

孔垂见, 余德伟, 夏道勋. 融合坐标注意力与自适应加权学习的数据非均衡微表情识别方法研究[J]. 智能计算机与应用, 2026, 16(2): 15-23. DOI: 10.20169/j.issn.2095-2163.25082103

# 融合坐标注意力与自适应加权学习的数据非均衡 微表情识别方法研究

孔垂见<sup>1</sup>, 余德伟<sup>1</sup>, 夏道勋<sup>2</sup>

(1 贵州师范大学 大数据与计算机科学学院, 贵阳 550025; 2 贵州师范大学 网络空间安全学院, 贵阳 550025)

**摘要:** 现有典型的微表情识别数据集存在严重的数据样本分布不均衡问题, 数据集中的快乐和中性微表情样本占比较大, 而恐惧和厌恶的微表情样本却很少, 属于微表情子类中的小样本问题。这种数据失衡严重制约了微表情识别模型的多类别表征学习能力, 现有方法因面部关键特征提取不完整, 导致识别准确率受限。为此, 本文基于 ConvNeXt V2 架构提出了一种改进的微表情识别模型。具体贡献包括: 改进了坐标注意力机制, 强化模型对全局特征的提取能力; 引入一种自适应加权损失函数, 提升模型对非均衡数据的鲁棒性。该模型旨在强化全局特征提取能力并解决数据失衡问题, 从而有效提升模型的识别性能。实验结果表明, 本文提出的模型在公共表情识别数据集 FERPlus 和 RAF-DB 上达到了 92.96% 和 90.87% 的识别准确率, 具有良好的鲁棒性和泛化能力。

**关键词:** 微表情识别; 重新加权; 注意力机制; 损失函数

中图分类号: TP391.4

文献标志码: A

文章编号: 2095-2163(2026)02-0015-09

## Integrating coordinate attention and adaptive weighted learning for imbalanced Micro-Expression Recognition

KONG Chuijian<sup>1</sup>, YU Dewei<sup>1</sup>, XIA Daoxun<sup>2</sup>

(1 School of Big Data and Computer Science, Guizhou Normal University, Guiyang 550025, China;

2 School of Cyber Science and Technology, Guizhou Normal University, Guiyang 550025, China)

**Abstract:** In the context of computer vision, existing typical micro-expression recognition datasets suffer from a severe issue of imbalanced data sample distribution. Samples of happy and neutral micro-expressions occupy a larger proportion in these datasets, while samples of fear and disgust are scarce, representing a few-shot problem within micro-expression subcategories. This data imbalance significantly restricts the multi-class representation learning capability of Micro-Expression Recognition models. Due to incomplete extraction of facial key features, existing methods exhibit limited recognition accuracy. To address this problem, this paper proposes an improved micro-expression recognition model based on the ConvNeXt V2 architecture. Specific contributions include: enhancing the coordinate attention mechanism to strengthen the model's ability to extract global features; introducing an adaptive weighted loss function to improve the model's robustness to imbalanced data. The model aims to enhance global feature extraction and mitigate data imbalance, thereby effectively improving recognition performance. Experimental results demonstrate that the proposed model achieves recognition accuracy of 92.96% and 90.87% on the public expression recognition datasets FERPlus and RAF-DB, respectively, exhibiting strong robustness and generalization capability.

**Key words:** facial expression recognition; re-weighting; attention mechanism; loss function

## 0 引言

微表情识别 (Micro-Expression Recognition, MER) 是指利用计算机提取检测到的人脸特征, 以

便计算机能够理解人类情绪产生的表情并对人们的需求做出响应。研究旨在通过分析人类面部瞬间、无意识流露的细微表情, 来探测和解读其内心的真实情绪、意图或隐藏的信息。对于自动情感分析系

**基金项目:** 国家自然科学基金 (62566013); 贵州省科学计划项目 ([2022]4054); 贵阳市科技计划项目 ([2023]48-9 号)。

**作者简介:** 孔垂见 (2000—), 女, 硕士研究生, 主要研究方向: 教育大数据; 余德伟 (1999—), 男, 硕士研究生, 主要研究方向: 教育大数据。

**通信作者:** 夏道勋 (1980—), 男, 博士, 教授, 博士生导师, 主要研究方向: 计算机视觉, 教育大数据。Email: dxia@gznu.edu.cn。

收稿日期: 2025-08-21

统而言,尤其在应对光照变化和非线性视角变化等实际情况时,这一研究尤其重要。许多关于微表情识别的研究已经取得显著成果。因此,微表情识别在医疗诊断、公共安全以及智慧教育等多个领域有着广泛的应用。

近年来,不受约束的表情识别数据集,即直接从现实世界或互联网等渠道收集到的、更接近现实中复杂环境的数据集,如 RAF-DB<sup>[1]</sup>、FER2013<sup>[2]</sup>、FERPlus<sup>[3]</sup>以及 AffectNet<sup>[4]</sup>等大量的人脸图像为算法提供了数据支持。然而,对于从互联网上收集的表情识别数据集而言,获取足够的负面表情(如厌恶、恐惧和愤怒)极为困难,主要原因是这些负面表情的细微差异和难以准确标注。具体来说,负面表情通常伴随着相同的肌肉变化,因此标注者很难以高置信度确定标签,这直接导致负面表情数量的减少,使得在自然环境下收集的表情识别数据集中存在样本分布严重失衡的问题,如图1所示。现有关于微表情识别<sup>[1,5]</sup>的研究旨在通过监督信号提升表情表示的区分能力。然而,为了达到这一目的,卷积神经网络(Convolutional Neural Networks, CNN)泛化能力受到数据样本分布不均衡的影响。因此,如何解决微表情识别中的数据失衡问题,值得进一步探索。此外,在复杂场景下的人脸表情常常处于遮挡状态,遮挡部分丢失的表情信息不可避免地会对表情表征的学习产生影响,无法获取有效的表情特征。

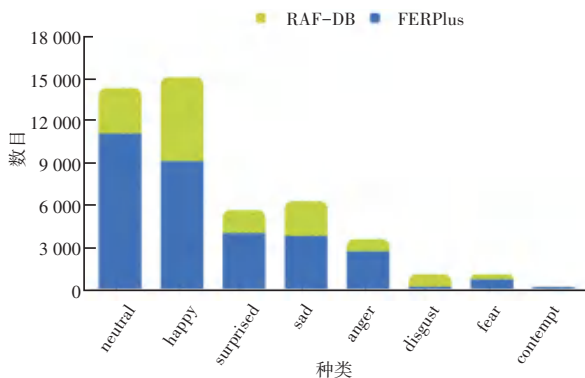


图1 RAF-DB和FERPlus数据集的数据样本分布不均衡对比图

Fig. 1 Comparison of data imbalance distribution of RAF-DB and FERPlus datasets

针对上述问题,本文基于 ConvNeXt V2 架构提出了一种改进的微表情识别模型。该模型致力于解决大规模数据集中的数据失衡问题(某些类别样本占比较多,而其他类别样本极少),并充分提取微表情的全局特征,而经过预训练的微表情识别模型在图像遮挡和姿势变化较大的情况下识别能力和优势也更为突出。

本文的贡献和创新点如下:

(1)提出了一种基于 ConvNeXt V2 架构的微表情识别模型,该方法解决了数据样本分布不均衡的问题,并能有效提取面部表情显著特征,在图像遮挡和数据失衡的情况下获得良好的识别率。

(2)引入了一个自适应加权损失函数,重新加权类别的权重系数以增强少数类别与其他类别之间的特征差异,即通过特征差异的自适应加权来解决数据失衡问题;同时加入改进的坐标注意力机制,实现了动态硬样本和类分布的表情识别。

(3)经过一系列的实验表明,本文提出的模型在公共表情识别数据集 FERPlus 和 RAF-DB 上准确率分别为 92.96% 和 90.87%,该模型有效提升了全局特征提取能力与在数据失衡场景下的识别准确率,具有良好的鲁棒性和泛化能力。

## 1 相关工作

在本节中,系统地梳理了与本研究密切相关的3个研究内容,并结合相关文献进行阐述。首先,介绍微表情识别的研究进展,总结了传统特征提取与深度学习方法;其次,是注意力机制的演进与应用,重点分析其在特征增强和关系建模中的优势。最后,阐述在非均衡数据中学习的方法,主要包含数据预处理、重新加权的损失函数和模型集成等方法,探讨其在数据失衡场景下的适用性。

### 1.1 微表情识别

随着深度学习的迅速发展,其在计算机视觉的研究领域,如目标检测<sup>[6-8]</sup>和高动态范围成像<sup>[9]</sup>等方面已取得了可观成就。然而在微表情识别方面,传统方法主要基于图像处理、特征提取和机器学习技术。例如,Zeng 等学者<sup>[10]</sup>提出了 Meta-Face2Exp 框架,利用大量未标记的表情识别数据来实现表情识别。Li 等学者<sup>[11]</sup>则采取卷积神经网络和支持向量机(Support Vector Machines, SVM)来识别表情。尽管这些方法能够实现微表情识别,但其识别准确率和模型的鲁棒性仍有待提高。在深度学习领域,卷积神经网络等模型已成为表情识别的主流方法。研究中能够从原始像素中自动学习更高级的特征,并在大规模数据集上取得重大进展,比如一些经典的深度学习框架,包括 VGG<sup>[12]</sup>、ResNet<sup>[13]</sup>和 Inception<sup>[14]</sup>等。Choi 等学者提出了 DCNN 集成分类器<sup>[15]</sup>,这是一种将深度学习模型集成起来以解决具有挑战性的表情识别问题并提高在自然环境下捕获的表情识别准确率的方法,其效果远优于仅使用

单一深度学习模型。为了解决单一特征表征能力不足的问题,王成宝等学者<sup>[16]</sup>提出了一种称为异构多视图特征融合网络(HMF-Net)的模型来进行微表情识别,通过互注意力融合能实现多视图信息的交互和补充,提高微表情识别的准确率。微表情识别中的数据失衡问题一直是众多计算机视觉领域面临的主要技术挑战。为了解决微表情识别中数据失衡、面部模糊和遮挡问题,Fan等学者<sup>[17]</sup>提出了CUNN模型;然而该模型的识别率仍有待提高。因此,解决数据失衡和面部关键特征缺失等问题的方法,仍旧值得探究学习。

## 1.2 注意力机制

注意力机制是一种重要的深度学习技术,在不同领域的模型中得到了广泛应用,用于增强模型对重要信息的关注和学习,从而提高模型的性能和准确率。Huang等学者<sup>[18]</sup>设计了一种交叉注意力机制,为每个像素收集其十字路径上的所有像素的上下文信息,并通过循环操作捕获整幅图像的依赖性。Guo等学者<sup>[19]</sup>设计了一种名为外部注意力的新型注意力机制,通过2个线性层和2个归一化层替代自注意力,实现线性复杂度。针对多模态信息冗余导致融合不充分和模型参数量大的问题,张海侠等学者<sup>[20]</sup>提出了一种基于稀疏注意力的多任务多模态情感识别模型。相比之下,本文的工作主要侧重于增强网络对不同空间位置的特征建模能力,从而更好地捕捉微表情中的全局信息并提升特征表达能力。

## 1.3 在不均衡数据中学习

在计算机视觉领域,针对现实数据集中存在的数据失衡问题,现有研究主要从3个角度进行解决,分别是数据预处理、重新加权的损失函数和模型集成。本文重点研究前2类方法,因为模型集成方法<sup>[21-22]</sup>通常计算成本较高。数据预处理方法通常采用重采样技术<sup>[23-24]</sup>,但研究<sup>[25-26]</sup>表明,重采样可能对表征学习产生负面影响,过采样可能引入重复样本,增加过拟合风险,欠采样则可能丢失有价值样本。另一种技术是数据增强<sup>[27-28]</sup>,通过对少数类样本施加预定义变换来扩充数据集。然而,由于微表情识别数据具有与表情相关的特定局部特征,寻找有效的增强方法具有挑战性。重新加权的损失函数<sup>[29-30]</sup>通过在训练过程中为不同类别或样本分配差异化权重,旨在为所有类别传播合适的梯度值。本文通过以下策略来解决数据集不均衡问题,在小批量训练中自适应估计权重系数,并通过引入调节

因子来调整样本数量和样本难易程度,以便模型专注于学习少数类。

## 2 基于ConvNeXt V2架构的微表情识别模型设计

本节将详细阐述所提出的微表情识别模型的核心方法。该模型以ConvNeXt V2为骨干网络进行构建,其核心在于2项关键技术。首先,是改进的坐标注意力机制,旨在精准捕捉面部关键区域的细微特征,强化模型的空间感知能力;其次,是自适应加权损失函数,用以提升模型对非均衡数据的鲁棒性。通过上述技术的协同作用有效解决了数据失衡和人脸遮挡的问题,并显著提升了模型在复杂场景下的识别准确率。

### 2.1 微表情识别模型框架

在微表情识别任务中,识别准确率往往会受到图像噪声、人脸遮挡和数据集样本分布不均衡的影响。因此,如何优化模型在微表情识别中的准确率和鲁棒性是值得进一步去深究的问题。本文从解决数据失衡问题的角度出发,同时考虑提取更多显著特征,以应对复杂场景下的微表情识别,提出了一种基于ConvNeXt V2架构,并融合了改进的坐标注意力机制与自适应加权损失函数的微表情识别模型。模型框架如图2所示,该模型包括了3个部分。首先,使用ConvNeXt V2<sup>[31]</sup>作为模型的骨干网,将预处理后的图像送到骨干网络中,通过卷积神经网络进行下一步的特征提取。然后,加入了改进的坐标注意力机制,通过注意力模块在一个空间方向上捕获长依赖关系,并增强网络的表现力。最后,引入自适应加权损失函数,用于解决数据失衡以及图像边际损失的问题。

### 2.2 改进的坐标注意力机制

当使用卷积神经网络处理图片时,如何让网络自适应地提取显著区域特征显得尤为关键,而注意力机制就是实现网络自适应关注显著区域的一个方式,其核心思想就是让模型定位到更需要关注的地方。因此,本文加入了改进的坐标注意力机制。与传统的坐标注意力机制不同之处在于,研究中将坐标注意力机制<sup>[32]</sup>与简单无参注意力机制<sup>[33]</sup>并行操作,通过水平和垂直方向的池化来分解空间信息,生成方向感知的注意力权重,并基于特征能量的无参注意力,增强显著区域。通过这种并联设计,模型能同时利用空间位置先验和自适应能量注意力,显著提升特征表达能力。

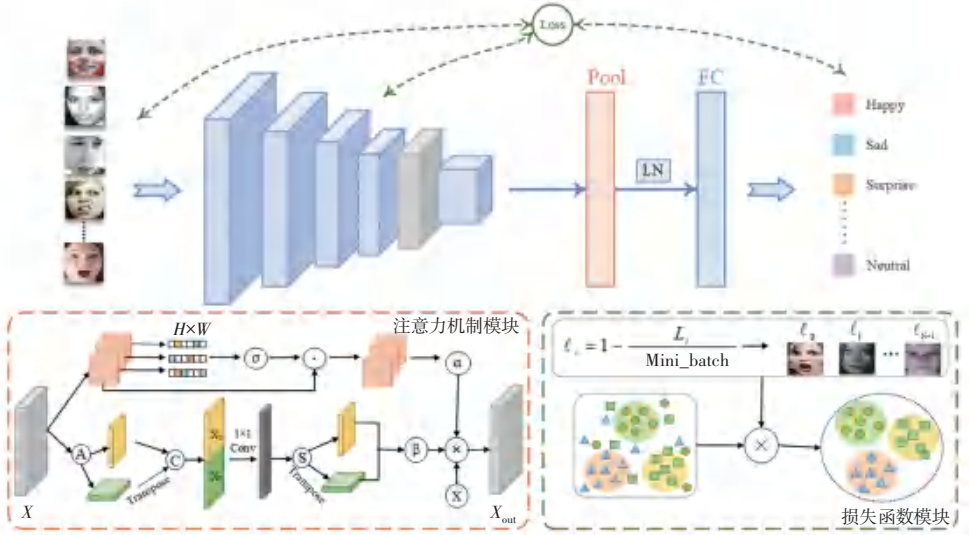


图2 模型框架示意图

Fig. 2 Schematic diagram of the model framework

对于坐标注意力机制,可以解决传统卷积无法捕获长距离空间依赖关系的问题,即通过捕获空间位置信息(水平和垂直)来动态调整特征图的权重,从而提高模型的性能。具体来说,设输入特征  $X \in \mathbb{R}^{B \times C \times H \times W}$ ,其中  $B$ 、 $C$ 、 $H$  和  $W$  分别表示是小批次大小、通道数、空间高度和宽度。并对输入特征进行水平和垂直两个方向的池化。对于水平方向,压缩宽度维度,保留高度方向的空间信息,反映垂直边缘;对于垂直方向则是压缩高度维度,保留宽度方向的空间信息,反映水平边缘。计算公式如下:

$$X_h = \frac{1}{W} \sum_{j=1}^W X[:, :, :, j] \in \mathbb{R}^{B \times C \times H \times 1} \quad (1)$$

$$X_w = \frac{1}{H} \sum_{i=1}^H X[:, :, i, :] \in \mathbb{R}^{B \times C \times 1 \times W} \quad (2)$$

其中,  $X_h$  表示水平方向池化后的特征,  $X_w$  表示垂直方向池化后的特征。

将水平和垂直信息融合后,通过卷积拼接和降维学习两者之间的交互关系,计算公式如下:

$$Y = \text{Concat}(X_h, X_w \in \mathbb{R}^{B \times C \times (H+W) \times 1}) \quad (3)$$

为了计算图像中每行和每列的重要性,计算其注意力图。首先,将得到的信息拆分为水平和垂直方向。这里,可表示为:

$$Y_h, Y_w = \text{Split}(Y, [H, W]),$$

$$Y_h \in \mathbb{R}^{B \times \frac{C}{r} \times H \times 1}, Y_w \in \mathbb{R}^{B \times \frac{C}{r} \times W \times 1} \quad (4)$$

然后,计算水平和垂直方向的注意力图,即:

$$A_h = \sigma(W_1 \cdot Y_h), A_w = \sigma(W_2 \cdot Y_w) \quad (5)$$

其中,  $A_w$  和  $A_h$  分别表示每行和每列的重要性,  $\sigma(\cdot)$  表示 Sigmoid 函数。最后,通过坐标注意力机制计算得到的输出特征为:

$$X_{\text{Coord}} = X \cdot A_h \cdot A_w \quad (6)$$

对于简单无参注意力机制,可以突出重要特征并抑制不重要的特征,即通过计算每个空间位置的“能量”来动态调整特征图的权重,从而增强网络的表现力。这里的“能量”是指对每个特征点,计算其与周围特征的统计差异。具体而言,这里设输入特征  $X \in \mathbb{R}^{B \times C \times H \times W}$ ,其中  $B$ 、 $C$ 、 $H$  和  $W$  分别表示小批次大小、通道数、空间高度和宽度。首先,计算空间均值,表示全局上下文,计算公式如下:

$$\mu = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X[:, :, i, j] \quad (7)$$

然后,计算特征图的能量,量化每个位置的“显著性”,且高能量区域代表需要增强的特征。能量计算函数如下:

$$E = \frac{(X - \mu)^2}{\sum_{i,j} (X[:, :, i, j] - \mu)^2 + 4 \left( \frac{1}{n} + \lambda \right)} + 0.5 \quad (8)$$

其中,  $n = H \times W - 1$ ,  $\lambda$  表示一个防止除零的极小数,本文取值为  $1e-8$ 。

进一步地,利用 Sigmoid 函数将能量映射到  $[0, 1]$ ,作为自适应权重对注意力图进行加权,自动增强高能量区域。则通过简单无参注意力机制计算得到的输出特征为:

$$X_{\text{Simam}} = X \cdot \sigma(E) \quad (9)$$

其中,  $\sigma(\cdot)$  表示 Sigmoid 函数。最后,通过并联计算得到的最终输出特征为:

$$X_{\text{CSA}} = \alpha \cdot X_{\text{Coord}} + \beta \cdot X_{\text{Simam}} \quad (10)$$

其中,  $\alpha$  和  $\beta$  表示可学习的参数,满足  $\alpha + \beta = 1$ 。

### 2.3 自适应加权损失函数

当数据集中存在严重的数据失衡时,一般的深度学习算法表现可能不太理想,受文献[34]的启发,文中引入了一个自适应加权损失函数来重新加权尾部表情表示的贡献。

具体而言,设输入 $Y=[y_1, y_2, \dots, y_M]$ 表示一个小批次的原始标签,并输出每种表情的权重系数,其公式为:

$$e_i = 1 - \frac{L_i}{M}, \quad i = 0, 1, 2, \dots, N-1 \quad (11)$$

其中, $L_i$ 表示第*i*个小批次中表情的数量; $N$ 表示最大标签索引; $M$ 表示小批次大小。在实验中分析可知,即使不使用大批次也能实现较好的性能,并且在后续实验部分会进行比较说明。

焦点损失是Lin等学者<sup>[35]</sup>针对单阶段目标检测器提出的改进方法。对于分类问题,最常用的损失函数是交叉熵损失函数,其平等对待所有类别,即在学习过程中赋予相同权重。而焦点损失的核心在于通过引入调节因子来改进交叉熵损失,从而动态调整样本数量与样本难易程度的影响,使模型更专注于学习少数类别。具体公式如下:

$$L_F = -\alpha_i(1 - P_i)^\gamma \log(P_i) \quad (12)$$

其中, $P_i$ 表示模型预测样本属于真实类别的概率,对于正样本 $P_i = P$ ,对于负样本 $P_i = 1 - P$ ;  $\gamma$ 和 $\alpha_i$ 是控制难易样本的权重差异,本文分别取值为0.1

和0.5。

最后,利用权重系数 $e_i$ 对不同类别之间的表示差异进行加权,使模型更多地关注尾部表情特征和其他表情特征之间的差异。最终模型的损失函数为:

$$L_{AF} = -e_i \alpha_i (1 - P_i)^\gamma \log(P_i) \quad (13)$$

## 3 实验结果与分析

在本节中,通过一系列实验对所提方法进行全方面评估,以验证模型的有效性和鲁棒性。具体而言,首先详细地介绍了实验设置,包括训练策略、超参数配置及硬件环境,并说明所使用的数据集及其划分方式。随后,通过系统的消融实验分析验证各模块的贡献,以阐明模型性能提升的来源。最后,为了进一步评估所提方法的先进性,在2个公开数据集上与其他先进方法进行了对比实验。

### 3.1 数据集

为了评估本文所提出的微表情识别模型的性能,本文使用了2个公开的大型人脸表情数据集(RAF-DB和FERPlus)进行实验。

(1) RAF-DB<sup>[1]</sup>:包含了12 271张训练图像和3 068张验证图像,这些图像由40名训练有素的人类编码员标注了基本或复合表情。该数据集包含7种表情(惊讶、恐惧、厌恶、快乐、悲伤、愤怒和中性),为其赋予了0~6的数值标签。每个表情对应的图像数量都已列出,见表1。

表1 FERPlus数据集和RAF-DB数据集的情感分布

Table 1 Sentiment distribution in FERPlus datasets and RAF-DB datasets

项目内容	标签							
	0	1	2	3	4	5	6	7
情感	自然	高兴	惊喜	伤心	生气	厌恶	害怕	蔑视
FERPlus	11 012	9 045	3 960	3 749	2 660	161	677	148
RAF-DB	3 204	5 957	1 619	2 460	867	877	355	—

(2) FERPlus<sup>[3]</sup>:这个数据集重新标注了FER2013数据集中错误标记的图像,并移除了几个未知图像。数据集中包含大约31 412张人脸图像,其中25 060张用于训练,3 199张用于验证,3 153张用于测试。由表1可知,FERPlus数据集包含了8种微表情,文中为每种表情赋予了0~7的标签,每个表情对应的图像数量也已列出。

### 3.2 实验设置

(1) 数据处理:在所有数据集的训练和验证过程中,每张图像的尺寸均调整为224×224像素,模型以 $2e-5$ 的学习率训练了60个周期。实验是在深度学习框架PyTorch上开发的,编程语言为Python 3.9。

(2) 实验环境:本文的实验环境包括硬件环境和

软件环境。硬件环境主要配置为服务器(Intel(R) Xeon(R) Gold5218CPU@ 2.30 GHz)和Quadro RTX8000显卡,显存48 GB)。使用的软件环境是Ubuntu18.04.5 LTS操作系统。

### 3.3 关于自适应加权损失函数中M的取值

$M$ 表示小批次中的微表情数量,与权重系数和损失函数的效果直接相关。文中研究了从6到60的不同 $M$ 值对模型准确率的影响,结果如图3所示。默认情况下, $M$ 设置为36并实现了最佳的分类性能。 $M$ 值越小,每个类别的数量就越少,尤其是负面表达的数量,这不利于算法的泛化,导致性能下降。此外, $M$ 值越高,性能下降越明显,因为算法过度考虑了负面情绪。

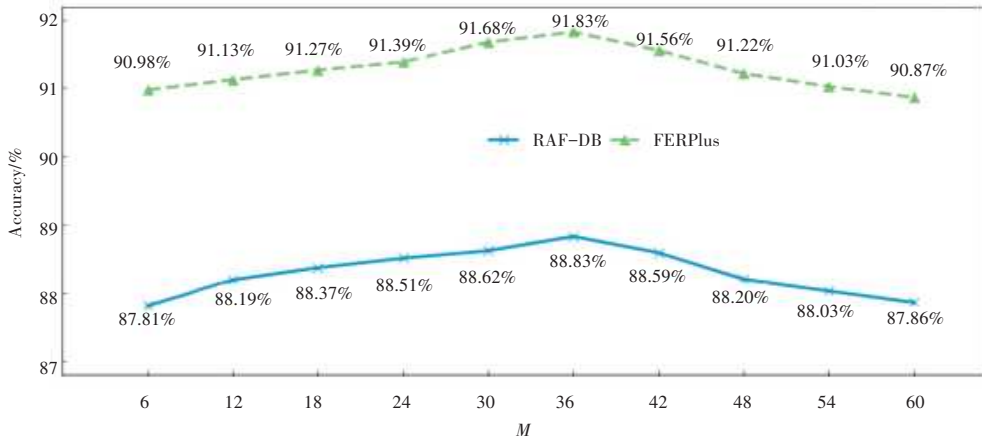


图 3 在 RAF-DB 和 FERPlus 上对小批量大小  $M$  进行评估

Fig. 3 Evaluation of the mini-batch size  $M$  in RAF-DB datasets and FERPlus datasets

### 3.4 消融实验

为了评估每个模块的贡献,在 RAF-DB 和 FERPlus 数据集上进行了消融实验。本文将改进的方法与 ConvNeXt V2 进行对比,并在 RAF-DB 和 FERPlus 数据集上进行验证,结果见表 2。ConvNeXt V2 在这 2 个数据集上的识别准确率分别为 87.35% 和 90.47%。在加入了自适应加权损失函数和改进的坐标融合注意力机制后,本文的新模型分别取得了 90.87% 和 92.96% 的准确率,较原始模型提升了 3.52% 和 2.49%。实验结果表明,本文所研发的每个改进模块都对模型性能的提升至关重要。更重要的是,这些改进并未显著增加模型的参数量,从而验证了本文改进的模块在提升性能的同时基本不会增加额外的计算复杂度。

表 2 不同方法在 FERPlus 和 RAF-DB 数据集上的准确率比较

Table 2 Comparison of accuracy of different methods in FERPlus datasets and RAF-DB datasets

坐标注意力 模块	损失函数 模块	准确率/%		参数/M
		RAF-DB	FERPlus	
×	×	87.35	90.47	8.76
×	✓	88.93	91.83	8.80
✓	×	88.54	91.56	8.92
✓	✓	90.87	92.96	8.98

### 3.5 对比实验

研究中将提出的微表情识别模型与微表情识别领域的其他先进方法进行比较,这些方法同样在公开数据集 FERPlus 或 RAF-DB 上进行了训练与验证。实验结果见表 3。表 3 中,加粗数值代表与前沿方法对比时的最优平均准确率,下划线数值则为次优结果。根据表 3 的数据显示,本文所提的模型在 FERPlus 和 RAF-DB 数据集上较现有最先进的

表情识别技术具有显著准确率优势。

表 3 与其他先进方法在数据集 FERPlus 和 RAF-DB 上的准确率对比

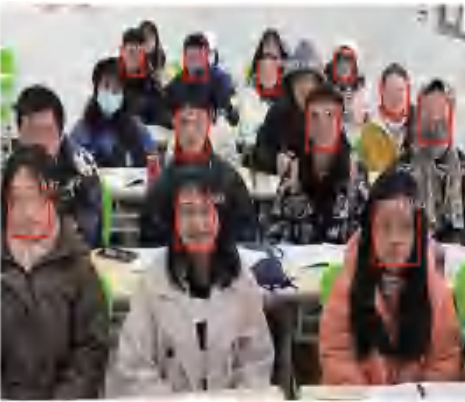
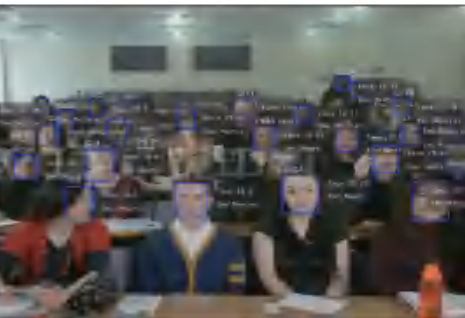
Table 3 Comparison with the accuracy of other advanced methods in FERPlus datasets and RAF-DB datasets

模型	年份	准确率/%	
		FER	RAF
ResNet <sup>[13]</sup>	CVPR 2016	83.85	74.87
ResNet+VGG16 <sup>[36]</sup>	URTC 2017	87.40	—
ResNet50 <sup>[37]</sup>	ACM 2018	88.80	—
GACNN <sup>[38]</sup>	TIP 2018	—	85.07
IPA2LT <sup>[39]</sup>	ECCV 2018	—	86.77
SCN <sup>[40]</sup>	CVPR 2020	88.01	87.03
RAN <sup>[41]</sup>	CVPR 2020	88.55	86.90
KTN <sup>[34]</sup>	TIP 2021	90.49	88.07
STSN <sup>[34]</sup>	TIP 2021	89.66	88.07
EfficientFace <sup>[42]</sup>	AAAI 2021	—	88.36
AVT <sup>[43]</sup>	ICIP 2022	90.45	—
AMP-Net <sup>[44]</sup>	TCSVT 2022	74.48	89.25
ResNet+PAL <sup>[45]</sup>	CVPR 2022	89.62	—
NAGNet <sup>[46]</sup>	TII 2023	89.30	—
SSF-ViT(B) <sup>[47]</sup>	TCSVT 2023	74.95	89.72
FG-AGR <sup>[48]</sup>	TCSVT 2024	91.09	<u>90.81</u>
LCANet <sup>[49]</sup>	TNNLS 2025	<u>91.43</u>	90.03
本文模型	2025	<b>92.96</b>	<b>90.87</b>

### 3.6 不同方法对不同表情的识别效果对比

为验证本文提出的微表情识别模型在复杂场景中行人脸检测和表情识别的性能,本文在自然复杂环境中进行了实验。研究中定义的复杂场景包含以下因素:环境背景干扰、后排人脸图像分辨率低、高图像噪声、环境光线不均、面部遮挡以及面部姿势

变化(如侧脸、托腮、头部倾斜、低头等)。实验结果表明,即使在如此复杂的环境中,本文提出的模型仍能成功检测并框选大部分人脸。为了更直观地展示提出的模型的识别有效性与先进性,与其他先进方法的表情识别效果做对比,结果如图4所示。其中,图4(c)为本文模型的识别结果,为了保护隐私,对人脸表情做了保护处理。总体而言,本文的模型在复杂场景中仍保持强大的鲁棒性,性能未出现显著下降。因此,本文提出的模型为复杂场景下数据失衡时微表情的识别提供了有效的技术解决方案。

(a) NAGNet<sup>[46]</sup>(b) LCA Net<sup>[49]</sup>

(c) 本文模型

图4 不同方法对不同表情的识别对比图

Fig. 4 Comparison results of different methods for recognizing different expressions

## 4 结束语

在本文中,基于 ConvNeXt V2 架构,提出了一种融合坐标注意力机制与自适应加权损失函数的微表情识别模型。用以解决微表情识别中的数据失衡和人脸遮挡等问题。自适应加权损失函数为少数类别赋予更高的权重,让模型能够更多地关注少数类别,有效地利用所有训练样本,提升了在数据失衡场景下的识别准确率。同时,改进的坐标注意力机制增强了显著特征的提取并抑制了无关特征,从而提升了模型提取复杂特征的能力。通过消融实验证明,将改进的坐标注意力机制和自适应加权损失函数融入模型对于提高模型的识别性能至关重要。实验结果表明,本文提出的模型在 FERplus 和 RAF-DB 数据集上的识别率分别提高了 2.49% 和 3.52%,这证明每个模块都对模型性能的提升有显著作用。与现有的最先进的微表情识别方法相比,本文提出的模型在公开数据集上的表现也非常出色,在 FERplus 数据集和 RAF-DB 数据集上分别达到了 92.96% 和 90.87% 的识别准确率。一系列实验表明,本文提出的方法有效解决了表情识别中的数据失衡和人脸遮挡问题,在复杂的场景下也能很好地识别微表情。在后续工作中,将继续探索在数据失衡和更复杂场景下的特征提取方法,以进一步提高在真实的复杂场景中微表情识别的准确率。

## 参考文献

- [1] LI Xue, ZHU Chunhua, ZHOU Fei, et al. Facial expression recognition via joint loss constraining attention - modulated contextual spatial information network[J]. *Multimedia Systems*, 2025, 31: 43.
- [2] JAE Y C, BUMSHIK L. Combining deep convolutional neural networks with stochastic ensemble weight optimization for facial expression recognition in the wild [J]. *IEEE Transactions on Multimedia*, 2023, 25: 100-111.
- [3] XU Jie, LI Yang, YANG Guanci, et al. Multiscale facial expression recognition based on dynamic global and static local attention[J]. *IEEE Transactions on Affective Computing*, 2024, 16 (2): 683-696.
- [4] TAN Yumei, XIA Haiying, SONG Shuxiang. Learning informative and discriminative semantic features for robust facial expression recognition[J]. *Journal of Visual Communication and Image Representation*, 2024, 98: 104062-104074.
- [5] MARAH H, FLORIAN B, PIA B, et al. Multi-task multi-modal self-supervised learning for facial expression recognition [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway, NJ: IEEE, 2024: 4604-4614.
- [6] XIA Daoxun, LIU Haojie, XU Lili, et al. Visible-infrared person

- reidentification with data augmentation viacycle – consistent adversarial network[J]. *Neurocomputing*,2021,443:35–46.
- [7] LIU Haojie, XIA Daoxun, JIANG Wei, et al. Towards homogeneous modality learning and multi-granularity information exploration for visible infrared person re-identification[J]. *IEEE Journal of Selected Topics in Signal Processing*, 2023, 17(3): 545–559.
- [8] HU Yaoqi, NIU Axi, SUN Jinqiu, et al. Dynamic center point learning for multiple object tracking under severe occlusions[J]. *Knowledge-Based Systems*, 2024,300:112130–112144.
- [9] YAN Qingsen, WANG Haishen, MA Yifan, et al. Uncertainty estimation in HDR imaging with Bayesian neural networks[J]. *Pattern Recognition*,2024,156: 110802–110814.
- [10] ZENG Dan, LIN Zhiyuan, YAN Xiao, et al. Face2Exp: Combating data biases for facial expression recognition [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway, NJ: IEEE, 2022: 20259–20268.
- [11] LI Ming, XU Hao, HUANG Xingchang, et al. Facial expression recognition with identity and emotion joint learning [J]. *IEEE Transactions on Affective Computing*, 2018, 12(2): 544–550.
- [12] KAREN S, ANDREW Z. Verydeep convolutional networks for large-scale image recognition [J]. *arXiv preprint arXiv*, 1409.0556, 2014.
- [13] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway, NJ: IEEE, 2016:770–778.
- [14] CHRISTIAN S, VINCENT V, SERGEY I, et al. Rethinking the inception architecture for computer vision [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway, NJ: IEEE, 2016: 2818–2826.
- [15] JAE Y C, BUMSHIK L. Combining deep convolutional neural networks with stochastic ensemble weight optimization for facial expression recognition in the wild [J]. *IEEE Transactions on Multimedia*, 2023, 25: 100–111.
- [16] 王成宝, 孙宁, 刘佳鑫, 等. 基于异构多视图特征融合的人脸微表情识别[J]. *智能计算机与应用*, 2025, 15(8): 25–35.
- [17] FAN Jiaxiang, ZHOU Jian, DENG Xiaoyu, et al. Combating uncertainty and class imbalance in facial expression Recognition [C]// *Proceedings of 2022 IEEE Region 10 Conference (TENCON)*. Piscataway, NJ: IEEE, 2022:1–4.
- [18] HUANG Zilong, WANG Xinggang, HUANG Lichao W, et al. CCNet: Criss-cross attention for semantic segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(6): 6896–6908.
- [19] GUO Menghao, LIU Zhengning, MU Taijiang, et al. Beyond self-attention: External attention using two linear layers for visual tasks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(5): 5436–5447.
- [20] 张海侠, 张星, 车雷, 等. 基于稀疏注意力的多任务多模态情感识别模型[J]. *智能计算机与应用*, 2025, 15(8): 18–24.
- [21] ASRIA K, LEE H Y, CHEN Yuling, et al. A machine learning-based ensemble model for estimating diurnal variations of nitrogen oxide concentrations in Taiwan [J]. *Science of The Total Environment*, 2024, 916: 170209.
- [22] DEBAJYOTI D, ARIJIT M, PARTHA P, et al. Emotion detection in social robotics: Empath – Obscura – an ensemble approach with novel face augmentation using SPIGA [C]// *Proceedings of 2023 Seventh IEEE International Conference on Robotic Computing (IRC)*. Piscataway, NJ: IEEE, 2023: 94–101.
- [23] ARNAUD L, SCOTT W, THIERRY C, et al. Estimating animal density using the space-to-event model and bootstrap resampling with motion-triggered camera-trap data [J]. *Remote Sensing in Ecology and Conservation*, 2024, 10(2): 141–155.
- [24] XIAO Yangyang, ZHANG Tiewi, CAO Juan, et al. Accelerated Lloyd’s method for resampling 3D point clouds [J]. *IEEE Transactions on Multimedia*, 2025, 27: 1033–1046.
- [25] XU Zhaozhao, SHEN Derong, KOU Yue, et al. A synthetic minority oversampling technique based on Gaussian mixture model filtering for imbalanced data classification [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(3): 3740–3753.
- [26] PERETTI S, JUDGE R, HINDMARCH I. Real-time tilt under sampling optimization during electron tomography of beam sensitive samples using golden ratio scanning and RECAST3D [J]. *ACTA Psychiatrica Scandinavica*, 2023, 15(11): 5391–5402.
- [27] SHI Jian, GHAZZAI H, MASSOUD Y. Differentiable image data augmentation and its applications: A survey [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024, 46(2): 1148–1164.
- [28] FRANK S, JURGEN S, ANDRE K. Multispectral snapshot image registration using learned cross spectral disparity estimation and a deep guided occlusion reconstruction network [J]. *IEEE Transactions on Image Processing*, 2025, 34: 2338–2350.
- [29] SONG Youyi, TEOH J Y, CHOI K S, et al. Dynamic loss weighting for multiorgan segmentation in medical images [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(8): 10651–10662.
- [30] HE Rundong, HAN Zhongyi, LU Xiankai, et al. SAFER – STUDENT for safe deep semi-supervised learning with unseen-class unlabeled data [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2024, 36(1): 318–334.
- [31] SANGHYUN W, SHOUBHIK D, RONGHANG H, et al. ConvNeXt V2: Co-designing and scaling ConvNets with masked autoencoders [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway, NJ: IEEE, 2023: 16133–16142.
- [32] DING Zhiquan, ZHANG Yuejin, ZHU Chenxin, et al. CAT-Unet: An enhanced U-Net architecture with coordinate attention and skip-neighborhood attention transformer for medical image segmentation [J]. *Information Sciences*, 2024, 670: 120578.
- [33] NING Zhipeng, WANG Hui, LI Shenglin, et al. YOLOv7-RDD: A lightweight efficient pavement distress detection model [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2024, 25(7): 6994–7003.
- [34] LI Hangyu, WANG Nannan, DING Xinpeng, et al. Adaptively learning facial expression representation via C – F labels and distillation [J]. *IEEE Transactions on Image Processing*, 2021, 30: 2016–2028.
- [35] LIN T, PRIYA G, ROSS G, et al. Focalloss for dense object detection [J]. *Computing Research Repository*, 2018, 42(2): 318–327.
- [36] CHRISTINA H. Combining convolutional neural networks for emotion recognition [C]// *Proceedings of 2017 IEEE MIT*

- Undergraduate Research Technology Conference (URTC). Piscataway, NJ: IEEE, 2017; 1-4.
- [37] SAMUEL A, ARSHA N, ANDREA V, et al. Emotion recognition in speech using cross-modal transfer in the wild [C]// Proceedings of the 26<sup>th</sup> ACM International Conference on Multimedia. New York: ACM, 2018; 292-301.
- [38] LI Yong, ZENG Jiabei, SHAN Shiguang, et al. Occlusion aware facial expression recognition using CNN with attention mechanism [J]. IEEE Transactions on Image Processing, 2019, 28(5): 2439-2450.
- [39] ZENG Jiabei, SHAN Shiguang, CHEN Xilin. Facial expression recognition with inconsistently annotated datasets [C]// Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer, 2018; 227-243.
- [40] WANG Kai, PENG Xiaojiang, YANG Jianfei, et al. Suppressing uncertainties for large-scale facial expression recognition [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2020; 6896-6905.
- [41] WANG Kai, PENG Xiaojiang, YANG Jianfei, et al. Region attention networks for pose and occlusion robust facial expression recognition [J]. IEEE Transactions on Image Processing (TIP), 2020, 29: 4057-4069.
- [42] ZHAO Zengqun, LIU Qingshan, ZHOU Feng. Robust lightweight facial expression recognition network with label distribution training [C]// Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI, 2021, 35(4): 3510-3519.
- [43] JIN Rijin, ZHAO Sirui, HAO Zhongkai, et al. AVT: Assisted visual transformer for facial expression recognition [C]// Proceedings of the IEEE International Conference on Image Processing (ICIP). Piscataway, NJ: IEEE, 2020; 2661-2665.
- [44] LIU Hanwei, CAI Hailing, LIN Qingcheng, et al. Adaptive multilayer perceptual attention network for facial expression recognition [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(9): 6253-6266.
- [45] JULES B, ARNAUD D, FERDINAND D, et al. Privileged attribution constrained deep networks for facial expression recognition [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2022; 1055-1061.
- [46] ZHU Huihui, HU Pengyun, TANG Xianpiao, et al. NAGNet: A novel framework for real-time students' sentiment analysis in the Wisdom classroom [J]. Concurrency and Computation: Practice And Experience, 2023, 35(22): e7727.
- [47] CHEN Xuanchi, ZHENG Xiangwei, SUN Kai, et al. Self-supervised vision Transformer-based few-shot learning for facial expression recognition [J]. Information Sciences, 2023, 634: 206-226.
- [48] TANG Xianpiao, HU Pengyun, YU Dewei, et al. A novel relation-aware global attention network for sentiment analysis of students in the wisdom classroom [J]. Concurrency and Computation: Practice and Experience, 2024, 36(18): e8148.
- [49] HU Pengyun, TANG Xianpiao, LIU Yang, et al. LCArNet: A model for analysis of students real-time sentiment by integrating attention mechanism and joint loss function [J]. Complex & Intelligent Systems, 2025, 11: 27.