference on Machine Learning and Applications*

- (ICMLA). December 14-17, 2020. Miami, FL, USA. IEEE, 2020: 1036-1043.
- [23] 马艳军, 于佃海, 吴甜, 等. 飞桨: 源于产业实践的开源深度学习平台[J]. 数据与计算发展前沿, 2019(1): 105-115.
- MA Y J, YU D H, WU T, *et al.* PaddlePaddle: an open-source deep Lear[J]. *Frontiers of Data & Computing*, 2019(1): 105-115. (in Chinese)
- [24] SMITH L N. Cyclical learning rates for training neural networks [C]. 2017 *IEEE Winter Conference on Applications of Computer Vision (WACV)*. March 24-31, 2017. Santa Rosa, CA, USA. IEEE, 2017.
- [25] GAO W, CAO B, SHAN S G, *et al.* The CAS-PEAL large-scale Chinese face database and baseline evaluations [J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 2008, 38(1): 149-161.
- [26] 刘悦婷. 基于LWSFLA-SVM的人脸识别算法[J]. 宝鸡文理学院学报(自然科学版), 2016, 36(3): 40-44.
- LIU Y T. Face recognition algorithm based on LWSFLA-SVM[J]. *Journal of Baoji University of Arts and Sciences (Natural Science Edition)*, 2016, 36(3): 40-44. (in Chinese)
- [27] DING Y, QI L, TIE Y, *et al.* Single sample per person face recognition based on sparse representation with extended generic set [C]. 2018 *International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*. October 18-20, 2018. Zhengzhou, China. IEEE, 2018: 37-375.
- [28] 马果东, 吕非, 童莹, 等. 基于核扩展混合块字典的单样本人脸识别研究[J]. 计算机技术与发展, 2022, 32(1): 104-110, 116.
- MA G D, LÜ F, TONG Y, *et al.* Face recognition with a single training sample per person based on kernel extended hybrid block dictionary learning [J]. *Computer Technology and Development*, 2022, 32(1): 104-110, 116. (in Chinese)
- [29] 潘惠苹, 任艳, 徐春. 基于核典型相关分析和支持向量机的图像识别技术[J]. 南京理工大学学报, 2022, 46(3): 284-290.
- PAN H P, REN Y, XU CH. Image recognition technology based on kernel canonical correlation analysis and support vector machine[J]. *Journal of Nanjing University of Science and Technology*, 2022, 46(3): 284-290. (in Chinese)

作者简介:



孟祥周(1999—),男,山东济宁人,硕士研究生,2021年于济南大学获得学士学位,主要研究方向为图像识别和处理技术。E-mail: 743141727@qq.com

通讯作者:



李映君(1982—),男,山东莱阳人,博士,教授,硕士生导师,2004年于山东科技大学获得学士学位,2007年于沈阳理工大学获得硕士学位,2010年于大连理工大学获得博士学位,主要研究方向为传感器与执行器测控技术、工业机器人技术、智能化仪器仪表技术。E-mail: me_liyj@ujn.edu.cn