

基于 SSD 网络模型重构的表情检测算法

陈平安¹ 邓琦^{1,2}

¹ 湖北汽车工业学院电气与信息工程学院 湖北 十堰 442002

² 圣心大学商业科技学院 美国 康涅狄格州 06825

(wuju_an@qq.com)

摘要 人脸识别与表情检测是计算机视觉和深度学习领域的热门研究方向,广泛应用于各种场景。然而,传统的表情检测方法在非受限条件下的表现差,深度学习方法则面临特征区分度低和识别精度容易受到姿势和表情变化影响等问题。对此,提出了基于 SSD 的网络模型重构与中心损失优化算法(IML-SSD),以提升面部表情检测的准确性和鲁棒性。首先,提出了一种基于网络重构优化的 SSD 面部表情快速检测算法,通过重构 SSD 算法模型中的基础层和辅助层,提高了识别速度、准确率和鲁棒性。随后,结合中心损失函数对 SSD 算法进行了进一步优化,使得同一类别的表情特征更加聚合,不同类别的特征则更加分离,从而增强了面部表情特征的判别能力。测试结果表明,所提算法优于对比算法,且在数据集 FERPlus 上的 mAP 值提升了约 6.5 个百分点。

关键词:表情检测;SSD 重构;深度神经网络

中图分类号 TP183

Expression Detection Algorithm Based on SSD Network Model Reconstruction

CHEN Ping'an¹ and DENG Qi^{1,2}

¹ Electrical and Information Engineering College, Hubei University of Automotive Technology, Shiyan, Hubei 442002, China

² Jack Welch College of Business and Technology, Sacred Heart University, Fairfield, USA 06825

Abstract Facial expression recognition and detection are prominent research areas in computer vision and deep learning, widely applied across various scenarios. However, traditional expression detection methods perform poorly under unconstrained conditions, while deep learning approaches face challenges such as low feature discriminability and susceptibility to posture and expression variations. To address these issues, this paper proposes an IML-SSD algorithm based on SSD network model reconstruction and center loss optimization to enhance the accuracy and robustness of facial expression detection. Firstly, this paper introduces a fast SSD-based facial expression detection algorithm optimized through network reconstruction. By restructuring the base and auxiliary layers of the SSD model, the algorithm achieves improved recognition speed, accuracy, and robustness. Subsequently, the SSD algorithm is further optimized by incorporating a center loss function, which enhances the aggregation of features within the same category while increasing separation between different categories, thereby strengthening the discriminative capability of facial expression features. Test results demonstrate that the proposed algorithm outperforms comparative methods, achieving approximately a 6.5 percentage points increase in mAP values on the FERPlus dataset.

Keywords Facial expression detection, SSD Refactoring, Deep neural network

1 引言

尽管人类能够轻松完成面部识别,但对计算机而言,这项任务充满挑战。由于人体是非刚性结构,面部形态会受到光照、表情变化、姿势调整等多种因素的影响而随时变化,这为计算机面部识别带来了困难。因此,首先需要设计准确的表情检测算法,尤其是在复杂背景下提取面部特征,这是表情检测中的关键一步。面部表情识别是计算机视觉中的一个重要领域,旨在通过计算机自动分析人脸表情,识别情绪状态和面部表现。与传统的人脸识别不同,表情识别更注重捕捉由面部肌肉运动引起的变化,需要精确分析眼睛、眉毛、嘴巴等部

位的微小动作,同时应对光照、姿势和表情变化的影响。有效的表情检测算法不仅要求提取静态图像中的关键特征,还要在复杂环境和各种表情变化中保持高准确性和鲁棒性。随着深度学习的发展,新的神经网络架构不断出现,推动了表情识别技术的进步。

20 世纪末,经典算法 Fisherface 和 Eigenface 问世,为表情识别奠定了基础。随后,在 20 世纪末和 21 世纪初,研究者们开始对非理想环境下的人脸数据进行研究,提出了分类器与特征提取融合的方法,典型算法包括 SVM、SIFT 特征法和 Gabor 特征法。近年来的研究逐渐结合神经网络,探索更加复杂的算法。例如, Li 等^[1]提出结合 Gabor 小波和分区选择

基金项目:湖北省重点研发计划项目国际科技合作类(2023EHA018)

This work was supported by the Hubei Provincial Key R&D Program of International Science and Technology Cooperation Category(2023EHA018).

通信作者:邓琦(dq@huat.edu.cn)

的遮挡表情识别方法;Hou 等^[2]开发了非监督的人脸图像旋转与渲染算法;Zhang 等^[8]研究了复杂环境下的表情识别;Wei 等^[4]提出了适用于光照不一致条件下的正面化特征学习方法;Zhang 等^[5]提出了基于 DNN 模型的人脸去噪算法。虽然深度卷积神经网络显著改善了表情检测算法的性能,但在现实场景中应用的检测算法,仍存在因人脸大小、姿势、遮挡、外观和光照变化等因素而导致无法准确检测的问题。Yao 等^[6]指出表情标注的主观性和差异性导致数据集中的标签噪声难以避免。对此,Wang 等^[7]提出自修复网络(SCN)来抑制不确定性,增强表情识别的鲁棒性;Zhao 等^[8]提出 Efficient-Face 算法,通过标签分布学习(LDL)优化训练,提升识别效果;Zeng 等^[9]提出的 Face2Exp 利用元优化框架消除数据偏差,通过引入大规模未标记的人脸数据集来增强表情识别模型,并通过反馈损失机制来解决数据集间类别分布不匹配的问题;Xia 等^[10]提出了一种基于条件随机场的多尺度时空网络(MS-CRFNet),通过多尺度特征来解释面部表情;Li 等^[11]提出了一种基于多分支融合和深度可分离卷积的新型面部表情识别框架 FERmc,从图像中自适应地捕获不同尺度的特征,从而提高了特征表示的鲁棒性;Jiao 等^[12]提出了一种基于多尺度注意力残差网络的表情识别方法,引入了面部多尺度注意力模块,通过通道分组来保持信息有效性和降低计算成本;Peng 等^[13]引入了一种新的特征提取器-邻域坐标注意力 Mamba(NCA-Mamba),以替代特征金字塔网络(FPN)中的原始特征提取器。这些方法共同致力于提高表情识别的准确性和鲁棒性。然而,在现实场景中,仍面临因各种复杂因素而导致难以检测的挑战。同时,人脸数据存在分布偏斜的问题,尤其是在常用的国际通用人脸数据库中,白人样本的数量显著多于亚洲和非洲人种样本,导致模型在不同族群上的表现不平衡。对于非白人族群,误检率可能高出 3 至 4 倍。现有的参数调整和特征提取方法多以主流人种为参考,进一步加剧了模型的群体偏差。此外,在极端条件下(如超低分辨率和高度遮挡),表情识别的准确率依然难以保障,表明网络模型的表达能力和泛化性仍需提升。同时,模型的鲁棒性、抗扰动能力、对抗样本的适应性也需进一步加强,以提高在复杂环境中的表现。

SSD(Single Shot MultiBox Detector)算法作为被广泛应用于表情识别领域的一种 one-stage 实时目标检测模型,结合了 YOLO 的回归检测思想和 Faster-RCNN 的候选框机制。通过多尺度特征图和先验框设计,SSD 直接预测目标的类别和边界框,大幅降低了计算复杂度。不同尺度的特征图负责检测不同大小的目标:大尺度特征图用于检测小目标,小尺度特征图则用于检测大目标。SSD 采用 VGG-16 作为基础结构,通过卷积操作从多尺度特征图中提取特征,并利用非极大值抑制获得最终分类结果,兼顾了检测速度和精度的平衡。在复杂场景下的表情识别中,SSD 的多尺度特征图和先验框设计具有显著优势,是一种高效且准确的检测方法。

然而,作为一种表情识别算法,SSD 虽然具备诸多优势,但仍存在精度不足、鲁棒性欠佳等问题,难以适用于各种场景。为此,本文首先选取了 FERplus 和 RAF-DB 作为测试数据集。FERplus 数据集来源于 FER-2013 数据集,是一个包含面部表情的情感识别数据集,提供了不同情绪的面部图像,广泛应用于表情识别任务。RAF-DB 数据集来源于“北京航空

航天大学面部表情数据集”(Beijing University of Aeronautics and Astronautics Facial Expression Database),该数据集包含了大量的面部表情图像,覆盖了多种表情类别,且具有较高的标注质量,广泛应用于面部表情识别研究。本文对 SSD 算法进行了分析与优化,并提出了一种基于 SSD 的网络模型重构与中心损失优化算法。该算法首先改进了其网络架构,增加了包含更多信息的特征层,减少冗余层,从而提升了算法的效率与识别精度。同时,对损失函数进行了优化,以增强对面部表情特征的辨别能力,从而提升算法的整体性能。通过这些改进,有望提高面部表情检测的准确性和鲁棒性。

2 本文方法

为提升 SSD 算法的准确性、速度和鲁棒性,本文对其网络结构进行了重构,并改进了损失函数,将两者结合提出新的表情识别算法。新的网络结构有效减少了冗余特征对识别的干扰,从而提升了算法的精度和识别速度;改进后的损失函数进一步增强了算法在恶劣环境下的稳定性及对非主流人种的适应能力。

2.1 SSD 网络模型重构

首先,针对 SSD 网络模型进行了优化与重构,删除对表情特征提取效果较弱的 conv8 和 conv9 层,保留了特征提取能力更强的卷积层,以减少冗余特征的干扰,从而提升识别的准确性和处理速度。其次,在基础层的特征提取过程中,保留了更多细节信息,以便更好地检测细微特征,提升在恶劣条件下的识别成功率并增强算法的鲁棒性。最后,针对原有损失函数进行了改进,引入了损失调节参数,在基础损失函数中调节原始损失与中心损失的比例。这一优化显著增强了模型在复杂环境下的稳定性,并有效减小了不同人种之间的特征提取差异。

本文基于深度学习框架 Caffe,使用 SSD 目标检测算法作为网络模型训练的基础框架。如图 1 所示,基于 Caffe 深度学习框架构建的 SSD 表情识别算法采用了 conv4_3 层、fc7 层、conv6_2 层、conv7_2 层和 conv9_2 层作为特征提取层。其中 conv4_3 层、fc7 层作为基础层,conv6_2 层、conv7_2 层、conv8_2 层和 conv9_2 层作为辅助层。

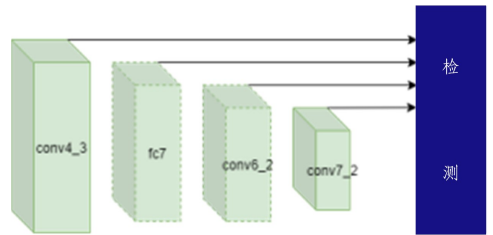


图 1 SSD 特征提取层

Fig. 1 SSD feature extraction layer

SSD 算法的特征提取能力随着网络层次的增加逐渐减弱,且特征图的分辨率也随之降低。为了提升其在面部表情检测中的表现,本文对网络结构进行了优化。具体而言,保留了特征提取能力较强的中间层和浅层,而删除了 conv8 层和 conv9 层,这两层在表情特征提取中的效果较弱。删除这两层的原因如下:conv8 层和 conv9 层的输出特征图分辨率过低,分别为 5×5 和 3×3 ,无法有效捕捉面部表情的细微变

化,如嘴角弯曲和眼睑收缩等微小的表情变化;其次,conv9层的感受野过大,已达到 450×450 像素,远超人脸区域的实际尺寸,导致其特征响应覆盖到背景区域,难以聚焦于局部的肌肉运动变化;此外,实验表明,这两层对表情关键区域的响应强度显著下降,进一步证明它们对表情识别贡献较小。删除这两层后,能够减少18%的计算量,提升1.2G FLOPs运算量,并提升推理速度23%,从而显著提升了模型的计算效率和实时性。通过这一剪枝优化,SSD网络能够更加聚焦于关键表情特征,减少冗余特征对分类结果的干扰。删减冗余层后的SSD架构如图2所示。

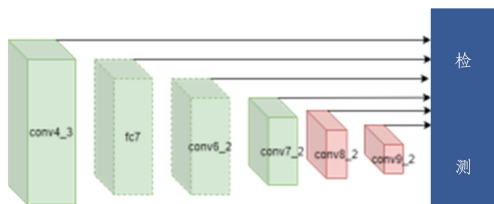


图2 删减冗余层后的SSD架构

Fig.2 SSD architecture after redundancy reduction

SSD通过多尺度特征层的级联预测增强了网络在不同形态和尺寸目标检测中的覆盖能力。接近底层的特征层保留了较多细节信息,但由于多次卷积和池化操作,该层的特征图每个网格像素对应的感受野较大,导致其难以捕捉恶劣环境中的细节特征,进而影响在不同光照、背景、年龄和种族条件下的检测表现。尽管这是SSD架构中细节保留最丰富的层,但表情检测任务的鲁棒性仍受到一定限制。

目前,大多数目标检测架构仅能有效检测约300像素的小目标。为了平衡检测速度与精度,本文提出了一种适用于实时场景的改进策略,保留更多细节信息并增强底层特征层的表现,同时提高对不同尺度目标的覆盖能力。

为了在删减深层冗余网络结构的同时,进一步提升模型对细微表情特征的捕捉能力,尤其是在复杂场景和小尺度人脸下的检测鲁棒性,本文对SSD的浅层特征提取部分进行了增强。

原始SSD网络使用conv4_3作为最底层的特征图,其感受野较大,且经过多次下采样后,对于面部肌肉的微小运动(如嘴角轻微上扬、眉梢微蹙等)所对应的底层细节信息(如纹理、边缘)保留不足,而这些细节特征对于区分相似表情(如“厌恶”与“愤怒”)至关重要。为此,本文引入分辨率更高、细节更丰富的conv3_3层作为新增的基础特征层。该层位于网络更浅处,具有以下优势:

1)更高的空间分辨率。conv3_3的特征图尺寸更大,每个网格点对应的原图区域更小(相当于 4×4 像素),使其对表情的局部细微变化更加敏感。

2)更强的细节表征能力。浅层网络天然擅长捕获边缘、角点、纹理等底层特征,这些正是构成细微表情的基本单元。

3)优化小目标检测。在远距离或部分遮挡情况下,人脸区域在图像中占比小,高分辨率的conv3_3层能有效防止表情关键区域的特征在初始阶段即被淹没。

通过引入conv3_3,并与原有的conv4_3及更高层语义特征进行融合,本文构建了一个跨度更大、兼顾细节与语义的多尺度特征金字塔,从而在不显著增加计算负载的前提下,显著增强了模型对复杂环境下表情特征的判别能力。在此基础上,本文还保留了SSD架构中顶层的特征用于级联预测。人

脸在恶劣环境或不同人种条件下对细节信息非常敏感,而顶层特征则提供了丰富的语义信息,有助于网络理解全局图像内容。本文的特征融合方案跨度较大,专门针对人脸检测进行了设计,新增了两个靠近底层的特征层,以实现细节与语义信息的兼顾。在确保网络计算成本可控的前提下,该改进方案显著提升了对小目标的检测效果。

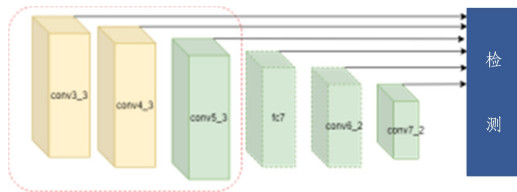


图3 展开信息细节的SSD架构

Fig.3 SSD architecture for unfolding detailed information

2.2 损失函数设计

SSD目标检测算法的损失函数由位置误差和置信度误差的加权和构成,如式(1)所示。该损失函数为非负实值,用于衡量模型与训练样本的匹配程度。损失值越小,模型的鲁棒性越强。中心损失(Center Loss)是一种旨在优化Softmax Loss的损失函数,主要解决同类别特征之间距离过大的问题。它通过为每个类别维护一个特征中心点,在特征提取过程中惩罚样本与类别中心之间的距离,从而增强同类特征的聚集能力。为有效获取这些中心点,网络在每个batch的迭代中随机生成并更新类别中心,计算特征与中心点之间的距离平方和。这一方法提升了模型在特征提取过程中的一致性和分类性能。

$$L(x, c, l, g) = \frac{1}{N} (L_{\text{conf}}(x, c) + aL_{\text{loc}}(x, l, g)) \quad (1)$$

在此基础上,本文对上一节提出的SSD网络模型进行了进一步优化。原始模型使用的Softmax Loss损失函数主要要求特征在最后一层分类器中正确分类。尽管Softmax Loss能够约束不同类别之间的特征,但它不能有效约束同一类别特征之间的关系。相比之下,Center Loss通过计算样本特征与其所属类别中心点之间的欧氏距离的平方和,能够约束同一类别特征之间的距离。如果距离过大,样本将受到惩罚,以确保同一类别的特征与中心点之间的距离足够小,从而提升特征的聚合能力。

基于此,本文结合Center Loss与原始损失函数,引入参数 λ 用于调节两者的比例,提出了基于改进网络模型的Center Loss SSD面部表情识别算法——IML-SSD。最终的损失函数如下所示:

$$\begin{aligned} L &= L(x, c, l, g) + \lambda L_C \\ &= \frac{1}{N} (L_{\text{conf}}(x, c) + aL_{\text{loc}}(x, l, g)) + \frac{\lambda}{2} \sum_{i=1}^m \|x_i - C_{y_i}\|_2^2 \end{aligned} \quad (2)$$

其中, λ 表示原始损失函数 $L(x, c, l, g)$ 和中心损失函数 L_C 的比例权重,具体参数设置通过实验设置。

3 实验设置与结果分析

3.1 数据集

本文对改进后的SSD面部表情识别算法进行了深入的实验分析,基于FERPlus和RAF-DB数据集,从数据预处理、模型训练及模型测试表现等多个方面全面评估了算法的改进效果。FERPlus数据集以其丰富的多样性和复杂性,为算法

提供了面部角度、光照变化和遮挡等多维度挑战;而 RAF-DB 数据集则通过情绪逐渐变化的序列,为模型理解和识别情绪提供了有力的支持。这两者的结合使得实验能够在识别模型的鲁棒性和泛化能力方面获得更全面的反馈。具体如下:

FERPlus^[14]:该数据集是 2016 年在 FERPlus 数据集的基础上重新标注并扩展的情感识别数据集。该数据集包含



图 4 FERPlus 数据集中的基础表情分类

Fig. 4 Basic facial expression classification in the FERPlus dataset

RAF-DB^[15]:该数据集是一个面向情感识别的公开人脸图像数据集,于 2017 年发布,共包含 29 672 张从网络采集的真实人脸图像。该数据集分为单标签和多标签子集,其中单标签子集用于情感分类实验,包含 7 种基本情绪:快乐、中性、惊讶、愤怒、悲伤、恐惧和厌恶。实验中使用的单标签子集由 12 271 张训练样本和 3 068 张测试样本组成。每张图像均提供了 5 个精确的人脸关键点(如眼角、嘴角等),并包含年龄范围和性别标注,以便进行进一步的细化分析和模型训练。此外,RAF-DB 还包含复杂情绪组合的多标签子集,适用于更高难度的情绪识别任务。

在 FERPlus 和 RAF-DB 测试集中,面部表情识别的特征各具特点,影响着算法的评估和改进。FERPlus 数据集的多样性体现在生气、害怕、厌恶、高兴、悲伤、惊讶和中性 7 种表情,同时图像质量较低,存在角度和光照变化的复杂性,面临识别过程中遮挡物和细节缺失的挑战。相较之下,RAF-DB 数据集提供高质量的图像和从弱到强的渐变表情序列,要求模型捕捉动态变化,并涵盖不同年龄段参与者的多样情感表达。这些特征为验证面部表情识别算法的效果提供了重要的参考,有助于鉴别算法的识别精度和鲁棒性。本文以这两个测试集为实验基础,验证改进 SSD 算法在复杂环境中人脸识别的精度与鲁棒性。

3.2 实验配置及评估指标

本实验环境配置如表 1 所列。初始学习率设为 0.000 1,迭代总数为 120 000,每 12 000 次迭代触发“stepValue”策略,将学习率按 0.1 倍逐步衰减。*momentum* 为 0.9,*weight_decay* 为 0.000 5,以减少过拟合风险。图像尺寸设为 227 × 227,确保了数据输入的规范性和训练的稳定性。

表 1 实验环境配置

Table 1 Experimental environment configuration

名称	环境配置
操作系统	Ubuntu20.04
处理器	Intel(R)Xeon(R)Gold 6248 CPU @ 2.50 GHz
CUDA 版本	8.0
cuDNN 版本	cuDNN 5.1
深度学习框架	Caffe
Python 版本	3.9
显存	80 GB

实验中,本文采用各类别准确率的平均值作为模型训练效果的评估标准。在目标检测领域,识别精度通常通过平均

28 709 幅训练样本、3 589 幅验证样本和 3 589 幅测试样本,所有图像均来源于谷歌搜索引擎。FERPlus 在原有 7 种基本情绪标签(快乐、中性、惊讶、愤怒、悲伤、恐惧和厌恶)的基础上新增了“蔑视”类别,使情感分类更加丰富。此外,FERPlus 采用多标签标注策略,由多个标注者对每张图像进行打分,提升了标签的准确性和情感细微差异的捕捉能力。

准确率 (mean Average Precision, mAP) 来衡量,该值由精度和召回率计算得出。精度 (Precision) 表示预测为正例的样本中,实际为正例的比例;而召回率 (Recall) 则表示在某一类别下,正确预测的样本占该类别所有样本的比例。通过精度和召回率计算每个类别的精度值,再对所有类别的精度值求平均,得到最终的 mAP 值。

$$mAP = \frac{\sum_{k=1}^N P(k) \Delta r(k)}{m} \quad (3)$$

其中, N 表示测试集中图像的数量, $P(k)$ 表示识别 k 个图像的精度值, $\Delta r(k)$ 表示从 $k-1$ 到 k 的识别图像数量变化时 Recall 的变化量, m 为所有图像的类别数。

3.3 实验结果分析

本节主要围绕 IML-SSD 算法进行了一系列的实验和比较,以提升面部表情识别的准确性。在对比实验中,本文 IML-SSD 算法与原始 SSD 算法进行了详细的性能评估,以验证中心损失函数在提高分类精度方面的有效性。

首先对改进损失函数比例权重 λ 进行测试,以确定 $L(x, c, l, g)$ 与 L_c 之间的最优权重。本文设置了多种权重值并在 RAF-DB 数据集上进行了测试,具体结果如表 2 所列。实验结果表明,当比例权重 $\lambda=0.5$ 时,训练得到的模型在测试集上 mAP 值最高,得出最佳权重比例值为 0.5。因此,本节提出的改进损失函数的最佳比例权重设为 0.5,后续所有实验中 λ 均为 0.5。

表 2 不同权重值下的 mAP 值

Table 2 mAP value under different weight

权重	mAP
0.1	0.855 7
0.3	0.882 3
0.4	0.877 1
0.5	0.905 0
0.55	0.885 7
0.6	0.861 8
0.75	0.872 8
1	0.870 9

确定 λ 的值后,将 IML-SSD 与原始 SSD 算法在 FER-Plus 数据集上进行了检测准确度的对比,结果如表 3 所列。IML-SSD 方法在各个类别的检测精度上均优于原始 SSD 算法,尤其在“厌恶”和“害怕”表情上的准确率提升最多。IML-SSD 实现了 72.4% 的 mAP 值,其中“高兴”类别的准确度达

到了 90.3%,远高于 SSD 检测模型的 83.5%,精确度提升了 6.8 个百分点。IML-SSD 在多个类别的表情识别中相较于 ViT-CL 展现了显著的优势,尤其在“高兴”类别的准确度上,IML-SSD 达到了 90.3%,明显高于 ViT-CL 的 85.7%。这一

提升源于中心损失函数的引入,使 IML-SSD 在不同类别表情的特征分布上有所改进,从而使同类表情特征更加聚合、不同类别表情特征更加分离,显著提高了表情特征的判别能力,从而提升了面部表情检测的准确度。

表 3 不同模型在 FERPlus 数据集上的检测准确度对比

Table 3 Comparison of detection accuracy of different models on the FERPlus dataset

模型	mAP	生气	厌恶	害怕	高兴	悲伤	惊讶	正常
SSD	0.659	0.652	0.538	0.417	0.835	0.633	0.823	0.7210
IML-SSD	0.724	0.709	0.697	0.587	0.903	0.622	0.830	0.7180
ViT-CL	0.634	0.636	0.560	0.554	0.800	0.644	0.616	0.5509

IML-SSD 与原始 SSD 算法在 RAF-DB 数据集上进行了检测准确度的比较,结果如表 4 所列。在 RAF-DB 数据集上,IML-SSD 依然表现出色,整体优于原始 SSD 算法。IML-SSD 达到了 70.4%的 mAP 值,其中“高兴”类别的准确度为

88.3%,显著高于 SSD 检测模型的 81.5%,提升了 6.8 个百分点,而 ViT-CL 的 mAP 值为 68.1%，“高兴”类别的准确度为 83.2%。虽然 IML-SSD 个别类别上的精度有略微下降,但对整体准确度的影响微乎其微,可以忽略不计。

表 4 不同模型在 RAF-DB 数据集上的检测准确度对比

Table 4 Comparison of detection accuracy of different models on the RAF-DB dataset

模型	mAP	生气	厌恶	害怕	高兴	悲伤	惊讶	正常
SSD	0.639	0.6320	0.518	0.397	0.815	0.613	0.803	0.701
IML-SSD	0.704	0.6890	0.677	0.567	0.883	0.602	0.810	0.708
ViT-CL	0.679	0.6567	0.640	0.590	0.880	0.586	0.799	0.600

此外,探讨了 IML-SSD 在不同噪声比例下的鲁棒性表现,并与 SSD 原算法进行了对比。按照文献[16-18]的方法生成噪声标签,将数据集的标签按不同噪声比例随机更改为其他类别,然后在生成的噪声数据集上重新实现了 SSD 原算法,以确保公平比较。实验结果取最后 5 轮迭代的精度平均值作为性能评估指标。如表 5 所列,IML-SSD 在不同噪声比例下均表现出较好的鲁棒性。例如,在 10%的标签噪声下,IML-SSD 在 RAF-DB 和 FERPlus 数据集上的精度分别为 69.28%和 72.26%,相比 SSD 原算法分别提升了 6.13 个百分点和 4.27 个百分点;在 20%噪声比例下,IML-SSD 在 RAF-DB 和 FERPlus 上的精度分别达到 67.15%和 69.56%,比 SSD 原算法分别高出 7.36 个百分点和 3.21 个百分点。这些结果表明,IML-SSD 在处理噪声数据方面具有显著的鲁棒性,即使在较高噪声情况下,模型依然能够稳定处理数据并获得良好结果。

表 5 不同模型的鲁棒性对比

Table 5 Comparison of robustness of different models

模型	噪声比例/%	RAF-DB	FERPlus
SSD	10	63.15	67.99
IML-SSD	10	69.28	72.26
ViT-CL	10	65.33	70.94
SSD	20	59.79	66.35
IML-SSD	20	67.15	69.56
ViT-CL	20	61.52	69.12

图 5 和图 6 是部分实验结果截图,方框左上角为识别的表情类型。具体来说,本文首先使用数据集上的照片进行表情检测,其次将系统开发成在线视频表情检测系统。如图 6 所示,IML-SSD 可以实现在线视频表情检测。

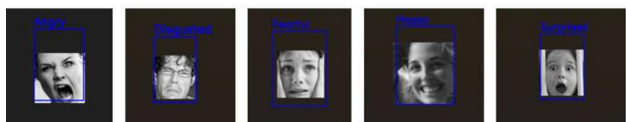


图 5 样本输入图像的表情识别

Fig. 5 Emotion recognition of sample input image

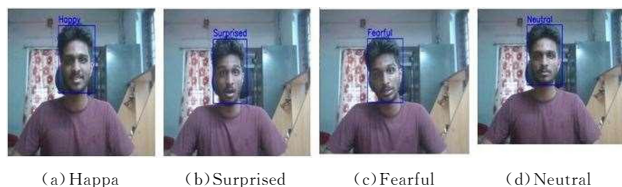


图 6 基于实时视频的表情识别

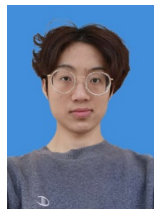
Fig. 6 Emotion recognition based on real-time video

结束语 面部表情检测技术是人工智能情感识别的重要组成部分。本文在深度学习中的 SSD 算法基础上进行了改进,通过增加有效层、删除无效层,提高了表情检测的准确性和算法的鲁棒性。此外,本文对损失函数进行了优化,使同类特征更紧密地聚集于其中心点,不同类别特征之间的分布更加分离,从而增强了表情特征的辨别能力。为验证所提 IML-SSD 算法的有效性,本文在不同测试集上进行了性能评估,结果表明,所提算法优于基础 SSD 算法。

参考文献

- [1] LI X, YANG M, DU Y T. Occlusion face recognition based on partition selection and gabor wavelet [J]. *Microelectronics & Computer*, 2023, 40(5): 39-46.
- [2] ZHOU H, LIU J H, LIU Z W, et al. Rotate-and-render unsupervised photorealistic face rotation from single-view images [C] // *Proceedings of International Conference on Computer Vision and Pattern Recognition*. Amsterdam: Elsevier, 2020.
- [3] ZHANG Y S, NIE Z Y, SUI L L. Deep Learning-based Facial Recognition in Complex Mining Areas [J]. *Energy Science and Technology*, 2022, 20(5): 3-8.
- [4] WEI Y, LIU M, WANG H, et al. Learning flow-based feature warping for face frontalization with illumination inconsistent supervision [C] // *Proceedings of European Conference on Computer Vision*. Berlin: Springer, 2020.
- [5] ZHANG R R, MIN G, LUO T, et al. Face Recognition Algorithm Based on Deep Neural Networks Model Introducing Image

- Denoising Processing Mechanism[J]. Digital Printing, 2022(5): 26-36.
- [6] YAO H X, DENG W H, LIUH H, et al. An overview of research development of affective computing and understanding [J]. Journal of image and graphics, 2022, 27(6): 2008-2035.
- [7] WANG K, PENG X J, YANG J F, et al. Suppressing Uncertainties for Large-Scale Facial Expression Recognition[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020.
- [8] ZHAO Z Q, LIU Q S, ZHOU F. Robust Lightweight Facial Expression Recognition Network with Label Distribution Training [C]//AAAI-21 Technical Tracks 4, 2021.
- [9] ZENG D, LIN Z, YAN X, et al. Face2exp: Combating data biases for facial expression recognition[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022; 20291-20300.
- [10] XIA M, ZHENG H, PENG H, et al. Enhanced multi-Scale Dynamic Facial Expression Recognition via Conditional Random Fields[J]. The Visual Computer, 2024, 10; 1-12.
- [11] LI J, LIU Z, WANG Z, et al. FERmc: Facial expression recognition framework based on multi-branch fusion and depthwise separable convolution [J]. Information Fusion, 2025, 124; 103416-103416.
- [12] JIAO Z, FU B, MAO Y, et al. Emotion Recognition Method Based on Multiscale Attention Residual Network [J]. Pattern Recognition and Image Analysis, 2025, 34(4); 1000-1006.
- [13] PENG C, SUN M, ZOU K, et al. Facial Expression Recognition-You Only Look Once-Neighborhood Coordinate Attention Mamba: Facial Expression Detection and Classification Based on Neighbor and Coordinates Attention Mechanism [J]. Sensors, 2024, 24(21); 912-6912.
- [14] GOODFELLOW I J, ERHAN D, CARRIER P L, et al. Challenges in Representation Learning: A Report on Three Machine Learning Contests [J]. Lecture Notes in Computer Science, 2013, 8228(1); 125-132.
- [15] LUCEY P, COHN J F, KANADE T, et al. The Extended Cohn-Kanade Dataset (RAF-DB): A complete dataset for action unit and emotion-specified expression [C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW 2010). 2010.
- [16] FARZANEH A H, QI X. Facial expression recognition in the wild via deep attentive center loss [C]// Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021; 2402-2411.
- [17] ZHANG Y, WANG C, DENG W. Relative uncertainty learning for facial expression recognition [J]. Advances in Neural Information Processing Systems, 2021, 34; 17616-17627.
- [18] SHE J, HU Y, SHI H, et al. Dive into ambiguity: Latent distribution mining and pairwise uncertainty estimation for facial expression recognition [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021; 6248-6257.



CHEN Ping'an, born in 1997, postgraduate. His main research interests include object detection and image segmentation.



DENG Qi, born in 1982, Ph.D. His main research interests include artificial intelligence and financial technology.