胡蝶, 唐敏, 朱永梁,等. 基于注意力机制的人脸情绪识别[J]. 智能计算机与应用,2024,14(8):65-69. DOI:10.20169/j.issn.2095-2163.240811

基于注意力机制的人脸情绪识别

胡 蝶[1], 唐 敏[1], 朱永梁[1], 朱耀东[2]
(1 浙江理工大学 信息科学与工程学院, 杭州 310018; 2 嘉兴学院 信息科学与工程学院, 浙江 嘉兴 314001)

摘 要: 针对传统情绪识别模型在面对复杂情境、小目标特征提取以及通道信息整合方面的不足,本文提出了一种基于注意力机制的人脸情绪识别方法。使用 ResNet50 作为主干网络,融合通道注意力机制 CA 和 SimAM 无参数注意力机制,使网络关注图像的重点区域,忽略背景区域,同时用间隔采样模块 SC 替换 ResNet50 中的第一个下采样模块,来增强网络对小目标的检测能力。实验结果表明,本文提出的方法能够获取有效的特征,忽略背景区域,有利于面部表情的识别和判断,并且在自采集数据集上最高达到了 76.45%的准确率,相对于传统的人脸情绪识别方法,准确率提高了 3.1%。

关键词: 情绪识别; 注意力机制; 间隔采样; 特征提取
中图分类号: TP391 **文献标志码**: A **文章编号**: 2095-2163(2024)08-0065-05

Facial emotion recognition based on attention mechanism

HU Die[1], TANG Ming[1], ZHU Yongliang[1], ZHU Yaodong[2]
(1 College of Information Science and Engineering, Zhejiang Sci-Tech University, Hangzhou 310018, China;
2 College of Information Science and Engineering, Jiaxing University, Jiaxing 314001, Zhejiang, China)

Abstract: Aiming at the shortcomings of traditional emotion recognition models in facing complex situations, small target feature extraction and channel information integration, this paper proposes a face emotion recognition method based on the attention mechanism. In this paper, ResNet50 is used as the backbone network, and the channel attention mechanism CA and SimAM parameter-free attention mechanism are fused so that the network focuses on the focus region of the image and ignores the background region, and it is also proposed to replace the first downsampling module in ResNet50 with the interval sampling module SC to enhance the network's ability to detect small targets. The experimental results show that the proposed method is able to acquire effective features and ignore background regions, which is beneficial to the recognition and judgment of facial expressions, and it achieves the highest accuracy of 76.45% on the self-collected dataset, which is a 3.1% improvement in the accuracy rate relative to the traditional face emotion recognition method.

Key words: emotional recognition; attention mechanism; interval sampling; feature extraction

0 引 言

人脸情绪识别是计算机视觉领域的一个重要研究方向,其目的是通过分析人脸表情来推断人的情绪状态。人脸情绪识别在许多领域都有重要的应用,如疲劳驾驶、智能医疗、犯罪测谎等。传统的人脸情绪识别方法主要有基于几何的方法与基于整体两种方法。传统识别方法依赖于前期人工提取特征的优劣,人为干扰因素较大。与传统方法不同的是深度学习方法,这种方法有许多网络框架,如 VGG、AlexNet、ResNet 等,被广泛应用于人脸情绪识别。Cheng 等[1]在 VGG19 基础上优化了网络结构及其参数,并且采用迁移学习技术克服训练样本不足的问题,提高了情绪识别的准确率;卢官明[2]提出了一种基于深度残差网络的人脸情绪识别方法,利用残差单元解决网络深度和模型收敛性之间的矛盾,能够提升表情识别的准确率;Zhong 等[3]在 ResNet 的基础上移除了 Softmax,引入丢弃层并对全连接进行修改,使网络参数减少,同时将激励模块(SE)加入到网络中,取得了较高的准确率;Shahzad 等[4]提

基金项目: 浙江省医学电子与数字健康重点实验室开放课题(MEDH202206)。
作者简介: 胡 蝶(1998-),女,硕士研究生,主要研究方向:深度学习。
通讯作者: 朱耀东(1970-),男,博士,教授,主要研究方向:智能机器人与健康物联网。Email:zhuyaodong@163.com
收稿日期: 2023-05-04

出一种基于深度学习方法,使用面部特征分区,划分和定位主要面部图像,用于识别深层面部情绪。

注意力机制最初是用于机器翻译,但 Marrero 等[5]提出一种新的带有注意力机制的人脸情绪识别网络结构,该网络的注意力机制模块的使用基于 U-Net,对输入的人脸图像进行人脸分割并输出一张掩膜,再把特征提取模块输出的特征图张量和掩膜进行融合,以去除特征图张量中与表情无关的特征,从而取得更好的情绪分类效果;Wang 等[6]提出一个新型区域注意力网络,能够较好地处理遮挡和姿势变化下的面部表情识别,通过主干卷积网络产生的各个区域特征,聚合并嵌入到紧凑的固定长度特征中,以提高其准确性;Gan 等[7]提出了一种多注意力机制融合的网络,处理复杂条件下的人脸表情识别问题,该网络包含 2 个模块:区域感知模块和表情识别模块,通过区域感知模块学习掩码,用于定位与表情相关的重要区域,再通过表情识别模块学习具有强区分度的特征;兰凌强等[8]以 ResNet 为基础网络,融合了瓶颈注意力机制及全局二阶池化层,平衡和改善特征数据分布情况,提高表情识别准确率;Vats 等[9]提出一个面部情绪识别框架,该框架建立在 Swin Vision Transformers 和激励块(SE)的基础上,提出了一种基于注意力机制的转换器模型来解决视觉任务。

但在实际应用中人脸图像采集环境比较复杂,采集到的人脸图像会包含大量的非人脸表情信息,如光线、姿态、角度、遮挡等,这些干扰信息会极大地增加人脸情绪识别的难度,导致情绪识别准确率下降。由于环境比较复杂,现存的识别方法大多无法有效的从背景复杂的人脸图像中准确剥离出重要表情区域,导致特征中往往会包含较多冗余信息,影响识别的准确率。

因此,本文提出了基于注意力机制的人脸情绪识别方法,通过关注图片中重要的信息而忽略不相关的背景区域,从而使网络模型获取更好的性能。

1 基于注意力机制的人脸情绪识别模型

近年来,随着深度学习技术的发展,基于卷积神经网络的人脸情绪识别方法取得了显著的进展,主要工作可分为 3 个方面:设计新型网络架构,增加网络广度和深度等来提高性能;研究不同的优化方法,如加入注意力机制来优化特征提取;探索不同损失函数监督网络训练。然而,由于不同表情的面部特征之间存在巨大的差异,难以充分利用所有的面部信息来准确地预测情绪。因此,本文引入注意力机制来提高模型性能。

在神经网络中加入注意力机制模块,使得卷积神经网络能够更加关注输入图像中的重要区域的信息,拥有更好的性能。

1.1 通道注意力机制(CA)

本文设计了一种通道注意力机制(CA),目的是让网络更加关注特征图张量中的通道信息,其结构如图 1 所示。

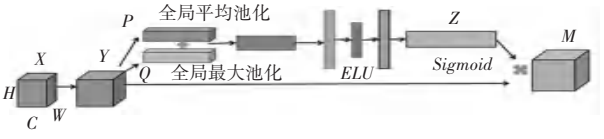


图 1 CA 模块

Fig. 1 Channel attention block

在 CA 模块中,首先对于输入的特征图 $X \in \mathbb{R}^{W' \times H' \times C'}$ 进行一个普通的卷积操作,通过卷积核的映射,得到新的特征图 $Y \in \mathbb{R}^{W \times H \times C}$,旨在通过空间卷积提取更加丰富的特征信息;其次,对 $Y \in \mathbb{R}^{W \times H \times C}$ 进行全局平均池化和全局最大池化,得到特征图 $P \in \mathbb{R}^{1 \times 1 \times C}$ 和 $Q \in \mathbb{R}^{1 \times 1 \times C}$,采用两个池化层是为了弥补空间信息在迁入通道时的不足之处;将全局平均池化和全局最大池化得到的特征描述向量相加,再通过两个全连接层和 ELU 激活函数进行特征融合和非线性变换,最终得到通道注意力向量 $\boldsymbol{Z} \in \mathbb{R}^{1 \times 1 \times C}$,该向量代表了各个通道的重要性权重;最后,与 $Y \in \mathbb{R}^{W \times H \times C}$ 进行点乘运算,得到特征图 $M \in \mathbb{R}^{W \times H \times C}$。这一过程相当于对原始特征图进行了自适应的重新加权,使得网络能够更加关注于重要的通道信息,从而提高了特征的表达能力和网络的性能。

通道注意力是一种软注意力机制,在卷积神经网络中,特征图张量经过卷积层中不同的卷积核后,每一个卷积核都会输出一张新的特征图,比如特征图经过一个包含 n 个卷积核的卷积层之后,就会输出 n 个通道的新特征图。本文中将设计的通道注意力机制模块放在 ResNet50 网络的最大池化层之前,加入通道注意力机制 CA 后的网络架构如图 2 所示。

1.2 无参注意力机制(SimAM)

SimAM 无需向原始网络添加参数,而是在一层中推断特征图的 3D 关注权重。该模块的另一个优点是大多数算子是根据定义的能量函数的解来选择的,避免在结构调整上花费太多精力。CA 模块缺乏空间和通道同时变化的灵活性,因此本文提出 CA

模块和 SimAM 结合，改善了网络结构，添加在网络的第一层和最后一层，较好的提升了网络的性能。融合 CA 和 SimAM 后的网络架构如图 3 所示。

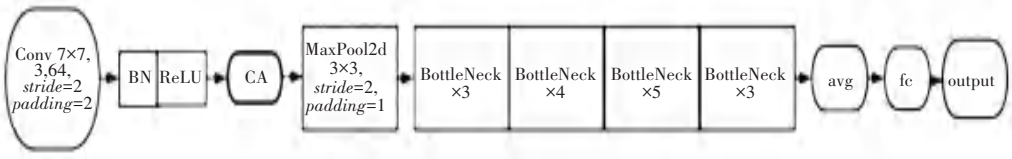


图 2　加入通道注意力机制 CA 后的网络架构

Fig. 2　Network architecture after adding the channel attention mechanism CA

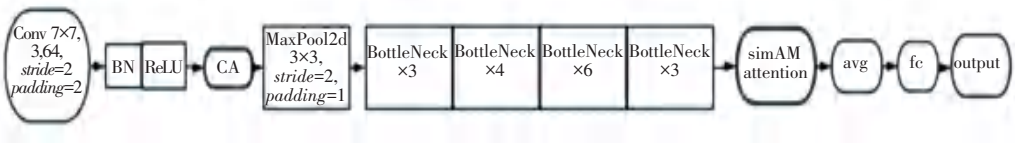


图 3　融合 CA 和 SimAM 后的网络架构

Fig. 3　Network architecture after convergence of CA and SimAM

1.3　间隔采样模块(SC)

ResNet50 网络框架中，网络前端有一个步距为 2 的下采样过程，这个过程会丢失一些小目标的信息，甚至会导致网络忽略一些有效信息。因此本文提出了 SC 模块，以此来增强网络对小目标的检测能力，结构如图 4 所示。将输入特征图分成 4 组，每组特征在 x 轴和 y 轴方向上分别间隔取样。假设原始特征图的尺寸为 $H \times W \times C$，则分组卷积后得到 4 个尺寸为 $1/2H \times 1/2W \times C$ 的特征图；将这 4 个特征图在通道维度上进行拼接，得到尺寸为 $1/2H \times 1/2W \times 4C$ 的特征图。为了保持通道数与原始特征图一致，对拼接后的特征图进行一次卷积操作，将通道数降低至 C。本文在以 ResNet50 为主干网络，将干结点单元中的第一个 7×7 下采样模块的步长修改为 1，并在其后加入 SC 模块。

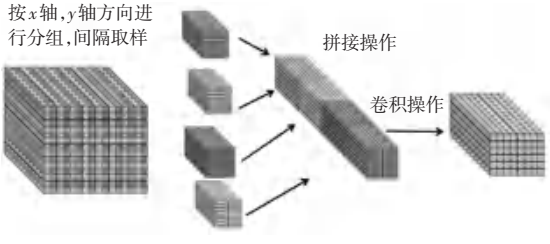


图 4　SC 模块

Fig. 4　SC block

2　实验

2.1　数据集

本文采用的是自采集的数据集，总共 1 880 张图片，包含 7 种表情标签，即厌恶、开心、惊讶、恐惧、悲伤、中性、愤怒，其中训练集一共 1 501 张图片，验证集一共 379 张图片。为了构建所需的数据集，采用网络搜索和人工拍摄两种方法，获得了多样的图片来检验模型训练效果。为了增强数据集的多样性，还特意收集了一些年轻人和儿童的表情图片，数据集图片示例如图 5 所示。

图 5　数据集图片示例

Fig. 5　Examples of pictures on dataset

2.2　实验参数

实验在 linux 5.15.0 系统，以 pytorch1.7.1 作为基础框架来编写程序，在训练过程使用随机梯度下降来优化交叉熵损失，学习率最开始设置为 0.001，epoch 为 30，60，90 时调整学习率，冲量为 0.9，权重衰减 weight_decay 为 0.000 1。

2.3　评价指标

2.3.1　准确率 (*Accuracy*)

正确分类的测试实例个数占测试实例总数的比例，公式(1)：

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

其中，TP 是指被检索到正样本，实际也是正样本；FP 是指被检索到正样本，实际是负样本；FN 是指未被检索到正样本，实际是正样本；TN 是指未被检索到正样本，实际也是负样本。

准确率的使用有一定的局限性。比如当样本不平衡时,假设有 99 个负例,一个正例。如果模型将这些样本全部预测成了负例,准确率为 99%,这显然是不合理的,此时准确率就失去了衡量模型性能的作用。因此本文中还引入了其他判断标准。

2.3.2 精确率 (*Precision*)

被正确检索的样本数与被检索到样本总数之比,公式(2):

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

精确率反映预测为正类的样本中有多少是预测对的,衡量模型预测的精确性。比如模型预测出了 100 个正类,但实际上只有 50 个是预测正确的,剩余 50 个是误分类,那么这个类别对应的精确率为 50%,性能就比较低下。

2.3.3 召回率 (*Recall*)

被正确检索的样本数与应当被检索到的样本数之比,公式(3):

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

2.3.4 *F*1 值(*F*1_Score)

综合考虑精准率与召回率,引入一个新指标 *F*1 -Score, 公式(4):

$$F1\_Score = 2 \times \frac{1}{\frac{1}{Precision} + \frac{1}{Recall}} \tag{4}$$

2.4 结果分析

为了验证本文所提方法的有效性,以 ResNet50 作为主干网络,加入本文提出的一种通道注意力机制 CA,将步距为 2 的下采样模块替换为 SC 模块,在自采集数据集上进行了对比实验和消融实验,不同模型在自采集数据集上的准确率见表 1。

表 1 不同模型在自采集数据集上的准确率

Table 1 Accuracy rate of different models on datasets

方法	准确率/%	精确率/%	召回率/%	*F*1 值
ResNet50	73.81	71.92	71.42	71.49
ResNet50+SC	75.13	72.57	72.41	72.44
ResNet50+CA	75.13	73.18	72.78	72.65
Resnet50+SC+CA	75.40	73.68	72.68	72.89

从表 1 中可以看出,本文在自采集数据集上加入 SC 模块后准确率达到了 75.13%,相较于基础模型提升了 1.32%;在基础模型中加入 CA 模块之后,无论是准确率还是召回率,相比于基础模型有所提高。实验结果表明所提出的通道注意力机制 CA 能很好地关注网络中的有效信息,提高了网络的非线性能力,从而提高了情绪识别的准确率。本文在加入 CA 模块的基础上加入 SC 模块,可以看到 *F*1 分数和准确率同样有所提升,证明改进方法能够防止网络丢失有效信息,辅助网络进行有效分类。

本文设计通道注意力机制模块后,还引入了其他注意力机制,希望进一步提高模型的分类能力,加入不同注意力机制在自采集数据集上的准确率见表 2。

表 2 加入不同注意力机制在自采集数据集上的准确率

Table 2 Accuracy of adding different attention mechanisms to self collected datasets

方法	准确率	精确率	召回率	*F*1 值
ResNet50	73.81	71.92	71.42	71.49
BAM_ResNet50	73.81	72.19	72.25	72.11
ECA_ResNet50	74.07	71.82	72.04	71.81
ACmix_ResNet50	74.87	72.77	72.09	72.18
coord_ResNet50	75.13	72.14	72.04	72.03
simAM_ResNet50	76.19	73.64	73.31	73.39

从表 2 中可以看出,在自采集数据集上,加入 BAM, ECA, ACmix, coordinate attention, simAM 后,准确率分别为 73.81%, 74.07%, 74.87%, 75.13%, 76.19%,改进后准确率均有所提升,证明部分图片含有遮挡、光照等干扰的前提下,加入注意力机制可以使网络更关注重要信息,抑制背景噪声,提高分类能力。

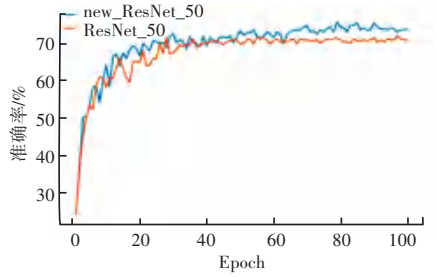
相较于 CA 模块,可以观察到 SimAM 对情绪识别的任务更具有影响力,相对于 ResNet50 基础模型准确率提升 2.38%。进一步做消融实验,将两种注意力机制进行结合,查看网络性能,实验结果见表 3。

表 3 混合注意力机制的准确率

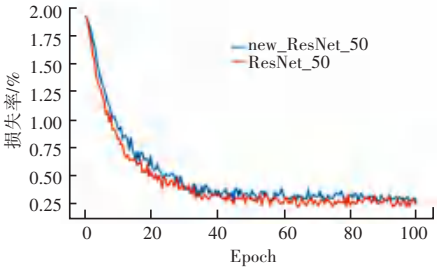
Table 3 Accuracy of mixed attention mechanisms

方法	准确率
ResNet50	73.81
ResNet50+SimAM	76.19
ResNet50+CA	75.13
ResNet50+SimAM+CA_	76.45
ResNet50+CA+SimAM+SC	76.91

本文融合通道注意力机制 CA 和 SimAM 无参注意力机制,同时用 SC 模块替换 ResNet50 中的第一个下采样模块,从表 3 中可以看到本文所提方法相较于其他模型准确率有所提升。改进模型和基础模型准确率和损失率对比如图 6 所示,可见改进后网络模型更加完善。



（a）准确率对比



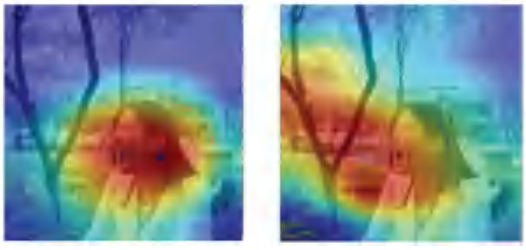
（b）损失率对比

图 6　基础模型与改进后模型准确率和损失率

Fig. 6　Accuracy and loss rate of basic model and improved model

2.5　Grad-CAM 可视化

本文应用 Grad-CAM 进行特征可视化，开心类别模型关注的区域如图 7 所示，可以明显看到改进后的网络模型更能关注图像中的重要区域，而不太关注图片中的背景区域。



（a）改进模型　　（b）基础模型

图 7　开心类别的热力图

Fig. 7　Heat map of the happy category

悲伤类别的热力图如图 8 所示。颜色越偏红色，说明模型更关注这块区域。

（a）基础模型　　（b）改进模型

图 8　悲伤类别的热力图

Fig. 8　Heat map of grief categories

3　结束语

本文主要探讨了基于注意力机制的人脸情绪识别方法，使用 ResNet50 作为主干网络，提出了一种融合通道注意力机制 CA 和 SimAM 无参注意力机制的方法，使网络关注图像的重点区域，忽略背景区域，同时提出用 SC 模块替换 ResNet50 中的第一个下采样模块，来增强网络对小目标的检测能力。实验结果表明，所提出的方法能够有效的提取面部特征，忽略背景区域，在自行采集的数据集上达到了 76.91% 的准确率，相对于传统的人脸情绪识别方法，改进模型的准确率提高了 3.1%。

然而，本文的研究也存在一些局限性。本文仅考虑了少数注意力机制，在未来的研究中，可以考虑引入更多的注意力机制，并结合其他技术，如迁移学习、深度强化学习等，探索更好的人脸情绪识别方法。

参考文献

[1] CHENG S, ZHOU G H. Facial expression recognition method based on improved VGG convolutional neural network [J]. International Journal of Pattern Recognition and Artificial Intelligence, 2020, 34(7): 2056003.

[2] 卢官明，朱海锐，郝强，等. 基于深度残差网络的人脸表情识别[J]. 数据采集与处理，2019，34(1)：50-57.

[3] ZHONG Y, QIU S, LUO X, et al. Facial expression recognition based on optimized ResNet[C]// Proceedings of 2020 2nd World Symposium on Artificial Intelligence (WSAI). IEEE, 2020: 84-91.

[4] SHAHZAD T, IQBAL K, KHAN M A, et al. Role of zoning in facial expression using deep learning[J]. IEEE Access, 2023, 11: 16493-16508.

[5] MARRERO F P D, GUERRERO P F A, REN T, et al. Feratt: Facial expression recognition with attention net[J]. arXiv preprint arXiv:1902.03284, 2019.

[6] 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.

[7] GAN Y, CHEN J, YANG Z, et al. Multiple attention network for facial expression recognition[J]. IEEE Access, 2020, 8: 7383-7393.

[8] 兰凌强，刘淇缘，卢树华. 基于注意力机制与特征相关性的人脸表情识别[J]. 北京航空航天大学学报，2022，48(1)：147-155.

[9] VATS A, CHADHA A. Facial emotion recognition[J]. arXiv preprint arXiv:2301.10906, 2023.