ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 41,No. 3,May 2026,pp. 841－853
DOI:10. 16337/j. 1004-9037. 2026. 03. 016

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

图卷积和关键点特征融合的D-GFK网络级联人脸表情识别

赵 藤，曹亚茹，闫厚儒，陈 荥，肖 湘，范 蕊，杨 慕，朱 红

（徐州医科大学医学信息与工程学院，徐州 221004）

摘 要：针对当前人脸表情在光照变化、存在遮挡等情况下难以识别，以及悲观情绪识别率较低的问题，本文提出一种基于改进稠密连接网络与人脸关键点特征融合的图卷积级联分类人脸表情识别算法。由于不同深度学习模型在人脸表情识别中各具优势，稠密连接网络在识别乐观和平静表情时准确率较高，而对悲观表情的识别效果较弱，因此本文首先采用小波变换、基于关键部位掩码注意力机制和二叉树分类器对稠密连接网络进行改进，构建了I-Densenet(Improved-DenseNet)模块，用于乐观、平静和悲观3类人脸表情的粗划分，提高粗划分识别率；其次使用基于人脸关键点特征融合的图卷积神经网络对人脸悲观表情细粒度划分，提高悲观表情的识别率。最后，通过将改进的稠密连接网络与基于关键点特征融合的图卷积神经网络进行级联，构建了D-GFK网络(DenseNet-GCN and face key point network)，结合不同模型的优势，综合提高了对人脸表情识别的准确率。实验表明，本文提出的模型在人脸表情识别任务中取得了较好的识别效果。

关键词：人脸表情识别；级联分类；稠密连接网络；特征提取；人脸关键点检测

中图分类号：TP391　　**文献标志码：**A

引用格式：赵藤，曹亚茹，闫厚儒，等.图卷积和关键点特征融合的D-GFK网络级联人脸表情识别[J].数据采集与处理，2026，41(3)：841-853. ZHAO Teng，CAO Yaru，YAN Houru，et al. D-GFK network cascaded facial expression recognition based on graph convolution and key point feature fusion[J]. Journal of Data Acquisition and Processing，2026，41(3)：841-853.

引　言

人的表情通常反映了一个人的内心情绪，人脸表情识别是一种通过分析人类面部表情来推断其情感状态的技术，利用人工智能技术识别和解释人脸上的表情。著名心理学家Ekman[1]提出人类6种基本情感的概念，后来又加入中性表情，构成人脸表情识别的7种基本表情，包括生气、害怕、厌恶、开心、悲伤、惊讶和平静等，并分别使用标签1～7表示。人脸表情识别在社交媒体分析、心理学研究、用户体验设计等领域有广泛应用。然而，由于被试者环境变化和面部外观多样性，人脸表情识别技术仍面临挑战，当被试者环境中有遮挡时或光线复杂时识别准确率会下降[2]，悲观的表情如厌恶和悲伤容易被混淆，这对人脸表情识别提出了重大挑战。

众多研究人员都尝试使用了不同的方法提升人脸识别准确度。例如Wang等[3]提出了将人脸图像分为高权重和低权重，分别通过卷积神经网络(Convolutional neural network，CNN)和图卷积网络

基金项目：国家自然科学基金(62102345)；江苏省卫生健康委员会医学科研项目(Z2020032)；徐州市重点研发计划(KC22117)；江苏摩尔声学技术研究院有限公司横向课题(MESX-202305001)；安徽方舟生物课题(240729001)。
收稿日期：2025-06-15；**修订日期：**2025-07-03

(Graph convolutional network,GCN)得到表情分布,再结合情感标签进行融合预测。这个方法很好地解决了人脸表情识别在复杂光线下和有部分遮挡导致识别困难的问题,但是对于性能有一定的要求。Zhao等[4]提出了一种基于几何和外观知识的感知算法框架,使用卷积神经网络应用于整个面部的观察,同时采用图卷积网络挖掘不同表情背后的面部结构信息,从视觉和结构方面进行面部表情推理,但是由于人脸几何信息容易被遮挡,在模糊或遮挡情况下没有较好的稳定性。Tao等[5]提出了一种具有递进特征融合的分层关注网络,设计了基于多个特征聚合块的多种特征提取模块,融合了不同的梯度特征,提高了网络在不同光照环境下的稳定性,同时增强了图像在人脸关键部分的特征判别,提高了人脸识别的准确性,但是这个网络模型还不够轻量化,并且在复杂环境下具有局限性。Liu等[6]提出了一种基于关联图的面部表情识别方法,将人脸识别任务表述为一个顶点预测问题。作者利用顶点置信度来寻找高阶邻居,利用图卷积对这些邻居进行推理,判断出顶点类别,使用较低的算力即可完成较好的精度。但是存在参数量过大以及过度依赖图结构的问题,需要加入神经网络提高泛化能力。

人脸表情识别虽然在许多领域具有广泛的应用前景,但是也存在一些局限性。人类表情是复杂多变的,同一个表情可能具有不同的解释,不同个体之间的表情表达也具有差异;其次,人脸表情识别对于光线、遮挡和清晰度等因素较为敏感,微小的变化也可能影响人脸表情识别的准确性。单模型的人脸表情识别虽然可以完成任务,但总体准确度低,对于差异大的表情具有较好的识别能力,例如开心、生气等,但是对于差异较小的表情识别能力很弱,极易混淆厌恶、悲伤等表情。同时单一模型对于不同光照条件下和存在遮挡情况下的表情识别也较低[7-8],光照条件的变化可能导致图像的对比度、亮度等特征发生变化,从而影响了表情特征的提取和识别。此外,在面部存在遮挡的情况下,部分面部特征可能被隐藏或模糊,导致算法识别偏差。因此,为了提高在这些场景下的表情识别准确性,通常需要采用多模型数据进行级联分层预测,以提高准确率。

本文提出了一种基于D-GFK网络(DenseNet-GCN and face key point network)的多模型级联分类技术用于人脸表情识别,解决了人脸表情识别在不同光照条件下和存在遮挡环境下难以识别的问题,同时提高了人脸表情在悲观表情下的识别性能。本文首先对稠密连接网络加入小波变换、掩码注意力机制和二叉树分类器进行改进,并使用改进的稠密连接网络对人脸表情数据集进行粗划分,将数据集划分为开心、平静和悲观3类。其次利用改进的稠密连接网络提取的特征与基于边缘引导方法提取的人脸关键点特征进行融合[9],并对融合之后的特征进行降维,使用降维之后融合特征构建一个关联图,然后在关联图中使用图卷积,对输入的人脸图像进行细划分,最后将悲观情绪细划分为生气、害怕、厌恶、惊讶和悲伤5类。由于不同的网络对应人脸表情识别准确率具有差异性,在稠密连接网络中,乐观表情的识别准确率达到了93.76%,平静表情识别准确率达到90.00%,但害怕的表情识别准确率只有55.41%,因此在稠密连接网络识别出乐观表情后即可输出预测结果。为了应对稠密连接网络在悲观表情识别准确率较低的问题,本文提出了级联分类网络,当改进的稠密连接网络预测3种类别进行粗粒度划分后,再结合改进的稠密连接网络与基于边缘引导方法提取的人脸关键点特征,对特征进行融合,利用图卷积神经网络实现悲观表情的细粒度划分,从而提升整体表情识别准确率。

1 相关工作

在原始图像中,存在较多与表情特征提取无关的干扰信息,因此需要对人脸图像进行预处理,例如图像裁剪、灰度化和统一尺寸等。先对图像进行人脸位置检测,框选出人脸位置并进行裁剪,再将彩色图像统一尺寸,最后转换为灰度图像,得到模型所需要的输入图像。同时由于训练图像数据存在分类不平衡问题,需要通过数据扩增增加样本数量,从而提高模型对不同类别数据的泛化能力和整体性能。

卷积神经网络是一种具有深度结构的前馈神经网络,专门用于处理具有网格结构的数据。核心思想是通过卷积层、池化层和全连接层等组件来自动学习输入数据的特征表示,从而实现对复杂模式的

识别和分类。

稠密连接网络(DenseNet)是一种深度卷积神经网络,由Huang等[10]于2017年提出。稠密连接网络通过密集连接的方式将每一层的输出与之前所有层的输出连接起来,从而实现了特征的复用和信息的流动,有助于缓解梯度消失问题,提高了网络的训练效率和参数利用率。在稠密连接网络中,每个层的输入由之前所有层的输出拼接而成,这种密集连接的结构使得网络更加深层,从而能够更好地捕捉和利用输入数据的特征信息,具有更强的表征能力。相较于传统的卷积神经网络,稠密连接网络通过密集连接的方式将每一层的输出与之前所有层的输出连接起来,有效地增强了特征的传播和重用,从而在面部表情识别中表现出更好的性能。相比于传统的卷积神经网络,稠密连接网络具有更少的参数量、更高的参数利用率以及更好的特征重用能力,具有较强的泛化能力,但是传统稠密连接网络存在特征提取效率低、分类器不够准确的问题。

GCN[11]是一种针对图数据结构进行深度学习的重要方法,已有的工作已经证明了图卷积网络对复杂图形模式建模的强大能力,在各种任务上,图卷积网络的使用带来了相当大的性能提升。与传统的卷积神经网络不同,图卷积网络能够有效地处理非结构化的人脸表情的特征数据。图卷积网络通过在图数据结构上定义卷积操作,实现了对图数据结构的特征提取和表示学习,从而可以应用于节点分类、链接预测和图数据结构推断等任务。其核心思想是通过邻居节点的信息传递来更新每个节点的特征表示,使得节点特征能够充分地利用图的局部信息。

人脸关键点识别旨在自动检测和定位人脸图像中的典型特征位置[12],如眼睛、鼻子、嘴巴和眉毛等,是面部几何结构的重要表征形式[13]。关键点信息不仅与表情变化密切相关,有助于提升表情识别的准确性,还具有较强的光照鲁棒性,从而增强人脸识别系统在复杂环境下的稳定性和适应性。

小波变换是一种利用特定小波基函数进行信号分析的数学框架,能够将图像在不同尺度上进行分解。它通过将信号或图像分解为不同频率和尺度的子带,实现了多分辨率分析。在图像处理中,小波变换将图像分解为4个子带:低频近似信息(Low-low,LL)、垂直边缘信息(Low-high,LH)、水平边缘信息(High-low,HL)和对角细节信息(High-high,HH),分别反映了不同方向与尺度下的图像特征。这种分解方式不仅有助于提取图像的多尺度特征,还能有效分离图像的背景和纹理信息。通过逆小波变换,这些子带被重构回原始图像,实现特征提取和图像恢复。将小波变换集成卷积神经网络中,可以增强模型对高频细节的敏感性,提高特征提取的丰富性和准确性,进而提升模型在图像分类任务中的性能。

2　本文方法

2.1　基于改进稠密连接网络I-DenseNet的人脸表情粗粒度识别

本文对稠密连接网络进行了改进,构建了基于小波变换、掩码注意力和二叉树分类器的改进稠密连接网络I-Densenet。

2.1.1　小波变换融入

在人脸表情识别中,应用稠密连接网络可以有效地提高识别性能。由于人脸表情具有复杂多变的特点,稠密连接网络能够更全面地捕捉面部表情中的微妙变化。通过稠密连接的方式,模型能够直接访问之前所有层的特征信息,包括微小的面部表情特征,从而提高了模型对表情的敏感度。

稠密连接网络由交替连接的稠密块和过渡层组成。在稠密块中,每一层都直接连接到所有后续层,以增强特征传递。因此,后续的每一层都会接收到来自前面所有层的特征图。使用X_r表示第r层输出层,有

$$X_r = H_r([X_0, X_1, \cdots, X_{r-1}]) \tag{1}$$

式中:$[X_0, X_1, \cdots, X_{r-1}]$为在层$0, 1, \cdots, r-1$中产生的特征映射的连接;$H_r$为第$r$层的复合变换函数,用于将前序所有层的拼接特征映射为当前层的输出特征图。

然而传统稠密连接网络存在对高频信息的处理不足、无法充分捕捉到图像中的多尺度特征和复杂的结构信息局限性。本文将小波变换集成到了稠密连接网络的每个密集层的卷积操作之间，用于增强特征提取能力。先将输入的特征图分解为LL、LH、HL和HH 4个子带，有

$$\begin{cases} \mathrm{LL} = \mathrm{conv2}d(x, \mathrm{dec\_lo} \otimes \mathrm{dec\_lo}) \\ \mathrm{LH} = \mathrm{conv2}d(x, \mathrm{dec\_lo} \otimes \mathrm{dec\_hi}) \\ \mathrm{HL} = \mathrm{conv2}d(x, \mathrm{dec\_hi} \otimes \mathrm{dec\_lo}) \\ \mathrm{HH} = \mathrm{conv2}d(x, \mathrm{dec\_hi} \otimes \mathrm{dec\_hi}) \end{cases} \tag{2}$$

式中：dec_lo和dec_hi分别为低通滤波器和高通滤波器的系数，“⊗”为张量积。然后对每个子带进行下采样，将特征图的分辨率降低一半，再将每个子带重构为原始特征图，有

$$\begin{cases} \mathrm{LL}_{\mathrm{up}} = \mathrm{upsample}(\mathrm{LL}_{\mathrm{down}}, 2) \\ \mathrm{LH}_{\mathrm{up}} = \mathrm{upsample}(\mathrm{LH}_{\mathrm{down}}, 2) \\ \mathrm{HL}_{\mathrm{up}} = \mathrm{upsample}(\mathrm{HL}_{\mathrm{down}}, 2) \\ \mathrm{HH}_{\mathrm{up}} = \mathrm{upsample}(\mathrm{HH}_{\mathrm{down}}, 2) \\ x_{\mathrm{rec}} = \mathrm{conv2}d(\mathrm{concat}(\mathrm{LL}_{\mathrm{up}}, \mathrm{LH}_{\mathrm{up}}, \mathrm{HL}_{\mathrm{up}}, \mathrm{HH}_{\mathrm{up}}), \mathrm{rec\_lo} \otimes \mathrm{rec\_lo}) \end{cases} \tag{3}$$

式中：经过小波分解和下采样之后得到$\mathrm{LL}_{\mathrm{down}}$、$\mathrm{LH}_{\mathrm{down}}$、$\mathrm{HL}_{\mathrm{down}}$和$\mathrm{HH}_{\mathrm{down}}$ 4个子带，再上采样得到$\mathrm{LL}_{\mathrm{up}}$、$\mathrm{LH}_{\mathrm{up}}$、$\mathrm{HL}_{\mathrm{up}}$和$\mathrm{HH}_{\mathrm{up}}$ 4个子带；upsample表示上采样操作，concat表示将4个子带特征图拼接在一起，最后经过上采样重构得到特征图x_{rec}。再将小波变换模块嵌入在第1个卷积层和第2个卷积层之间，使得稠密连接网络在处理具有复杂纹理和边缘信息，如图1所示。

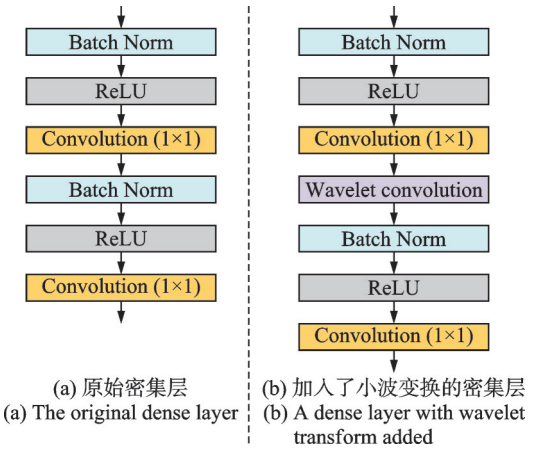


图1 原始密集层和加入了小波变换的密集层

Fig.1 Original dense layer and dense layer with wavelet transform added

2.1.2 基于关键部位掩码的注意力机制融入

在人脸表情识别中，基于关键部位掩码的注意力机制是一种有效的技术手段。该注意力通过使用人脸关键点检测模块，识别出人脸的关键部位，如眼睛、嘴巴等，再通过生成关键部位的掩码，模型能够将更多的注意力集中在这些区域，从而更准确地捕捉到表情变化的特征。例如，当识别快乐或悲伤等情绪时，嘴巴的形状变化是重要的特征。如图2所示，基于关键部位掩码的注意力机制能使模型在训练时专注于对表情识别任务更有帮助的关键区域，提高表情识别的准确率。

图2 基于关键部位掩码的注意力机制

Fig.2 Attention mechanism based on critical site masks

2.1.3 二叉树分类器改进

在传统的神经网络中，输入数据通常经过单一的全连接层或卷积层进行处理，然而，这种单一处理路径可能限制了模型对复杂特征的捕捉能力。为了提高模型的表现，本文使用了二叉树的分类器[14]结构。

二叉树分类器方法利用树结构的递归特性对输入数据进行分割和处理，从而捕获更深层次的特征关系，使得每个节点都能对数据进行特定的变换，从而实现分类任务。二叉树分类器主要有3个部分，树节点结构包含两个全连接层，分别对应左子节点和右子节点的输出；二叉树结构通过递归方式构建完整的树结构。树的根节点为1个树节点，后续的每层节点通过递归方式继续生成二叉树的深度决定

了树的层数;在前向传播结构中,左子节点的输出继续递归地传递给左子树,右子节点的输出递归地传递给右子树,整个过程通过深度优先遍历方式实现,最终将所有叶节点的输出相加,得到最终的分类结果,如图3所示,其中R_1为根节点,$R_i(i\geqslant 2)$为树的叶子节点,A_j为不同的决策,P_k为不同的分类结果。

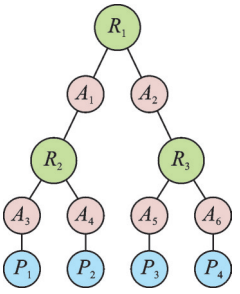


图3　二叉树分类器

Fig.3　Binary tree classifier

2.2　基于图卷积与关键点特征融合的悲观表情识别

本文将I-Densenet(Improved-DenseNet)提取的人脸表情特征,融合人脸关键点特征,使用图卷积构建了基于人脸关键点的图卷积模块KFP-GCN(Face key point-GCN)。先使用I-Densenet对人脸表情进行粗粒度划分,将其划分为开心、平静和悲观3类,再使用KFP-GCN对悲观情绪进行细粒度划分,将悲观情绪细划分为生气、害怕、厌恶、惊讶、悲伤5类,从而构建级联的D-GFK网络。

2.2.1　基于图卷积的特征提取

本文通过使用已经预训练的卷积神经网络模型,提取训练集中的人脸表情特征,对特征进行降维后组成一个关联图。本文将每个图像视为一个顶点,利用顶点之间的余弦相似度构造关联图[15],并将其构造为关联图模型。输入人脸图像数据时,使用预训练的模型进行特征提取,再将其融入关联图中,获得的与其关联的节点信息,通过关联节点的个数、标签、距离以及高阶邻域的个数、标签和距离综合预测人脸表情。

通过将不同人脸图像的特征提取出来,并以图的形式进行组合,捕获训练集图像之间的关系。利用图卷积神经网络进行标签预测,可以充分利用图结构中的拓扑信息,进一步提高了模型的性能和稳定性。

使用一个已经训练好的模型提取图像的特征,从全连接层输出的特征向量,利用t-SNE算法对提取特征进行降维,将高维特征降维到二维空间,KL散度公式为

$$C=\sum_i \mathrm{KL}(P_i\|Q_i)=\sum_i\sum_j p_{ij}\log\frac{p_{ij}}{q_{ij}} \tag{4}$$

式中:C表示KL散度,是优化的目标函数;P_i和Q_i分别表示高维空间和低维空间中的概率分布;p_{ij}和q_{ij}分别表示高维空间和低维空间中样本x_i和x_j之间的条件概率分布;$\sum_i$和$\sum_j$表示对所有样本对之间的概率分布差异进行全局累加;$\frac{p_{ij}}{q_{ij}}$表示两个空间中样本之间的相对距离。

当输入一张人脸图像时,提取其特征并将其放入已经训练好的关联图中,获得与其关联的节点信息,通过关联节点的个数、标签、距离以及高阶邻域的个数、标签和距离综合预测人脸表情。

2.2.2　基于人脸关键点的特征提取

基于人脸关键点的表情识别方法是一种利用人脸上的关键点信息来推断面部表情的技术。在这种方法中,首先通过人脸检测算法定位人脸,然后使用关键点检测算法识别出人脸上具有语义含义的关键点,例如眼睛、眉毛、鼻子和嘴巴等区域(图2)。利用关键点的位置信息进行人脸几何特征提取,实现对人脸表情的准确分类和识别。

首先使用Canny边缘检测算法定位人脸的边缘,进而辅助提取人脸的特征点,有

$$G=\sqrt{G_x^2+G_y^2} \tag{5}$$

Canny边缘检测算法用于计算图像中每个像素点的梯度幅值,式(5)中G_x和G_y分别表示图像在水平和垂直方向的梯度,G表示图像中每个像素点的梯度幅值。通过计算梯度幅值,可以找到图像中的边缘信息,进而得到人脸的几何信息。

2.2.3 基于改进的稠密连接网络与人脸关键点特征融合的图卷积悲观表情识别模型

在人脸表情识别中，卷积神经网络通过逐层卷积提取全局语义特征，而人脸关键点则提供局部几何信息，仅关注有限的信息并不能准确地识别表情，为了结合两者优势，因此本文构建了基于改进的稠密连接网络与人脸关键点特征融合的图卷积悲观表情识别模型。将从粗粒度划分中的改进稠密连接网络 V_1 与人脸关键点识别输出的特征向量 V_2 进行拼接，以获得更丰富和更具有表征性的特征表示，有

$$V_{\text{fusion}}=[V_1, V_2] \tag{6}$$

式中 V_{fusion} 为融合的特征。利用t-SNE算法对融合特征进行降维，将高维特征降维到二维空间。

最后使用图卷积模型对输入图像进行预测，每一层的图卷积操作为

$$H^{(l+1)}=\sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \tag{7}$$

式中：$H^{(l)}$ 为第 l 层的特征，$\tilde{A}$ 为邻接矩阵和单位矩阵的和，$\tilde{D}$ 为 $\tilde{A}$ 的度矩阵，$W^{(l)}$ 为第 l 层的权重矩阵，σ 为激活函数。最后一层输出节点的分类概率，有

$$Z=\text{soft}\max\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}H^{(L-1)}W^{(L-1)}\right) \tag{8}$$

式中 L 为网络中的层数。最后通过式(9)预测类别，有

$$\hat{y}_i=\arg\max(Z_i) \tag{9}$$

式中 Z_i 为第 i 个节点或样本的概率向量。通过选择最大概率对应的类别索引，从而确定每个节点或样本的预测类别。

2.3 基于D-GFK网络级联分类表情识别

由于改进的稠密连接网络，基于人脸关键点的图卷积网络识别的结果各有优劣势，因此本文提出了一种基于D-GFK网络级联分类方法，如图4所示。该方法首先使用改进的稠密连接网络对人脸表情数据集进行粗划分，将人脸表情划分为开心、平静和悲观3类；其次利用改进的稠密连接网络提取的特征与人脸关键点特征进行融合，并对融合之后的特征进行降维，使用降维之后融合特征构建一个关联图；然后在关联图中使用图卷积，对输入的人脸图像进行细划分；最后将悲观情绪细划分为生气、害怕、厌恶、惊讶和悲伤5类，从而提高多模型信息的综合利用和识别性能。

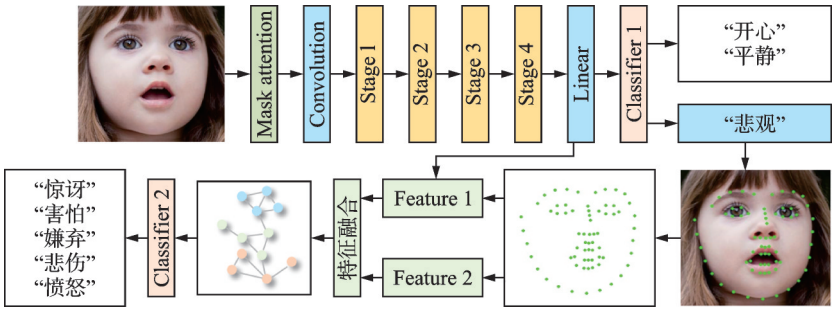


图4 基于D-GFK网络的级联分类人脸表情识别结构图

Fig.4 Structure diagram of cascaded classification facial expression recognition based on D-GFK network

该方法通过结合多模型的识别优势，对前一个网络的预测结果进行矫正，当稠密连接网络输出乐观或平静表情预测时直接输出预测结果；当输出悲观表情时则进入后续模型进行更加准确的预测。通过这种级联分类机制，能够更准确地融合多模型优势，提高识别结果的准确性。

因此对多个模型引入级联分类机制，通过对一级预测结果的判断，决定是否进入二级预测。首先使用稠密连接网络对输入图像进行第1次预测，有

$$R=\begin{cases}y & y=4\text{ or }y=7\\ \text{Continue} & \text{其他}\end{cases} \tag{10}$$

式中:y表示预测结果,y值为4时预测为开心,y值为7时预测为平静;R表示识别结果。根据第1次预测结果,当预测值不是开心(标签4)或平静(标签7)的表情时输入二级网络,最后使用图卷积输出最终结果。

本文的模型依然能够准确地聚焦在人脸图像的这些关键区域的主要原因有两个:(1)本文的D-GFK网络可以同时关注局部和全局特征;(2)本文的D-GFK网络可以有效地级联多个模型,调整级联模型输出结果。即使在面部遮挡和光线变化等复杂情况下,口腔区域以及眼睛眉毛区域对于面部表情识别具有极大的帮助。

3　实验结果及分析

3.1　数据集

为了分析D-GFK网络在表情识别任务中的性能,本文在公开数据集RAF-DB[16]与FER2013上进行实验。

真实世界的情感面孔数据库RAF-DB是一个大型面部表情数据库,包括29 672个真实世界图像及每个图像的标签。其中单标签数量15 339个,包含训练集12 271个和测试集3 068个。在单标签数据集中共有7种基本表情,分别使用表情标签1～7进行表示,具体如表1所示。

表1　RAF-DB数据集数据配置

Table 1　Configuration of the RAF-DB dataset

情绪类别	标签	训练集	测试集	总数
Surprise	1	1 290	329	1 619
Fear	2	281	74	355
Disgust	3	717	160	867
Happiness	4	4 772	1 185	5 957
Sadness	5	1 982	478	2 460
Anger	6	705	162	867
Neutral	7	2 524	680	3 204
总数	—	12 271	3 068	15 339

FER2013数据集是面部表情识别领域广泛使用的基准数据集,包含35 887张灰度人脸图像,涵盖7种基本情绪类别。数据分为28 709张训练图、3 589张验证图和3 589张测试集,如表2所示。

表2　FER2013原始数据集数据配置

Table 2　Configuration of the FER2013 dataset

情绪类别	训练集	验证集	测试集	总数
Anger	4 995	467	467	5 929
Disgust	436	56	56	548
Fear	4 097	496	496	5 089
Happy	8 989	895	895	10 779
Sad	6 077	653	653	7 383
Surprise	4 170	415	415	5 000
Neutral	6 198	607	607	7 412
总数	28 709	3 589	3 589	35 887

实验对所有人脸表情数据采取裁剪和归一化处理,最终实验图像大小调整为224×224像素。本文试验均在Windows环境中完成,环境配置为Intel E5-2620v4 CPU,内存64 GB,Geforce RTX 2080Ti 11G GPU。

3.2 数据扩增

为了提高图像数据的多样性和提升模型的泛化能力,本文对图像数据进行了多种预处理操作,包含旋转、翻转、平移和缩放等。同时,引入了随机遮挡操作,如图5所示。遮挡部分占图像的25%,其目的是模拟现实场景中可能出现的部分遮挡情况,增强模型对不完整图像的识别能力。通过数据增强生成多种变体的图像数据,进一步提升模型的鲁棒性。

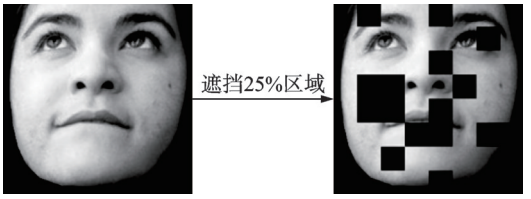


图5 图像加入25%的遮挡

Fig.5 Image with 25% occlusion

3.3 改进的稠密连接网络

稠密连接网络通过其稠密连接方式,能够有效缓解梯度消失问题,增强特征传播和重用,从而在人脸表情识别任务中表现出色。相较于其他卷积神经网络,能够提高人脸表情识别准确率,如表3所示。

表3 RAF-DB数据集在多种基础模型中的准确率

Table 3 Accuracy of RAF-DB dataset in basic models

基础模型	准确率/%
CNN	70.53
ResNet18[17]	65.84
ResNet50[18]	70.18
Vision Transformer[19]	69.04
VGG16	76.50
Vision mamba[20]	50.34
Vmamba[21]	63.02
DenseNet121	**76.76**

然而传统稠密连接网络存在对高频信息的处理不足、无法充分捕捉到图像中的多尺度特征和复杂的结构信息局限性。本文将小波变换集成到了稠密连接网络的每个密集层的卷积操作之间,用于增强特征提取能力。并且,本文使用二叉树分类器,利用树结构的递归特性对输入数据进行分割和处理,从而捕获更深层次的特征关系,使得每个节点都能对数据进行特定的变换,从而提高分类准确率。其次,通过人脸关键点检测,提取人脸的关键部位作为注意力区域,并生成相应的掩码,强化了模型对关键特征区域的聚焦,使得模型在训练过程中能够更高效地学习到关键信息,进而提升整体的分类准确率。为了研究改进的稠密连接网络各组成部分对模型性能的影响,本文分别对小波变换模块(Wavelets transform)、掩码注意力机制模块(Mask attention)和二叉树分类器(Binary tree classifier)模块依次进行消融实验,消融实验结果如表4所示。

表4 改进的稠密连接网络消融实验

Table 4 Ablation experiment for improved DenseNet network

Wavelet transform	Mask attention	Binary tree classifier	准确率/%	
			RAF-DB	FER2013
—	—	—	83.18	63.51
√	—	—	84.03	64.85
—	√	—	81.13	66.85
—	—	√	82.01	65.35
√	√	—	84.19	65.70
√	—	√	83.05	63.57
—	√	√	82.70	62.39
√	√	√	**84.60**	**66.91**

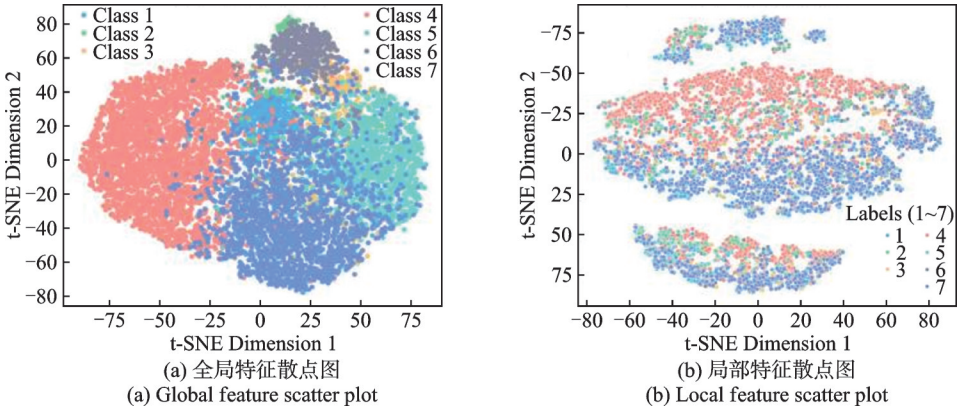
3.4　特征融合分析

本文采用了基于深度学习的卷积神经网络模型,使用预训练的模型,有效地识别并定位68个面部关键点,如图6所示,然后使用人脸关键点的坐标数据,得到人脸几何信息。

图6　人脸关键点识别

Fig.6　Recognition of face key points

传统卷积神经网络通过多层感受野聚合以提取全局特征,而人脸关键点识别则更侧重局部结构信息,但是仅关注有限的信息会降低人脸标签识别的准确度,如图7所示。



(a) 全局特征散点图
(a) Global feature scatter plot

(b) 局部特征散点图
(b) Local feature scatter plot

图7　全局特征散点图与局部特征散点图

Fig.7　Global feature scatter plot and local feature scatter plot

由图7可以看出,仅使用全局特征进行识别会导致不同标签边界模糊不清;仅使用局部特征散点图会使得不同标签之间聚集形成3个大簇,而标签之间的区别较小无法区分,难以进行预测。将全局特征和局部特征进行融合后,绘制出散点图如图8所示。特征融合后有效区分了不同标签之间的边界,不同标签之间的聚集程度也更加紧密。

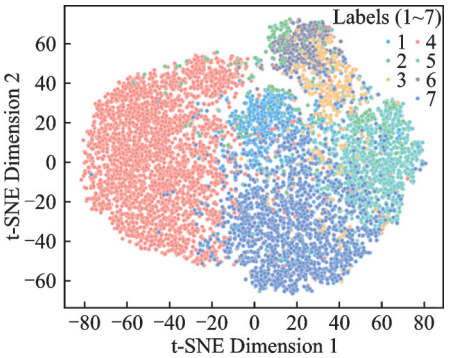


图8　特征融合t-SNE降维散点图

Fig.8　Dimensionality reduction scatter plot for feature fusion t-SNE

3.5　级联分类消融实验

传统的多模型融合是将多个独立训练模型的预测结果进行简单的组合,以获得更加稳健和可靠的预测,这种方法通常采用投票或平均的方式进行融合。传统多模型融合方法简单直观,易于实现,但可能会忽略不同模型之间的差异性,且无法自适应地调整每个模型的权重。多模型级联分类则更加灵活。此方法可以根据模型的性能和准确度调整输出,从而更好地利用每个模型的优势。本文首先使用改进的稠密连接网络进行粗粒度划分,再基于人脸关键点对人脸几何信息提取,使用图卷积网络进行细粒度划分,从而对识别结果进行矫正,提高准确率。为了研究D-GFK级联分类模型中各个组成部分对模型性能的影响,本文分别对改进的稠密连接模块I-Densenet、基于人脸关键点的图卷积模块FKP-GCN和级联模块依次进行消融实验,表5给出了级联分类模型在不同数据集上的消融实验结果。

表5 级联分类模型在不同数据集上的消融实验结果

Table 5 Ablation experiment results of cascading classification models on different datasets

模块		级联	准确率/%	
I-Densenet	FKP-GCN		RAF-DB	FER2013
√	—	—	84.60	66.91
—	√	—	80.67	63.75
√	√	—	82.40	65.35
√	√	√	**89.77**	**68.83**

3.6 实验结果

表6展示了多种人脸表情识别方法在RAF-DB数据集上的比较结果。本文的D-GFK网络方法在人脸表情识别上达到了89.77%的准确率，相比于DenseNet、Resnet18等传统模型有较大提高，相比于最优的FERGCN模型也有1.54%的提高。同时，在公开数据集FER2013上本文方法也有68.83%的准确率。

与传统方法相比，本文提出的D-GFK网络基于多模型的级联分类方法展示了更好的性能，并且由于同时使用了人脸表情的全局感知与几何信息，使得本文所提出的方法在光照变化和存在遮挡的环境中具有较好的稳定性。由于本文使用了级联分类的方法，开心和平静可在粗粒度阶段被有效区分，有效地提高了模型性能。

表6 不同模型在RAF-DB数据集上的性能对比

Table 6 Performance comparison of different models on RAF-DB datasets

方法	准确率/%
ACNN[22]	85.07
RAN[23]	86.90
SCN[24]	88.14
DACL[25]	87.78
KTN[26]	88.07
EAC[27]	88.02
FT-CSAT[28]	88.61
DRGFER[29]	84.41
MixCut[30]	87.88
FERGCN[31]	88.23
LVLM[32]	87.84
Ours	**89.77**

4 结束语

本文提出了一种基于图卷积与关键点特征融合的级联分类的人脸表情识别算法，并在光照变换和存在遮挡情况下具有较好的可靠性。该方法能够更全面地捕捉和分析人的表情变化，从而实现更精确的表情识别。综合考虑融合多模型特征信息的方法可以克服单一模型对不同光照条件和遮挡情况下表情识别的局限性，提高识别系统的可靠性。针对面对悲观情绪识别困难的问题，级联分类算法更偏向于基于人脸关键点的图卷积模型识别的结果，使得悲观情绪也能拥有较好的识别能力。同时图卷积相比于卷积神经网络具有较小的计算复杂度。

尽管本文的方法在一定程度上取得了显著的效果，但是依然存在一些局限性。首先，在处理模糊图像时，人脸识别的准确率仍然较低。其次，本方法在更复杂的场景下识别能力仍然具有挑战性，例如多光源场景、运动状态下等。未来将致力于优化模型的性能，使模型在复杂环境下保持鲁棒性。

参考文献：

[1] EKMAN P. Constants across cultures in the face and emotion[J]. Journal of Personality and Social Psychology, 1971, 17: 124-139.

[2] KARNATI M, SEAL A, YAZIDI A, et al. FLEPNet: Feature level ensemble parallel network for facial expression recognition[J]. IEEE Transactions on Affective Computing, 2022, 13(4): 2058-2070.

[3] WANG S, ZHAO A, LAI C, et al. GCANet: Geometry cues-aware facial expression recognition based on graph convolutional

networks[J]. Journal of King Saud University-Computer and Information Sciences, 2023, 35(7): 101605.

[4] ZHAO R, LIU T, HUANG Z, et al. Geometry-aware facial expression recognition via attentive graph convolutional networks [J]. IEEE Transactions on Affective Computing, 2023, 14(2): 1159-1174.

[5] TAO H, DUAN Q. Hierarchical attention network with progressive feature fusion for facial expression recognition[J]. Neural Networks, 2024, 170: 337-348.

[6] LIU T, LI J, WU J, et al. Facial expression recognition on the high aggregation subgraphs[J]. IEEE Transactions on Image Processing, 2023, 32: 3732-3745.

[7] LI Y, ZENG J, SHAN S, et al. Occlusion aware facial expression recognition using CNN with attention mechanism[J]. IEEE Transactions on Image Processing, 2019, 28(5): 2439-2450.

[8] WANG K, PENG X, YANG J, et al. Region attention networks for pose and occlusion robust facial expression recognition[J]. IEEE Transactions on Image Processing, 2020, 29: 4057-4069.

[9] HAMILTON W L, YING R, LESKOVEC J. Inductive representation learning on large graphs[C]//Proceedings of Advances in Neural Information Processing Systems (NeurIPS). Long Beach, USA: Curran Associates Inc., 2017: 1025-1035.

[10] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017: 4700-4708.

[11] KIPF T, WELLING M. Semi-supervised classification with graph convolutional networks[C]//Proceedings of the International Conference on Learning Representations. Toulon: ICLR, 2017.

[12] ZHANG Z, LUO P, LOY C C, et al. Facial landmark detection by deep multi-task learning[C]//Proceedings of European Conference on Computer Vision. Cham: Springer International Publishing, 2014: 94-108.

[13] YAN S, XIONG Y, LIN D. Spatial temporal graph convolutional networks for skeleton-based action recognition[C]// Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto:AAAI Press, 2018, 32(1): 7444-7452.

[14] JI R, WEN L, ZHANG L, et al. Attention convolutional binary neural tree for fine-grained visual categorization[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE, 2020: 10468-10477.

[15] SCHLICHTKRULL M, KIPF T N, BLOEM P, et al. Modeling relational data with graph convolutional networks[C]// Proceedings of European Semantic Web Conference. Cham: Springer International Publishing, 2018: 593-607.

[16] LI S, DENG W. Reliable crowdsourcing and deep locality preserving learning for unconstrained facial expression recognition[J]. IEEE Transactions on Image Processing, 2018, 28(1): 356-370.

[17] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(10): 2124-2138.

[18] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2016: 770-778.

[19] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: Transformers for image recognition at scale[C]//Proceedings of the International Conference on Learning Representations. Vienna: ICLR, 2021.

[20] ZHU L, LIAO B, ZHANG Q, et al. Vision mamba: Efficient visual representation learning with bidirectional state space model [EB/OL]. (2024-01-17) [2025-04-09]. https://arxiv.org/abs/2401.09417.

[21] LIU Y, TIAN Y, ZHAO Y, et al. VMamba: Visual state space model[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE, 2024: 21865-21875.

[22] LI Y, ZENG J, SHAN S, et al. Occlusion aware facial expression recognition using CNN with attention mechanism[J]. IEEE Transactions on Image Processing, 2018, 28(5): 2439-2450.

[23] WANG K, PENG X, YANG J, et al. Region attention networks for pose and occlusion robust facial expression recognition[J]. IEEE Transactions on Image Processing, 2020, 29: 4057-4069.

[24] WANG K, PENG X, YANG J, et al. Suppressing uncertainties for large-scale facial expression recognition[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]:IEEE, 2020: 6897-6906.

[25] 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. [S.l.]: IEEE, 2021: 2402-2411.

[26] LI H, WANG N, DING X, et al. Adaptively learning facial expression representation via CF labels and distillation[J]. IEEE Transactions on Image Processing, 2021, 30: 2016-2028.

[27] ZHANG Y, WANG C, LING X, et al. Learn from all: Erasing attention consistency for noisy label facial expression recognition[C]//Proceedings of European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 418-434.

[28] YAO H, YANG X, CHEN D, et al. Facial expression recognition based on fine-tuned channel-spatial attention transformer[J]. Sensors, 2023, 23(15): 6799.

[29] WANG S, LI X. Dynamic resolution guidance for facial expression recognition[C]//Proceedings of International Conference on Pattern Recognition. Cham: Springer Nature Switzerland, 2024: 412-426.

[30] YU J, LIU Y, Fan R, et al. MixCut: A data augmentation method for facial expression recognition[EB/OL]. (2024-05-17)[2025-04-09].https://arxiv.org/abs/2405.10489.

[31] LIAO L, ZHU Y, ZHENG B, et al. FERGCN: Facial expression recognition based on graph covolution network[J]. Machine Vision and Applications, 2022, 33(3): 40.

[32] YU J, LU X. Compound expression recognition via large vision-language models[EB/OL]. (2025-03-14)[2025-04-09]. https://arxiv.org/abs/2503.11241.

作者简介:

赵藤(2000-),男,硕士研究生,研究方向:智能信息处理,E-mail:303109110917@stu.xzhmu.edu.cn。

曹亚茹(2001-),女,硕士研究生,研究方向:深度学习、智能医学图像处理,E-mail:303109110913@stu.xzhmu.edu.cn。

闫厚儒(2001-)男,硕士研究生,研究方向:深度学习、智能人脸情感识别,E-mail:304109120866@stu.xzhmu.edu.cn。

陈荥(2001-),男,硕士研究生,研究方向:深度学习、智能医学图像处理,E-mail:304109110870@stu.xzhmu.edu.cn。

肖湘(2002-),硕士研究生,研究方向:大模型心理健康智能诊断,E-mail:304109110877@stu.xzhmu.edu.cn。

范蕊(2001-),硕士研究生,研究方向:深度学习、智能医学信号处理,E-mail:304109120868@stu.xzhmu.edu.cn。

杨慕(2001-),硕士研究生,研究方向:医学人工智能与大数据,E-mail:304109110873@stu.xzhmu.edu.cn。

朱红(1970-),**通信作者**,女,博士,教授,硕士生导师,研究方向:机器学习、模式识别、人工智能,E-mail:zhuhong@xzhmu.edu.cn。

(编辑:刘彦东)

D-GFK Network Cascaded Facial Expression Recognition Based on Graph Convolution and Key Point Feature Fusion

ZHAO Teng, CAO Yaru, YAN Houru, CHEN Ying, XIAO Xiang, FAN Rui, YANG Mu, ZHU Hong*

(School of Medical Information and Engineering, Xuzhou Medical University, Xuzhou 221004, China)

Abstract: In view of the current problems that facial expressions are difficult to recognize under conditions of lighting changes and occlusion, as well as the low recognition rate of pessimistic emotions, this paper proposes a facial expression recognition algorithm based on graph convolutional cascade classification based on improved dense connection network and fusion of facial key point features. Since different deep learning models have their own advantages in facial expression recognition, dense connection network has a high accuracy rate in recognizing optimistic and calm expressions, but has a weak recognition effect on pessimistic expressions. Therefore, this paper first uses wavelet transform, key part mask attention mechanism and binary tree classifier to improve the dense connection network, and constructs the I-Densenet (Improved-DenseNet) module for the rough division of optimistic, calm and pessimistic facial expressions to improve the recognition rate of rough division; Secondly, the graph convolutional neural network based on the fusion of facial key point features is used to fine-grainedly divide the pessimistic expression of the face to improve the recognition rate of pessimistic expression. Finally, this paper constructs the D-GFK network (DenseNet-GCN and face key point network) by cascading the improved dense connection network with the graph convolutional neural network based on key point feature fusion, combining the advantages of different models to comprehensively improve the accuracy of facial expression recognition. Experiments show that the model proposed in this paper has achieved good recognition results in facial expression recognition tasks.

Highlights:

1. Propose a cascaded facial expression recognition network named D-GFK (DenseNet-GCN and face key point network), which employs an improved DenseNet for rough division and a graph convolutional neural network based on facial key point feature fusion for fine-grained division, thereby improving facial expression recognition performance.
2. Construct a rough division module integrating wavelet transform, key-region mask attention mechanism, and binary tree classifier, enhancing the robustness of facial expression recognition under complex illumination and occlusion conditions.
3. Design a fine-grained graph convolutional classification module based on facial key point feature fusion, which combines global semantic features and local geometric features to improve the recognition ability of pessimistic facial expressions.

Key words: face expression recognition; cascade classification; dense connection network; feature extraction; face key point detection

Foundation items: National Natural Science Foundation of China (No.62102345); Medical Research Project of Jiangsu Provincial Health Commission (No.Z2020032); Key Research and Development Program of Xuzhou City (No.KC22117); Jiangsu Moore Acoustics Technology Research Institute Co., Ltd. Horizontal Project (No.MESX-202305001); Anhui Ark Biotechnology Project (No. 240729001).

Received: 2025-06-15; **Revised**: 2025-07-03

***Corresponding author**, **E-mail**: zhuhong@xzhmu.edu.cn.