工程科学学报, 第 47 卷, 第 6 期: 1272−1284, 2025 年 6 月
Chinese Journal of Engineering, Vol. 47, No. 6: 1272−1284, June 2025
https://doi.org/10.13374/j.issn2095-9389.2024.07.24.002; http://cje.ustb.edu.cn

基于 U-Net 架构改进 VGG19 模型的人脸表情识别方法

赵小虎[1,2)]，张景怡[1,2)✉]，焦明之[1,2)]，谢礼逊[1,2)]，王兰飞[1,2)]，孙维青[1,2)]，张　狄[1,2)]

1) 中国矿业大学信息与控制工程学院, 徐州 221008　2) 矿山互联网应用技术国家地方联合工程实验室(中国矿业大学), 徐州 221008
✉通信作者，E-mail: ts23060022a31@cumt.edu.cn

摘　要　针对传统面部识别技术中存在的诸多问题, 如网络模型对关键通道特征的关注不足、参数量过大以及识别准确率不高等, 本文提出了一种基于改进 Visual Geometry Group 19(VGG19)模型的全新方案. 该方案融合了 U-Net 网络架构的设计理念, 并引入了改进的 SEAttention 模块, 以期提高模型的收敛速度和对面部细节的关注程度. 在保持 VGG19 深层特征提取能力的基础上, 通过特定设计的卷积层和跳跃连接, 实现了对特征的高效融合与优化. 经过改进的 VGG19 模型, 不仅能更好地提取面部特征, 还能在保证准确率的前提下, 降低模型参数, 提高运算效率. 为了验证改进模型的效果, 利用 FER2013 数据集和 CK+两个数据集对本文提出的模型进行了测试. 实验结果显示, 改进后的 VGG19 网络在表情识别的准确率上分别取得了 1.58% 和 4.04% 的提升. 这一结果充分证明了本文提出的方法在解决传统面部识别问题方面的优越性, 也为面部识别技术的进一步发展提供了新的思路.

关键词　面部表情识别; 深度学习; 卷积神经网络; 情感分类; VGG19
分类号　TP391.41

U-Net-based VGG19 model for improved facial expression recognition

ZHAO Xiaohu[1,2)], *ZHANG Jingyi*[1,2)✉], *JIAO Mingzhi*[1,2)], *XIE Lixun*[1,2)], *WANG Lanfei*[1,2)], *SUN Weiqing*[1,2)], *ZHANG Di*[1,2)]

1) School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221008, China
2) National and Local Joint Engineering Laboratory of Internet Application Technology on Mine (China University of Mining and Technology), Xuzhou 221008, China
✉Corresponding author, E-mail: ts23060022a31@cumt.edu.cn

ABSTRACT　In response to the challenges faced by traditional facial recognition techniques, such as insufficient focus on key channel features, large number of parameters, and low recognition accuracy, this study proposes an improved VGG19 model that incorporates concepts from the U-Net architecture. While maintaining the deep feature extraction capability of VGG19, which is well-regarded in the field, the model employs specially designed convolutional layers and skip connections. The use of feature cropping and stitching techniques allows the model to efficiently integrate multi-scale features, thereby enhancing the robustness and effectiveness of facial expression recognition tasks. This design ensures the seamless integration of features from different layers, which is crucial for accurate facial expression recognition, as it maximizes the information yielded from each layer. Additionally, this paper introduces an improved SEAttention module, specifically designed for facial expression recognition tasks. The innovation of the SEAttention module lies in replacing the original activation function with the Mish activation function, which can dynamically adjust the weights of different channels to enhance performance. This adjustment ensures that important features are emphasized while redundant features are suppressed, streamlining the recognition process. This selective focus significantly speeds up the convergence of the network and improves the ability of the model to detect subtle changes in facial expressions, which is especially valuable in nuanced emotional contexts. Furthermore, modifications are made to the fully connected layers by substituting the first two layers with convolutional layers

收稿日期: 2024−07−24
基金项目: 江苏省自然科学基金资助项目(BK20200649); 国家自然科学基金资助项目(62001475)

while retaining the fully connected final layer. This change reduces the number of nodes in these layers from [4096, 4096, 1000] to just [7], effectively addressing the large parameter size in the VGG19 network. Additionally, this modification improves the resistance of the model to overfitting, making it more robust when applied to new data. Extensive experiments were conducted on the FER2013 and CK+ datasets, demonstrating that the improved VGG19 model significantly enhanced recognition accuracy by 1.58% and 4.04%, respectively, compared to the original version. Furthermore, the parameter efficiency of the model was thoroughly evaluated, which indicated a substantial reduction in the overall parameter count without compromising performance. This balance between model complexity and accuracy highlights the practical applicability of the proposed method in real-world facial recognition scenarios, ensuring that it can be deployed in environments with limited computational resources. In conclusion, integrating the U-Net architecture and enhanced SEAttention module into the VGG19 network led to significant advancements in facial expression recognition. The improved model not only boosts performance in terms of feature extraction and fusion but is also adept in solving the pressing problems of parameter size and computational efficiency. These innovations contribute to achieving state-of-the-art performance in facial expression recognition, making the proposed method an important contribution to advancing computer vision and deep learning. The robustness and efficiency of the proposed method highlight its potential for various applications requiring accurate real-time facial expression analysis, such as human-computer interaction, security systems, and emotion-driven computing. Future work will explore the adaptability of the model to other datasets and additional optimization techniques, aiming to further enhance its performance and expand its applicability across diverse use cases.

KEY WORDS facial expression recognition; deep learning; convolutional neural network; emotion classification; VGG19

面部表情是面对面交流中一种强有力的非语言交流手段, 能够让人类传递各种信息, 包含的情绪特征线索较多[1]. 面部表情识别(FER, Facial emotion recognition)在许多领域具有重要应用. 在安全领域, FER 可以帮助监控系统识别潜在的安全威胁或犯罪行为; 在医疗领域, FER 能够辅助诊断情感障碍和精神疾病, 提高诊疗效果; 在驾驶员疲劳监控中, FER 可以实时检测驾驶员的疲劳状态, 预防交通事故; 在人机交互中, FER 使机器更加智能化和人性化, 提供个性化服务和增强用户体验.

人脸表情识别过程中最为关键的就是对人脸特征的提取, 传统的表情识别研究主要依赖手工特征提取, 例如局部二值模式[2]、方向梯度直方图等. 这些传统方法由于将特征分析和推断过程分开, 往往无法有效处理包含大量变异源的复杂数据集. 在光照变化和部分遮挡的情况下, 这些方法也难以捕捉到细微表情变化. 为提高识别性能, 将特征提取与分类整合, 结合深度学习技术来实现自动 FER[3−4], 成为了新的研究趋势.

随着深度学习的发展, 基于卷积神经网络(CNN)的深度学习模型被广泛应用于 FER 任务[5−6]. 在 FER2013 数据集[7] 上, Khaireddin 等[8] 基于 VGG19 架构的 SOTA 模型证实了 VGG Net[9] 的优越性, 但该模型训练时间较长且参数量庞大, 可能导致准确性达到瓶颈. 为了解决深度网络训练中的效率问题, He 等[10] 提出了一种残差学习框架, 形成了 ResNet, 这使得高达 152 层的网络能有效调优. Pramerdorfer 等[11] 在 FER2013 数据集上对三种不同架构(VGG Net、Inception Net[12] 和 ResNet)的性能进行了比较分析, 结果显示在分类面部情绪方面 VGG Net 仍然表现优越. Li 等[13] 进行了胶囊网络和卷积网络在面部表情识别任务中的鲁棒性比较. 这项研究不仅探讨了两种网络结构的性能差异, 还为理解不同深度学习模型在表情识别中的适用性和稳定性提供了重要见解. Shan 等[14] 提出了一种 CNN 架构, 能够发现面部表情的更深层次特征表示. 该系统由输入模块、预处理模块、识别模块和输出模块组成. Shanthi 等[15] 通过整合局部区域的特征来增强对表情变化的识别能力. Yu 等[16] 提出一种基于多通道融合和轻量级神经网络的面部表情识别方法, 该方法通过将传统特征提取算法与深度学习特征提取算法相结合, 有效提取出更完整的图像特征, 提高了面部表情识别的准确性和鲁棒性.

近年来, 注意力机制在 FER 中的重要性与日俱增. Hu 等[17] 提出了 SE-Net, 这是一个简单而有效的模块, 突出了卷积块的更基本特征. Zhu 等[18] 提出的级联注意力网络结合了注意力机制与金字塔特征, 由 3 个模块组成:局部和多尺度立体空间上下文特征提取模块、级联注意力模块和时间序列特征提取模块. 该网络充分利用上下文信息来弥补空间特征的缺失, 增强了注意力机制的性能, 提高识别人脸的准确性. Zhong 等[19] 在其研究中将通道和空间注意力模块(CBAM)嵌入到深度残差网络(ResNet)中. 这种方法不仅增强了模型对特

征的感知能力, 还通过引入权重损失函数来优化网络, 有效解决了数据分布不均的问题. Ren 等[20]通过引入注意力机制, 使其能够自动地忽略不相关的信道, 并将重点放在关键信道上, 以增强鲁棒性和准确性. 然而, 该方法仅在 Maxpool 和 Avgpool 两个层次添加了注意力机制, 没有考虑 VGG 网络自身的多个参数特性, 因而在模型参数方面仍需进行改进.

多网络融合是采用足够多样的卷积神经网络来提取尽可能多的特征层, 通过合适的集成方法来高效融合这些神经网络[21]. 如 Verma 等[22]提出使用 Visual 和 Landmark 两个分支融合对输人数据进行特征提取, Visual 分支利用浅层神经网络提取图像序列中的中低层次特征, Landmark 分支处理更高层次的特征, 提取脸部坐标的位置信息. 实验结果表明, 该融合网络在 CK+数据集[23]上的性能明显优于其他方法. Jung 等[24]提出了两个相互协作的深度网络模型. 其一是基于多帧外观的 DTAN 网络, 其二是基于原始人脸坐标点提取时间几何特征的 DTGN 网络, 二者结合并采用新型集成方法以提升人脸表情识别性能. 最终实验结果表明该方法在相应数据集上取得了较好的识别效果. Yang 等[25]将 VGG19 与胶囊网络相级联, 以软池化层替换原网络中的最大池化层, 在训练中更好地保留了局部信息及细粒度特征, 具有抗过拟合的作用. Vignesh 等[26]提出了一种新颖面部情感识别(FER)模型, 该模型通过在视觉几何组(VGG)层之间接入 U-Net[27]构建分割层, 执行分割操作以提取和突出特征图中的关键特征, 同时控制通过 VGG 层的冗余信息流, 将分割块的输出与输入特征图结合, 形成最终的特征图, 显著提高了网络的特征提取效果.

综合上述分析可知, 虽然深度学习[28−30]尤其是卷积神经网络(CNN)在特征自动提取方面具有优势[31−32], 但许多现有模型仍然难以充分整合多层次、多尺度的特征信息. 传统的网络架构通常关注于提取高层次抽象特征, 而忽视了低层次细节的重要性, 导致在面对细腻的表情变化时性能受限. 并且, 现有方法在不同个体、文化背景和情感维度的多样性上, 往往缺乏适应性[33], 造成对特定群体或环境的识别效果不佳. 此外, 模型训练效率、参数量等问题也在一定程度上制约了其发展[34−35].

因此如何设计卷积神经网络来提取更加全面和更深层次的人脸表情特征, 仍然是当前人脸表情识别研究的热点问题. 针对这一问题, 本文采用多网络融合方法, 首先选择 VGG19 作为基础网络, 融合 U-Net 网络架构, 利用 U-Net 特有的跳跃连接机制, 能够更好地捕捉多尺度特征. 在技术实现上, 通过特征裁剪和拼接技术以有效的增强特征流动性, 此外, 引入了改进的 SE 注意力机制, 使得网络能够在特征提取过程中自适应地关注最具辨识性的区域. 综上所述, 本文提出的方法能够结合深度学习中的多个前沿技术, 通过系统性地优化网络结构、引入新的特征聚焦机制以及简化模型参数, 全面提高了人脸表情识别的性能, 可为其他相关研究提供新的思路.

1　相关理论

1.1　传统 VGG 网络

传统 VGG19 网络模型如图 1 所示. VGG19 有 19 个层, 由 16 个卷积层组成, 其中有 3 个全连接层 (Fully connected layer, FC Layer) 和 5 个池化层. 从图中可以看出, 该算法的特征提取模块包括 5 个子模块, 各子模块分别由 3×3 的卷积核和最大池化层叠加构成. 卷积核是特征抽取的关键, 它利用卷积函数对输入图像进行滑动操作, 从多个位置对输入图像进行特征提取, 并将其组合为一张特征图. 图像通过卷积后, 高度和宽度都没有改变, 而图像中的通道数量即为所用的卷积核数目. 由卷积核得到的特征是从底层到上层的深层语义. 最大池化层的主要功能是对原始矩阵进行降维, 使其长度和宽度各为其本身的 1/2. 最后是 3 个全连接层, 前两个具有 4096 个通道, 第 3 个具有 1000 个通道的全连接层进行类别的分类. 最后 Softmax 分类器给出每一种类的概率, 最大的就是预测结果.

VGG 网络将原来的 5×5 和 7×7 的大卷积核替换成 3×3 的小卷积核, 在获得同样的感知野的情况下, 不但能更好地捕获图像的局部细节, 还能增加神经网络的深度, 提高识别的准确率.

1.2　传统 U-Net 网络

传统 U-Net 网络模型如图 2 所示. 本文网络主要参考了 U-Net 网络的 U 形结构, 该结构不仅能够有效地降低训练样本的需求, 同时又能保持较高的准确率, 它由两大部分构成: 第一部分是主干特征提取网络, 它通过四次连续的下采样阶段来捕获输入图像的关键特征. 每个下采样阶段都包括两个卷积层和一个最大池化层, 以实现特征的逐步细化. 第二部分的特征融合网络通过上采样将具有高层抽象特性的低分辨率图像转化为高分辨图像, 它将这些特征与左侧的低层次但高分辨

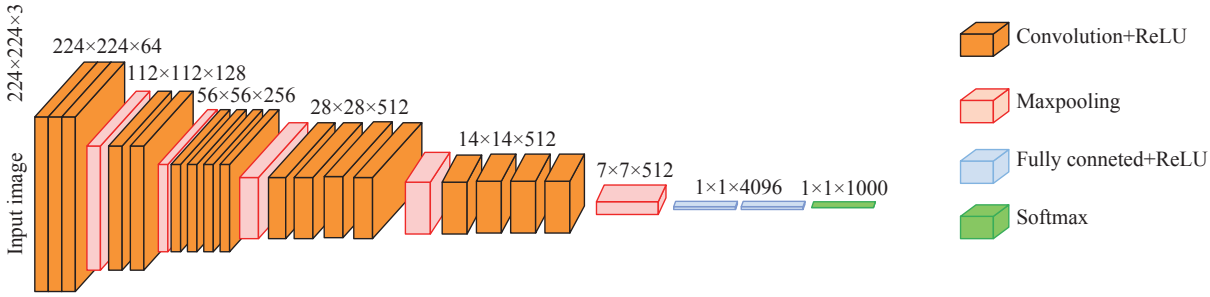

图 1 VGG19 网络结构图

Fig.1 VGG19 network structure diagram

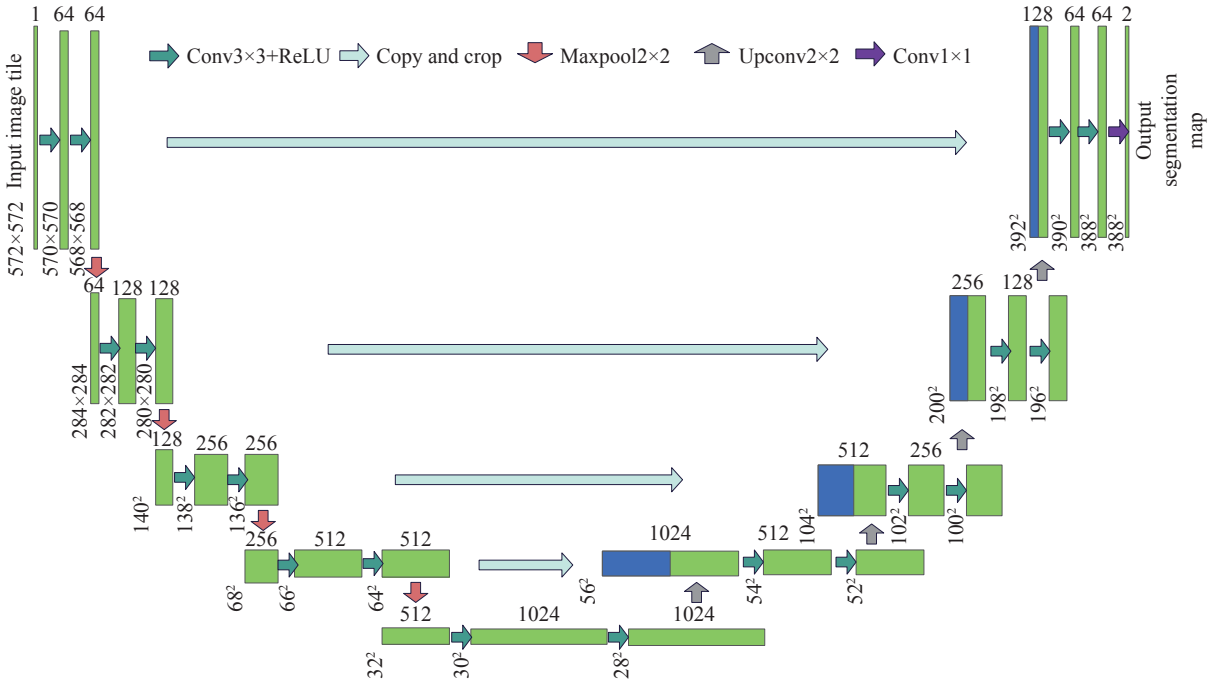

图 2 U-Net 网络结构图

Fig.2 U-Net network structure diagram

率的特征图像相结合，进行特征融合操作，以此来重建图像的细节特征并恢复其原始的维度，这一特性使得模型在分类时能够有效地识别不同类别的细微差异. 在此基础上，使用两个卷积核大小为 1×1 的卷积进行分类，以生成最终的两个热图（Heatmap）. 其中，第一张热图代表第一类的得分，第二张热图则表示第二类的得分. 这两张热图将作为 Softmax 函数的输入，用于计算类别概率. 通过比较得到较高概率的类别，我们再根据这些结果计算损失，并进行反向传播，从而优化模型参数，有效提升分类性能.

2 改进网络

2.1 改进的 U-VGG 网络结构

本文的 U-VGG 图像分类网络结构，是一种融合了 VGG19 和 U-Net 两种深度学习模型的经典特性的新型网络. 在 FER2013 数据集上，由于样本量大且多样性高，利用 VGG19 的深层次特征提取和 U-Net 的多尺度特征融合，有助于应对数据集的多样性，提高表情识别的鲁棒性；在 CK+数据集上，尽管数据集较小，但 VGG19 深层网络的强大特征提取能力和 U-Net 的跳跃连接仍能够捕捉复杂的表情细节.

本文对原始网络结构进行了改进，设计了一个由五个阶段组成的特征提取模块网络的总体结构如图 3 所示.

具体地，Stage1-Stage5 阶段，每个阶段都集成了多层卷积神经网络、BN 处理、PReLU 激活函数以及改进 SEAttention 模块. 这些组合模块协同工作，逐步提炼并加强输入图像的深层特征，显著提升了特征提取深度和精确度. 多层卷积神经网络能够有效提取从低级到高级的丰富特征，同时使

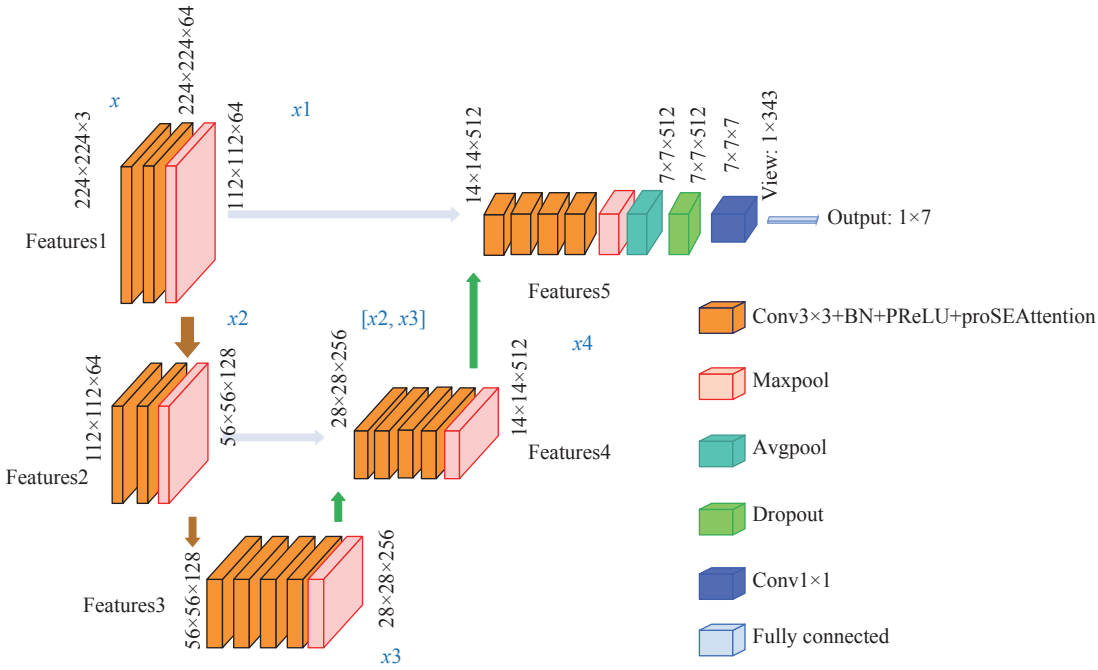

图 3　改进后的网络结构图

Fig.3　Improved network structure diagram

用多个 3×3 小卷积核实现细致特征捕获, 这不仅减少了参数数量, 还提升了模型性能. BN 处理将输出规范化为均值 0, 方差 1 的正态分布, 并将其作为神经网络输入, 使其远离激活函数的饱和区, 进入非线性函数的敏感区, 从而解决梯度消失的问题, 并且可以在样本的训练次序被打乱的情况下, 提升准确率. 为了使不同层次阶段特征得到灵活流动, 受 U-Net 网络思想启发, 在 Stage2 和 Stage4, Stage1 和 Stage5 之间, 采用了裁剪操作 (crop_tensor 函数) 和特征拼接 (torch.cat) 技术来实现不同层次特征的融合. 如公式(1)所示:

$$x4 = \text{Stage4}\{\text{cat}[\text{crop}(x2), x3]\} \tag{1}$$

裁剪操作主要是对较大尺寸特征图进行裁剪, 以确保其尺寸与目标特征图相匹配. 特征拼接与通道压缩则是将来自不同层次的特征图进行拼接, 并利用 1×1 卷积对各通道进行压缩、融合. 通过这种特征融合的策略, 网络能够充分整合来自不同层次的特征信息, 从而显著提升了特征的表达能力和辨识度.

针对原始 VGG19 网络使用 ReLU 作为激活函数存在的一些弊端, 如“死亡 ReLU”问题, 本文引入了 PReLU(Parametric rectified linear unit), 顾名思义, 它是带参数的 ReLU. 其数学表达式如(2)所示:

$$f(y) = \begin{cases} x_{(i)}, & x > 0 \\ a_{(i)}x_{(i)}, & x < 0 \end{cases} \tag{2}$$

当 $a_{(i)}$=0, 则 PReLU 退化为 ReLU; 当 $a_{(i)}$ 是一个很小的固定值 (如 $a_{(i)}$=0.01), 则 PReLU 退化为 Leaky ReLU(LReLU). PReLU 激励函数具有更强的抗过拟合能力, 更快的收敛速度和更高的精度. 尽管该方法引入了极少量的参数, 但它对网络计算负担的增加以及引发过拟合风险的可能性均极为有限. 特别的, 当不同通道使用相同的 $a_{(i)}$ 时, 参数的数量更是大幅减少.

2.2　改进的通道注意力模块

SEAttention 模块是本文改进网络的重要组成部分, 从注意力机制思想出发, 该方法通过捕捉全局信息来处理不同信道间的相互关系, 对特征权重比例再调节, 提高了从表情图片中提取特征的质量. SEAttention 模块最初由 Hu 等在 2018 年提出, 其灵感来源于人类快速识别图像全局信息, 以突出需要关注的候选区域, 从而获取更多目标细节. 其核心方法是自动地学习各通道间的相互关系, 并为每个通道动态地赋予相应的权重, 从而使模型能够更好地关注和利用重要的特征信息, 提高模型的表示能力和鲁棒性. 具体结构如图 4 所示.

SEAttention 模型中包含三个操作, 挤压 (Squeeze) 操作 (F_{sq})、激励 (Excitation) 操作 (F_{ex})、缩放 (Scale) 操作 (F_{scale}). 给定一个 input 特征图 $\boldsymbol{X}$, 经过 Transformation(F_{tr}) 操作生成特征图 $\boldsymbol{U}$, 尺寸为 $H\times W\times C$, 挤压操作对其进行了全局平均池化, 将特征图的 $H\times W$ 维度降至 1×1. 公式如(3)所示:

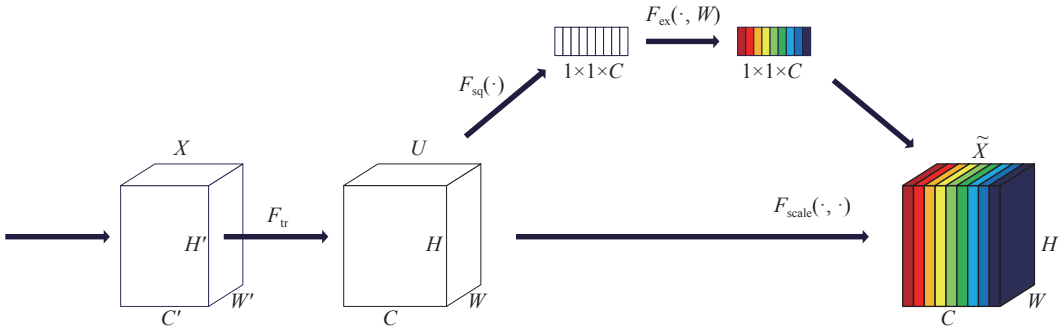

图 4 SEAttention 注意力机制原理图[13]

Fig.4 SEAttention attention mechanism schematic diagram[13]

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i, j) \qquad (3)$$

其中, z_c 是压缩后的通道特征, $u_c(i,j)$ 表示特征图 $\boldsymbol{U}$ 在位置 (i,j) 和通道 c 上的值.

激励操作通过两层全连接层完成, 通过权重 W 生成所需的权重信息, 其中 W 是通过学习得到的, 用来显示需要的特征相关性. 通过两个全连接层 W_1, W_2, 对上一步得到的向量 z 进行处理, 得到通道权重值 s, 经过两层全连接层后, s 中不同的数值表示不同通道的权重信息, 赋予通道不同的权重. 式如(4)所示:

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)) \qquad (4)$$

其中, σ 表示 Sigmoid 激活函数, g 表示全连接层的函数.

缩放操作通过注意力权重重新加权原始特征图, 强调重要性特征. 其尺寸大小与特征图完全一样, SE 模块不改变特征图的小大, 式如(5)所示:

$$\tilde{x}_c = F_{scale}(u_c, s_c) = s_c u_c \qquad (5)$$

其中, u_c 表示输入特征图 U 的第 c 个通道, s_c 表示激励步骤得到的第 c 个通道的权重.

注意力机制依赖于非线性函数来掌握信道之间的复杂关系. 但使用的激活函数为 ReLU (Rectified linear unit), 如图 5(a) 所示, 作为分段非线性的阈值激活函数, 其中横坐标代表函数输入 x, 纵坐标代表函数输出 y, 在输入大于 0 的情况下, 它的斜率是 1, 表现出线性特性; 当输入值为负时, ReLU 的梯度为 0, 对应的权重将不会得到更新, 导致神经元的“死亡”, 进一步影响关键特征的有效传递. 因此, 本文将 ReLU 激活函数替换为 Mish 激活函数, 以增强网络的表达力和信息传递.

Mish 激活函数如图 5(b) 所展示, 其中横坐标代表函数输入 x, 纵坐标代表函数输出 y, 其输出值在 [–0.31, +∞) 范围内, 没有上界, 这有助于防止梯度饱和. 对于正输入值, Mish 函数没有最大值限制, 从而允许梯度自由流动, 确保网络参数得到有效更新. 对于负输入值, Mish 函数避免了神经元失活的问题. Mish 激活函数通过消除 ReLU 的“Dying ReLU”问题, 提升了网络的学习能力. 尽管 Mish 函数的梯度在接近 0 时可能会减缓训练速度, 但其下界的存在提供了一种强正则化效应. 与 ReLU 相比, Mish 连续可微特性, 避免了奇点问题.

2.3 改进表情分类器

通常, 传统面部表情识别网络会通过全连接层将卷积层提取的特征图转换为一维向量, 并通过 softmax 函数来预测表情类别, 如公式(6)所示:

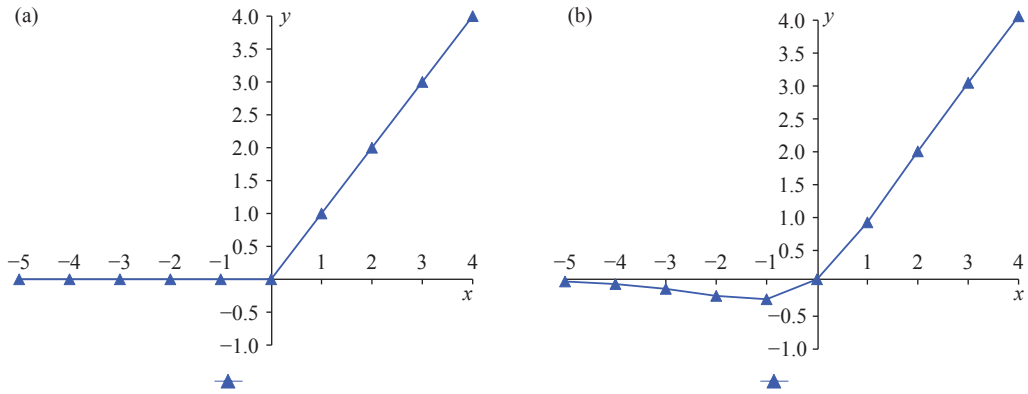

图 5 (a) ReLU 激活函数; (b) Mish 激活函数

Fig.5 (a) ReLU activation function; (b) Mish activation functions

$$p_i = \mathrm{soft\,max}(x_i) = \frac{\mathrm{e}^{x_i}}{\sum_{i=1}^{n} \mathrm{e}^{x_i}} \tag{6}$$

式中, p_i 表示预测的概率, x_i 表示待分类的样本. 所有概率之和被限制在 0 到 1 之间. 但这种方法可能会引起参数冗余、降低识别效率, 并增加过拟合的风险. 为了解决这些问题, 本文提出了改进型表情分类器, 如图 6 所示.

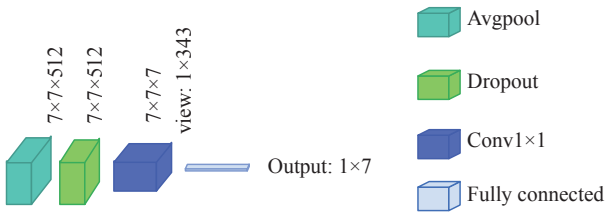

图 6　改进的表情分类器

Fig.6　Improved emotion classifier

首先采用全局平均池化处理卷积层的特征, 以提取区域内最具代表性的特征, 同时保持通道数不变, 减小特征图的尺寸. 其次, 引入 Dropout 正则化技术, 通过随机失活 50% 的神经元来抑制过拟合. 在每次反向传播中, 按一定比例随机选取隐藏节点, 使得全连通网络具有一定的稀疏性, 由于隐藏节点以一定的概率随机出现, 使得两个神经元不会同时出现, 减少了神经元之间的协作自适应关系, 提高了网络的鲁棒性. 接着 1×1 卷积层用于将特征提取部分 (Stage1 到 Stage5) 提取的高维特征映射到更低维的空间, 主要功能包括降维、特征混合和适应分类任务. 它将 512 个通道的高维特征图压缩到 7 个通道, 减少了参数数量和计算复杂度, 同时有效融合了不同通道间的信息. 最后, 全连接层将经过 1×1 卷积层处理后的特征图展平为一维向量, 通过线性变换实现分类映射, 输出具体的分类结果. 通过引入 1×1 卷积层和全连接层的组合, 改进后的分类器部分不仅降低了计算复杂度, 增强了特征融合能力, 还显著优化了分类性能, 为实现更高的图像分类准确率提供了有力支持.

3　实验及结果分析

3.1　实验环境

在本研究中, 选择了 Windows 10 操作系统, 并在 PyCharm 集成开发环境中使用 Python 3.8 版本来训练模型. 深度学习框架为 Pytorch1.1.0, 硬件平台 CPU 为 i7-11800H, 使用了一块 RTX 4090D 显卡, 显存大小为 24GB. 模型训练中, 使用了随机梯度下降作为优化器, 并采用交叉熵损失函数(Cross Entropy Loss). 在 FER2013 和 CK+数据集上, 训练的迭代周期(epoch)分别设置为 250 和 60.

3.2　数据集与数据预处理

在实验环节, 本文采用了 FER2013[7] 和 CK+[23] 两个面部表情数据集来评估所提出的面部表情识别模型的泛化能力, 确保其在多样化数据上的有效性. FER2013 数据集是在 2013 年 Kaggle 比赛时收集的大型表情数据库, 其中包含了 48×48 的灰度图人脸图像共 35887 张, 其中训练集 28708 张, 包括公有验证集 3589 张、私有验证集 3589 张, 共涉及 7 种情绪类别: 愤怒、厌恶、害怕、快乐、悲伤、惊讶和中性. 在 CK+数据集中有 123 个人, 共 593 个表情的视频序列样本. 在本实验中, 仅取 3 张有比较明显面部情绪的图片, 并将其剪切成 48×48 像素大小. 一共有 7 个表情, 依次是愤怒、厌恶、害怕、开心、悲伤、惊讶及蔑视. 两个数据集分布如表 1 所示.

表 1　数据集图片分布

Table 1　Image distribution of dataset

Emotion (class)	Class total (FER2013)	Class total (CK+)
Angry	4953	135
Disgust	547	177
Fear	5121	75
Happy	8989	207
Sad	6077	84
Surprise	4002	249
Neutral	6198	54
Total	35887	981

本研究采用了多种数据增强技术对 FER2013 数据集进行了预处理, 以提高模型泛化能力和性能. 训练集预处理包括随机裁剪至 44 像素、应用 Cutout 数据增强方法、随机水平翻转以及将图像转换为 PyTorch 的 Tensor 格式. 这些预处理手段提升了训练集多样性, 促进了模型深入学习和泛化能力. 测试集预处理则采用了十裁剪方式, 分别对图像进行中心裁剪、四角裁剪及其水平翻转后裁剪, 并将裁剪结果转换为 Tensor 并堆叠成一个批次, 以确保测试过程稳定性和一致性. 这些数据预处理策略有效地提升了模型在训练和测试阶段的表现.

3.3　实验及结果分析

3.3.1　可视化实验分析

首先, 本文所提出的模型在 FER2013 和 CK+

数据集上进行了训练, 并在训练趋于稳定时自动停止了当前的训练. 最终生成了训练过程中训练集和测试集的损失(Loss)曲线及准确率(Accuracy)曲线, 如图 7 所示.

在损失曲线上, 训练集和测试集的损失值随着训练轮次的增加逐渐下降, FER2013 数据集的损失曲线在第 200 轮时趋于收敛, CK+数据集在接近 30 轮时收敛, 表明模型在进一步优化过程中效果逐渐稳定. 与此相对应, 在准确率曲线上, 训练集和测试集的准确率也随训练轮次的提升而上升, FER2013 数据集在接近第 220 轮时达到了最高准确率 73.75%. 然而, 测试集的最终准确率比训练集低约 20%, 这与 FER2013 数据集中存在的样本不均衡和标注误差等问题有关. CK+数据集在 43 轮时达到了最高准确率 97.98%, 这些曲线共同反映了模型对 CK+数据集的良好适应性和较强的识别能力.

图 8(a) 和 (b) 分别呈现了 VGG 网络以及改进的轻量化面部表情识别模型在 FER2013 验证集上的混淆矩阵. 混淆矩阵的水平轴对应预测表情类别, 垂直轴对应实际表情类别, 对角线上的值表示模型对不同表情类别的识别准确率. 识别准确率越高, 对应单元格的颜色越深. 通过混淆矩阵的分析, 可以看出本研究所提出的模型在各类表情的平均识别率达到了 73.75%, 相较于传统 VGG 网络提高了 1.58%, 在当前研究中表现更为优异.

图 8(c) 和图 8(d) 分别展示了原 VGG 模型和改进后的人脸表情识别模型在 CK+数据集上的混淆矩阵验证结果. 改进模型在 CK+数据集上的识别准确率达到了 97.98%, 相较于原始 VGG 模型, 准确率提升了 4.04%. 这一结果验证了本研究所提出的模型在不同数据集上具有良好的泛化能力.

观察图 8(b) 和图 8(d) 可以发现, 在两个数据集中, 识别快乐表情的准确率均超过了 90%, 这可能是因为快乐表情的图像数量较多, 使得模型更容易学习并识别. CK+数据集的混淆矩阵分析显示, 除了轻蔑表情外, 其他表情的识别准确率均达到了 100%. 轻蔑表情容易与恐惧表情混淆, 分析测试集中的图像后发现, 轻蔑表情的样本数量较少, 这可能导致了在训练过程中的欠拟合. 由于恐惧和轻蔑表情可能同时伴有额头紧皱或皱眉的特征, 这可能会导致模型做出错误的判断, 影响检测

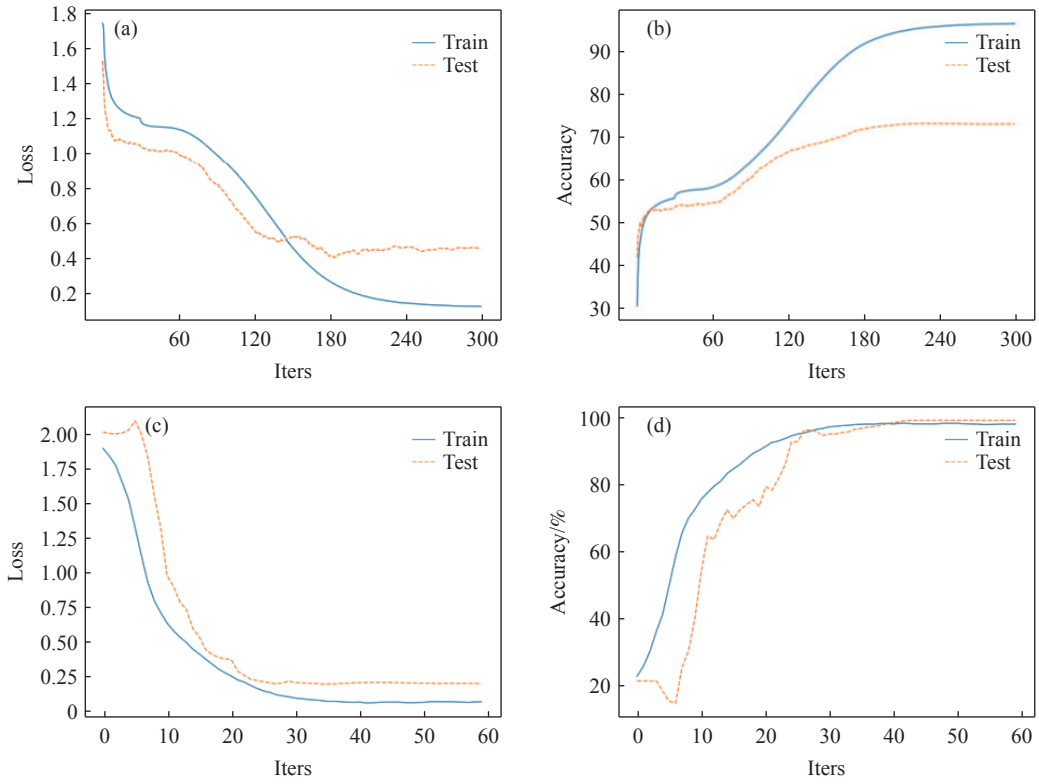

图 7 (a) FER2013 数据集上模型训练损失曲线; (b) FER2013 数据集上模型训练准确率曲线; (c) CK+数据集上模型训练损失曲线; (d) CK+数据集上模型训练准确率曲线

Fig.7 (a) Loss curve of the model training on the FER2013 dataset; (b) accuracy curve of the model training on the FER2013 dataset; (c) loss curve of the model training on the CK+ dataset; (d) accuracy curve of the model training on the CK+ dataset

的准确率.

本文网络的 Stage1 输出的 64 通道特征图如图 9 所示.

本文网络不同阶段的注意力特征图可视化如图 10 所示. 首先, 从输入层到输出层, 特征图的复杂性和抽象程度逐渐增加, 初期阶段主要捕捉低级特征(如边缘和纹理), 而后续阶段能够识别更复杂的模式和特征(如形状和对象), Stage1 的输出几乎保留了原始图像的所有信息, 而 Stage5 的输出已经超出了人类直观理解的范围. 可以这样理解, 随着层数的加深, 关于图像视觉内容信息越来越少, 而关于类别的信息就越来越多. 体现了特征提取的一个层次性过程: 从简单的低级特征到复杂的高级特征. 其次, 各阶段的注意力分布变化表明模型在不同特征上赋予了不同的关注度, 例如初期对背景或噪声有较强关注, 而后期则集中

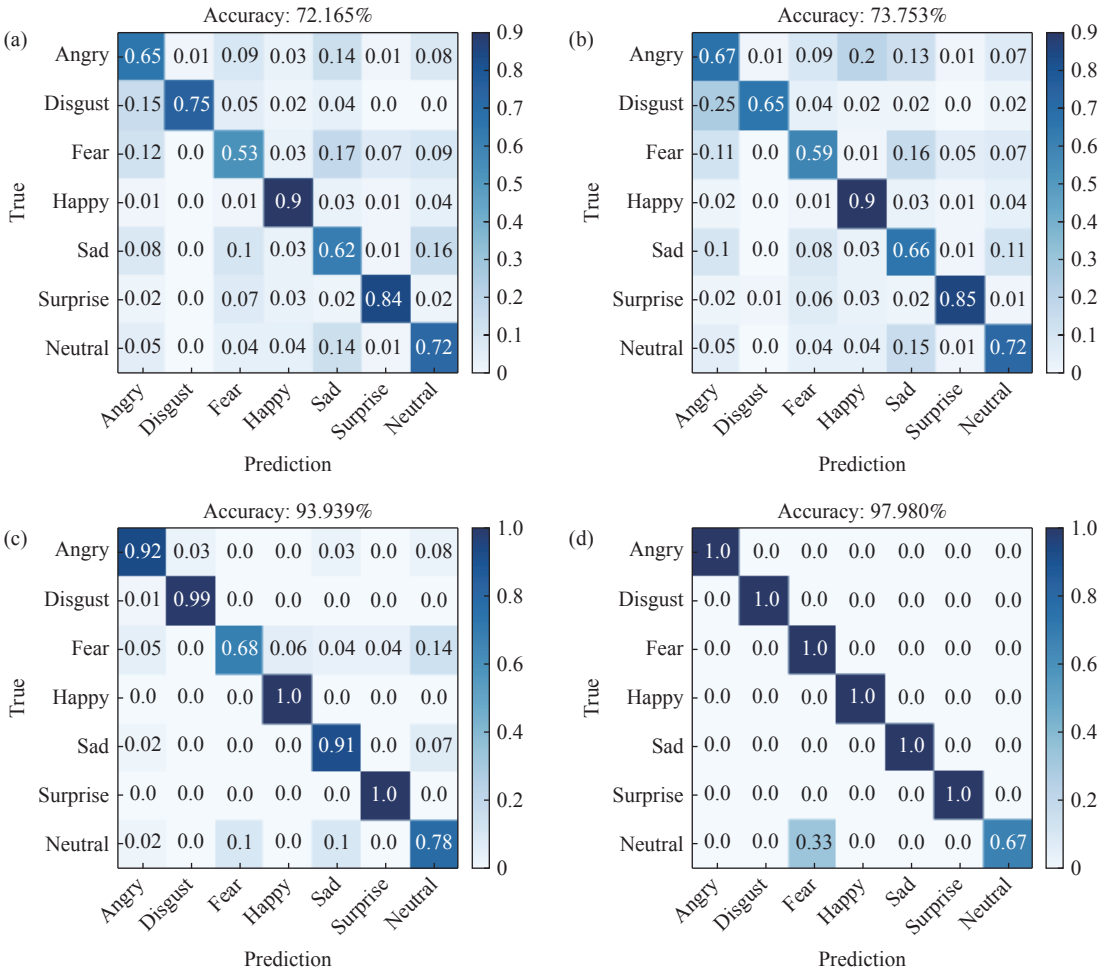

图 8　(a) FER2013 数据集上优化前网络的混淆矩阵; (b) FER2013 数据集上改进网络得到的混淆矩阵; (c) CK+数据集上优化前网络的混淆矩阵; (d) CK+数据集上改进网络得到的混淆矩阵

Fig.8　Confusion matrix obtained from the (a) original VGG network on the FER2013 dataset; (b) improved VGG network on the FER2013 dataset; (c) original VGG network on the CK+ dataset; (d) improved VGG network on the CK+ dataset

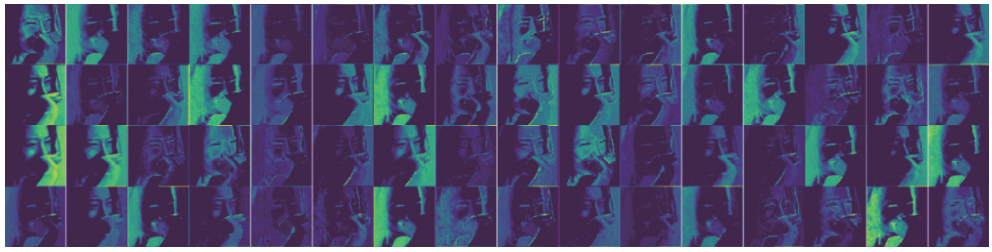

图 9　Stage1 输出的 64 通道特征图

Fig.9　64-channel feature map output from Stage 1

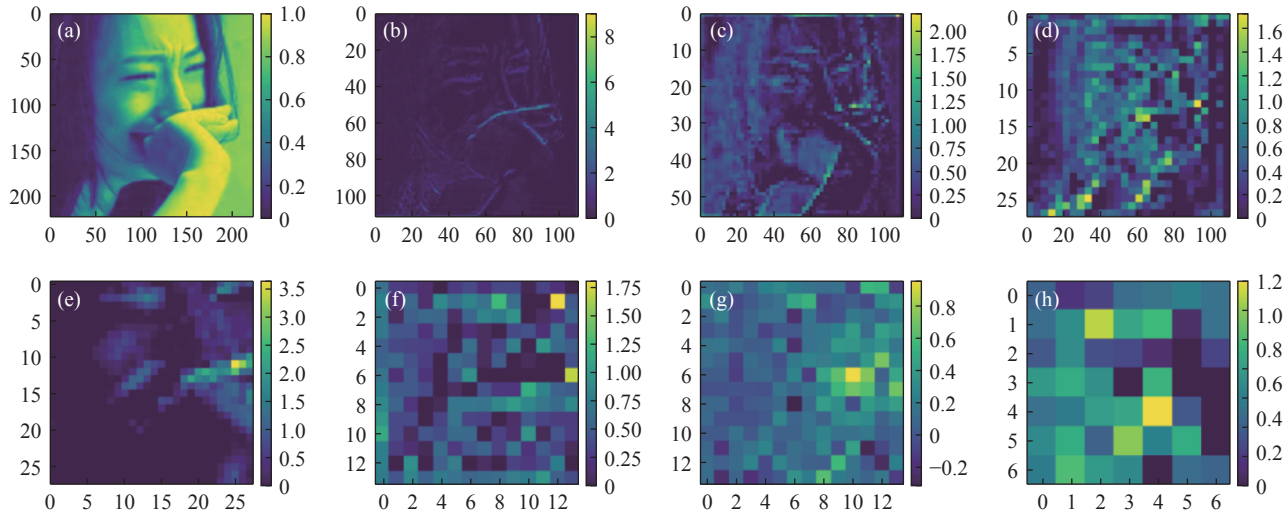

图 10 本文网络不同阶段的注意力特征图可视化. (a) 原始图像; (b) Stage 1 输出的特征图 x_1; (c) Stage 2 输出的特征图 x_2; (d) Stage 3 输出的特征图 x_3; (e) x_3 和 x_2 裁剪后拼接; (f) Stage 4 输出的特征图 x_4; (g) x_4 和 x_1 裁剪后拼接; (h) Stage 5 输出的特征图 x_5

Fig.10 Visualization of attention feature maps at different stages of the network: (a) Original image; (b) the feature map x_1 output from Stage 1; (c) the feature map x_2 output from Stage 2; (d) the feature map x_3 output from Stage 3; (e) concatenate x_3 and x_2 after cropping; (f) the feature map x_4 output from Stage 4; (g) concatenate x_4 and x_1 after cropping; (h) the feature map x_5 output from Stage 5

于关键信息, 如特定物体的边界. 综上所述, 图 10 不仅展现了网络不同阶段的特征提取能力, 也为模型的优化和可解释性提供了重要见解, 指引了今后在深度学习模型调优和解释性研究方面的方向.

3.3.2 对比实验分析

首先, 通过准确率指标对比了提出的模型与其他经典分类模型的表现, 准确率数学表达式如公式(7)所示.

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \tag{7}$$

式中, TP 是真阳性, TN 是真阴性, FP 是假阳性, FN 是假阴性. 本研究选取了一系列不同的网络架构作为对照组, 以便在 FER2013 数据集上进行比较实验, 结果如表 2 所示.

由表 2 可以看出, 本文所提出的模型在 FER2013 数据集的准确率为 73.75%, 与 VGGNET、S-ResNet、VGG19、UCNN、Deep-Emotion 这 5 种方法相比分别提升 1.26%, 1.25%, 1.06%, 5.10%, 3.73%. 在 CK+ 数据集上本文方法的准确率为 97.98%, 分别高于 JAIN 的 93.2% 和 Em-AlexNet 的 94.25%, 此外同样高于 MIANet 的 95.76%、SCAN 的 97.31%、DeRL 的 97.30%. 实验结果表明, 与现有的方法相比, 该算法在面部表情识别的准确率上实现了显著提升.

精确率表示模型预测为正类别的样本中有多少是真正的正类别, 计算方式如式(8)所示:

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}} \tag{8}$$

召回率, 也称为 True positive rate (TPR) 或灵敏度, 是指在所有实际为正类别的样本中, 模型能够正确预测为正类别的比例. 其计算方法如式(9)所示:

$$\text{Re call} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{9}$$

F-score 是一个综合指标, 通过调和平均的方式平衡了精确率与召回率, 从而全面反映了模型

表 2 FER2013 数据集和 CK+数据集上不同基干网络的准确率比较

Table 2 Accuracy of different backbone networks on FER2013 dataset and CK+dataset

Dataset	Models	Accuracy/%	Dataset	Models	Accuracy/%
FER2013	VGGNET[7]	72. 49	CK+	JAIN[39]	93.20
	S-ResNet[36]	72.50		Em-AlexNet[3]	94.25
	VGG19	72.69		MIANet[40]	95.76
	UCNN[37]	68.65		SCAN[4]	97.31
	Deep-Emotion[38]	70.02		DeRL[41]	97.30
	Method in this article	73.75		Method in this article	97.98

的性能. 其计算方式如式(10)所示:

$$\text{F-score}=2\times\frac{\text{Precison}\times\text{Recall}}{\text{Precision}+\text{Recall}}=\frac{2\text{TP}}{2\text{TP}+\text{FP}+\text{FN}} \tag{10}$$

表 3 展示了在微观、宏观和加权三种不同平均方法下, 两个面部表情识别数据集(FER2013 和 CK+)的 Precision、Recall 和 F-score. 从表格中可以看出, 在 FER2013 数据集上, Macro 平均方法下的 Precision、Recall 和 F-score 分别为 0.723、0.705 和 0.707, 高于 Micro 和 Weighted 平均方法. 这表明算法在各类别上表现较为均衡, 但在样本数量少的类别上的表现稍差. 在 Weighted 平均方法下的 Recall 为 0.727, 高于其他平均方法, 说明算法在样本较多的类别上的召回率较高. 对于 CK+数据集, 所有平均方法下的 Precision 和 Recall 均为 0.953, Weighted 平均方法下的 Precision 和 Recall 均为 0.982, F-score 略有不同, 但都高于 0.95, 表明在这个数据集上算法性能非常优越. 总体而言, FER2013 数据集更具挑战性, 算法在样本较多的类别上表现更好, 而在 CK+数据集上, 算法表现更加稳定且优越.

表 3　在不同类别平均方法下评估所提出方法的其他指标

Table 3　Evaluation of proposed method on other metrics under different class averaging methods

Metric	Macro	Micro	Weighted
Precision (FER2013)	0.723	0.702	0.683
Recall (FER2013)	0.705	0.702	0.727
F-score (FER2013)	0.707	0.702	0.702
Precision (CK+)	0.953	0.953	0.982
Recall (CK+)	0.953	0.953	0.982
F-score (CK+)	0.971	0.953	0.959

为了验证在保证准确率的前提下并不增加计算复杂度, 本文进行了模型复杂度分析实验, 实验结果如表 4 所示. 该表格比较了多种模型在参数量、准确率和浮点运算每秒(FLOPs)方面的表现. VGG16 和 VGG19 是经典卷积神经网络, 具有较高参数量, 分别为 138.3M 和 140.1M, 以及相对较好的准确率, 分别为 70.38% 和 72.49%, 但其计算复杂度也较高, FLOPs 分别为 15.8M 和 16.2M. 本文提出方法在参数数量上仅为 20.7M, 却达到了最高准确率 73.75%. 在 FLOPs 方面, 该方法也表现出优于 VGG16 和 VGG19 的性能. 因此, 本文方法在实际应用中具有更大的灵活性和可操作性, 为在计算资源受限的环境中应用深度学习技术提供了切实可行的解决方案.

表 4　不同模型的参数比较

Table 4　Comparison of parameter counts for different models

Models	The parameter count/10^6	Accuracy/%	FLOPs/10^6
VGG16	138.3	70.38	15.8
VGG19	140.1	72.49	16.2
FER-SoC[34]	102.4	66.0	9.6
KLS-Net[42]	1.95	69.21	0.03
LA-Net[35]	15.03	70.25	0.283
Method in this article	20.7	73.75	14.7

3.4　消融实验

消融实验数据见表 5. 从表中可以看出, 基线模型在 CK+数据集和 FER2013 数据集上的准确率分别为 94.747% 和 69.044%. 在测试中添加了十倍裁剪和 BN 层后进行实验, 准确率分别提升至 96.030% 和 71.293%, 相较于基线模型分别提高了 1.283% 和 2.249%. 将网络架构更改为 U-Net 后, 准确率达到 96.652% 和 72.264%, 相较于基线模型分别提升了 1.905% 和 3.22%. 添加改进的 SE 注意力机制后, 准确率进一步提升至 97.323% 和 72.699%, 分别提高了 2.576% 和 3.655%. 最终改进的网络在 CK+数据集和 FER2013 数据集上的准确率分别达到 97.980% 和 73.753%, 相较于基线模型分别提高了 3.233% 和 4.709%. 由此可以看出, 在网络中同

表 5　CK+和 FER2013 数据集消融实验结果

Table 5　Results of ablation experiments on the CK+ and FER2013 datasets

Ten-fold cropping	Cutout	Improved SE attention	Batch normalization	U-Net framework	CK+ dataset accuracy/%	FER2013 dataset accuracy/%
					94.747	69.044
√			√		96.030	71.293
√			√	√	96.652	72.264
√		√	√	√	97.323	72.699
√	√		√	√	97.379	73.264
√	√	√	√	√	97.980	73.753

时利用各个模块更加有效, 从而大大改善了网络模型的性能.

4 结论

综上所述, 本研究通过将 U-Net 结构和优化后的 SEAttention 模块嵌入至 VGG19 模型, 成功构建了一种先进的面部表情识别方法. 该方法在特征提取和融合能力方面取得了显著提升, 并有效提高了面部表情识别的准确率和模型的参数效率. 实验结果表明, 改进后的 VGG19 网络在 FER2013 数据集上的识别准确率达到了 73.75%, 在 CK+数据集上的识别准确率达到了 97.98%. 改进的网络结构在模型的鲁棒性和识别精度方面表现出更优异的性能, 展示了该方法在实际应用中的潜力. 未来研究将继续探索更高效的方法来进一步提升特征提取的能力.

参 考 文 献

[1] Zhang L, Tjondronegoro D. Facial expression recognition using facial movement features. *IEEE T Affect Comput*, 2011, 2(4): 219

[2] Shan C F, Gong S G, McOwan P W. Facial expression recognition based on local binary patterns: A comprehensive study. *Image Vision Comput*, 2009, 27(6): 803

[3] Yang X, Shang Z H. Facial expression recognition based on improved AlexNet. *Laser Optoelectron Prog*, 2020, 57(14): 141026
(杨旭, 尚振宏. 基于改进 AlexNet 的人脸表情识别. 激光与光电子学进展, 2020, 57(14): 141026)

[4] Gera D, Balasubramanian S. Landmark guidance independent spatio-channel attention and complementary context information based facial expression recognition. *Pattern Recognit Lett*, 2021, 145: 58

[5] Wang Y, Su W J, Liu H L. Facial expression recognition based on linear discriminant locality preserving analysis algorithm. *J Inform Comput Sci*, 2013, 9(11): 4281

[6] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *Comput Sci*, 2014, 1409: 1556

[7] Goodfellow I J, Erhan D, Luc Carrier P, et al. Challenges in representation learning: A report on three machine learning contests. *Neural Netw*, 2015, 64: 59

[8] Khaireddin Y, Chen Z F. Facial emotion recognition: State of the art performance on FER2013 [J/OL]. *arXiv preprints*, (2021–04–20) [2024–07–23]. https://arxiv.org/abs/2105.03588

[9] Cui Z Y, Pi J T, Chen Y, et al. Combining and improving the facial expression recognition of VGGNET and Focal Loss. *Comput Eng Appli*, 2021, 57(19): 171
(崔子越, 皮家甜, 陈勇, 等. 结合改进 VGGNet 和 Focal Loss 的人脸表情识别. 计算机工程与应用, 2021, 57(19): 171)

[10] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition // 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*). Las Vegas, 2016: 770

[11] Pramerdorfer C, Kampel M. Facial expression recognition using convolutional neural networks: State of the art [J/OL]. *arXiv preprints*, (2023–01–21) [2024–07–23]. https://arxiv.org/abs/1612.02903v1

[12] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions // 2015 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*). Boston, 2015: 1

[13] Li D H, Zhao X C, Yuan G J, et al. Robustness comparison between the capsule network and the convolutional network for facial expression recognition. *Appl Intell*, 2021, 51(4): 2269

[14] Shan K, Guo J Q, You W W, et al. Automatic facial expression recognition based on a deep convolutional-neural-network structure // 2017 *IEEE 15th International Conference on Software Engineering Research, Management and Applications* (*SERA*). London, 2017: 123

[15] Shanthi P, Nickolas S. An efficient automatic facial expression recognition using local neighborhood feature fusion. *Multimed Tools Appl*, 2021, 80(7): 10187

[16] Yu Y L, Huo H, Liu J Q. Facial expression recognition based on multi-channel fusion and lightweight neural network. *Soft Comput*, 2023, 27(24): 18549

[17] Hu J, Shen L, Sun G. Squeeze-and-excitation networks // 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, 2018: 7132

[18] Zhu X L, He Z L, Zhao L, et al. A cascade attention based facial expression recognition network by fusing multi-scale spatio-temporal features. *Sensors*, 2022, 22(4): 1350

[19] Zhong R, Jiang B, Li N X, et al. Facial expression recognition method embedded with attention mechanism residual network. *Comput Eng Appl*, 2023, 59(11): 88
(钟瑞, 蒋斌, 李南星, 等. 嵌入注意力机制残差网络的人脸表情识别方法. 计算机工程与应用, 2023, 59(11): 88)

[20] Ren Z Q, Chi X, Zhu Y J, et al. Research on facial expression recognition method combined with attention mechanism. *Comput Era*, 2022(3): 24
(任志强, 迟杏, 朱煜君, 等. 结合注意力机制的面部表情识别方法研究. 计算机时代, 2022(3): 24)

[21] Hong H Q, Shen G P, Huang F H. Summary of expression recognition technology. *J Front Comput Sci Technol*, 2022, 16(8): 1764
(洪惠群, 沈贵萍, 黄风华. 表情识别技术综述. 计算机科学与探索, 2022, 16(8): 1764)

[22] Verma M, Kobori H, Nakashima Y, et al. Facial expression recognition with skip-connection to leverage low-level features // 2019 *IEEE International Conference on Image Processing* (*ICIP*). Taipei, 2019: 51

[23] Lucey P, Cohn J F, Kanade T, et al. The extended Cohn-kanade dataset (CK): A complete dataset for action unit and emotion-

specified expression // 2010 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition – Workshops*. San Francisco, 2010: 94

[24] Jung H, Lee S, Yim J, et al. Joint fine-tuning in deep neural networks for facial expression recognition // 2015 *IEEE International Conference on Computer Vision* (*ICCV*). Santiago, 2015: 2983

[25] Yang S, Wang J D, Jiang Y J, et al. Research on expression recognition model cascading with VGG19 and CapsNet of SoftPool. *Semicond Optoelectron*, 2021, 42(6): 897
(杨双, 王敬东, 姜宜君, 等. 结合 SoftPool 的 VGG19 与 CapsNet 相级联的表情识别模型研究. 半导体光电, 2021, 42(6): 897)

[26] Vignesh S, Savithadevi M, Sridevi M, et al. A novel facial emotion recognition model using segmentation VGG-19 architecture. *Int J Inf Technol*, 2023, 15(4): 1777

[27] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation // *Medical Image Computing and Computer-Assisted Intervention – MICCAI* 2015. Munich, 2015: 234

[28] Debnath T, Reza M M, Rahman A, et al. Four-layer ConvNet to facial emotion recognition with minimal epochs and the significance of data diversity. *Sci Rep*, 2022, 12(1): 6991

[29] Sakthi D V, Ezhumalai P. RF-GCN: Residual fused-graph convolutional network using multimodalities for facial emotion recognition. *Trans Emerg Telecommun Technol*, 2024, 35(9): e5031

[30] Dong W M, Zheng X W, Zhang L F, et al. Attentional visual graph neural network based facial expression recognition method. *Signal Image Video Process*, 2024, 18(12): 8693

[31] Mohana M, Subashini P. Facial expression recognition using machine learning and deep learning techniques: A systematic review. *SN Comput Sci*, 2024, 5(4): 432

[32] Subathradevi S, Preethiya T, Santhi D, et al. Facial emotion recognition for feature extraction and ensemble learning using hierarchical cascade regression neural networks and random forest. *J Circuits Syst Comput*, 2024, 33(18): 2550011

[33] Chouhayebi H, Mahraz M A, Riffi J, et al. Human emotion recognition based on spatio-temporal facial features using HOG-HOF and VGG-LSTM. *Computers*, 2024, 13(4): 101

[34] Vinh P T, Quang Vinh T. Facial expression recognition system on SoC FPGA // 2019 *International Symposium on Electrical and Electronics Engineering* (*ISEE*). Ho Chi Minh, 2019: 1-4

[35] Ma H, Celik T, Li H C. Lightweight attention convolutional neural network through network slimming for robust facial expression recognition. *Signal Image Video Process*, 2021, 15(7): 1507

[36] Wu Y H, Chen X H. Facial expression recognition system based on improved ResNet. *Inf Commun*, 2020, 33(7): 37
(吴宇豪, 陈晓辉. 基于改进的 ResNet 的人脸表情识别系统. 信息通信, 2020, 33(7): 37)

[37] He C, Hou M. Facial expression recognition method based on improved convolutional neural network. *Inf Technol*, 2022, 46(5): 107
(何超, 侯明. 基于改进卷积神经网络的人脸表情识别方法. 信息技术, 2022, 46(5): 107)

[38] Minaee S, Minaei M, Abdolrashidi A. Deep-emotion: Facial expression recognition using attentional convolutional network. *Sensors*, 2021, 21(9): 3046

[39] Jain D K, Shamsolmoali P, Sehdev P. Extended deep neural network for facial emotion recognition. *Pattern Recognit Lett*, 2019, 120: 69

[40] Luo S S, Li M J, Chen M. Multi-scale integrated attention mechanism for facial expression recognition network. *Comput Eng Appl*, 2023, 59(1): 199
(罗思诗, 李茂军, 陈满. 多尺度融合注意力机制的人脸表情识别网络. 计算机工程与应用, 2023, 59(1): 199)

[41] Yang H Y, Ciftci U, Yin L J. Facial expression recognition by de-expression residue learning // 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, 2018: 2168

[42] Liu J, Luo X S, Xu Z X. Weight inference and label smoothing for lightweight facial expression recognition. *Comput Eng Appl*, 2024, 60(2): 254
(刘劲, 罗晓曙, 徐照兴. 权重推断与标签平滑的轻量级人脸表情识别. 计算机工程与应用, 2024, 60(2): 254)