

# 基于边云协同的 ECA 轻量化人脸表情识别<sup>①</sup>



李文浩<sup>1,3,4</sup>, 张宗帅<sup>2,3,4,5</sup>, 邹文浩<sup>2,5</sup>, 田霖<sup>2,3,4,5</sup>

<sup>1</sup>(南京信息工程大学 计算机学院, 南京 210044)

<sup>2</sup>(中国科学院 计算技术研究所, 北京 100190)

<sup>3</sup>(中国科学院大学南京学院, 南京 211135)

<sup>4</sup>(中科南京信息高铁研究院, 南京 211100)

<sup>5</sup>(中国科学院大学, 北京 100049)

通信作者: 张宗帅, E-mail: [zhangzongshuai@ict.ac.cn](mailto:zhangzongshuai@ict.ac.cn)

**摘要:** 人脸表情识别在日常生活中得到了越来越广泛的应用. 针对人脸表情识别模型参数复杂、易受背景干扰和高延迟等问题, 本文提出了一种边云协同融合高效通道注意力 (efficient channel attention, ECA) 机制的轻量化表情识别方法. 在云端部署通用模型并利用大规模数据集进行训练, 同时在边缘端迁移学习云端模型的浅层卷积层作为特征提取器, 提升特征提取和泛化能力, 降低过拟合风险. 在此基础上, 引入 ECA 机制, 使模型聚焦于人脸表情特征区域、抑制无用信息, 进一步提升识别准确性与鲁棒性; 此外, 采用深度可分离卷积, 有效减少模型参数量, 同时保证表达能力, 显著降低边缘设备的计算资源消耗. 最终, 在边缘端完成识别任务, 减少数据传输开销并提升响应速度. 实验结果表明, 该方法在 CK+数据集和 FER2013 数据集上的准确率分别达到了 98.76% 和 71.93%. 与传统方法相比, 在保证较高准确率的同时, 显著减少了模型参数量并降低了识别时延, 验证了该方法在边缘端表情识别任务中的准确性与高效性.

**关键词:** 人脸表情识别; 边云协同; ECA 机制; 深度可分离卷积; 迁移学习

引用格式: 李文浩, 张宗帅, 邹文浩, 田霖. 基于边云协同的 ECA 轻量化人脸表情识别. 计算机系统应用, 2026, 35(2): 92-102. <http://www.c-s-a.org.cn/1003-3254/10088.html>

## ECA Lightweight Facial Expression Recognition Based on Edge-cloud Collaboration

LI Wen-Hao<sup>1,3,4</sup>, ZHANG Zong-Shuai<sup>2,3,4,5</sup>, ZOU Wen-Hao<sup>2,5</sup>, TIAN Lin<sup>2,3,4,5</sup>

<sup>1</sup>(School of Computer Science, Nanjing University of Information Science & Technology, Nanjing 210044, China)

<sup>2</sup>(Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China)

<sup>3</sup>(University of Chinese Academy Sciences, Nanjing, Nanjing 211135, China)

<sup>4</sup>(Nanjing Institute of InforSuperBahn, Nanjing 211100, China)

<sup>5</sup>(University of Chinese Academy of Sciences, Beijing 100049, China)

**Abstract:** Facial expression recognition has been increasingly applied in daily life. To address issues such as complex parameters, vulnerability to background interference, and high latency in facial expression recognition models, this study proposes a lightweight expression recognition method based on edge-cloud collaboration and the efficient channel attention (ECA) mechanism. A general model is deployed in the cloud and trained on large-scale datasets. Meanwhile, the shallow convolutional layers of the cloud model are transferred to the edge as feature extractors, which enhance feature extraction, generalization, and reduce the risk of overfitting. On this basis, the ECA mechanism is introduced to enable the model to focus on facial expression feature regions and suppress irrelevant information, further improving recognition accuracy and robustness. Furthermore, using depthwise separable convolution effectively reduces model parameters while

① 基金项目: 江苏省重大科研设施预研筹建项目 (BM2021800); 青岛市关键技术攻关项目 (25-1-1-gjgg-31-gx)

收稿时间: 2025-07-26; 修改时间: 2025-08-19, 2025-09-17; 采用时间: 2025-10-09; csa 在线出版时间: 2025-12-29

CNKI 网络首发时间: 2025-12-30

maintaining expressive power, significantly lowering computational resource consumption on edge devices. Ultimately, the recognition task is performed at the edge, reducing data transmission overhead and improving response speed. Experimental results show that the accuracy of this method on the CK+ and FER2013 datasets reaches 98.76% and 71.93%, respectively. Compared with traditional methods, while maintaining high accuracy, the number of model parameters is significantly reduced and recognition latency is decreased, verifying the accuracy, efficiency, and deployment advantages of this method for facial expression recognition tasks at the edge.

**Key words:** facial expression recognition; edge-cloud collaboration; efficient channel attention (ECA) mechanism; depthwise separable convolution (DSC); transfer learning

人脸表情不仅是情感和意图的重要外在表现,更是反映个体内心状态、情感体验、沟通意图的关键因素. Mehrabia<sup>[1]</sup>的研究表明,人类情感表达中约有55%的信息是通过人脸表情传递,而通过声音与语言的方式进行传递的占比只有38%和7%. Frank等人<sup>[2]</sup>通过对来自不同国家和不同文化背景的学生进行实验和研究,提出了人脸运动编码系统(facial action coding system, FACS), FACS将人类的基本表情分为6种,包括伤心(sad)、害怕(fear)、厌恶(disgust)、开心(happy)、生气(angry)和惊讶(surprise),后续研究在此基础上补充了中性(neutral)状态<sup>[3]</sup>,进一步丰富了面部情绪表达的编码体系.

这些基本表情被认为是人类情感表达的核心要素,为面部表情识别任务提供了标准化、可量化的分析基础,也为智能系统中表情识别模块的设计与优化奠定了理论依据.随着人工智能和计算机视觉技术的快速发展,面部表情识别技术也进入了快速发展的阶段.传统的人脸表情识别算法一般是通过手工方法提取人脸特征,例如光流法<sup>[4]</sup>、主成分分析法<sup>[5]</sup>、Gabor小波变换<sup>[6]</sup>、局部二值法<sup>[7]</sup>等,但手工方法具有低效率和低精度等显著的缺陷.近几年随着卷积神经网络(convolutional neural network, CNN)出现,面部表情识别技术取得了显著进步.卷积神经网络可以自动学习并提取复杂的面部特征,克服了传统手工方法特征提取的缺陷. LeCun等人<sup>[8]</sup>于1998年提出的深度学习经典网络模型LeNet-5,标志着CNN在图像识别领域的实用化.随后GoogLeNet<sup>[9]</sup>、ResNet<sup>[10]</sup>、VGGNet<sup>[11]</sup>、MobileNet<sup>[12]</sup>和ShuffleNet<sup>[13]</sup>等主流的神经网络模型也相继被提出.在卷积神经网络在图像识别领域取得巨大进步的同时,也推动了人脸表情识别领域的快速发展.

刘经纬等人<sup>[14]</sup>提出了一种CNN的方法来解决面

部表情识别问题,并且在表情数据集中取得了不错的准确率.董翠等人<sup>[15]</sup>将类别注意力机制和上下文感知金字塔模块引入到VGG16模型中,通过扩大网络的感受野从而有效地提取人脸表情特征. Nur等人<sup>[16]</sup>通过迁移学习VGG16模型,并经过对比后使用Adam优化器,实现了对驾驶行为的高效特征提取,显著提高了驾驶行为检测的准确率.孙培源等人<sup>[17]</sup>提出了一种结合迁移学习和剪枝的轻量级表情识别方法,在保证识别精度的同时有效减少模型体积,提升了模型部署效率.潘东杰等人<sup>[18]</sup>提出一种使用注意力机制的改进残差网络,通过融合ECA注意力模块强化特征通道表达,配合图像预处理突出表情特征,实现高效的驾驶员异常情绪识别,在减少参数和提升速度的同时显著提高识别精度.罗思诗等人<sup>[19]</sup>提出了一种多尺度融合注意力机制网络(MIANet),通过引入Inception结构实现多尺度特征提取,利用ECA注意力机制使得聚焦面部关键区域,并采用深度可分离卷积显著降低了模型的参数.

上述方法虽然在提升人脸表情识别性能方面取得了一定进展,但在实际的互联网场景中会出现数据量激增、实时性差以及用户隐私等问题.边缘计算靠近用户,具有低时延、隐私保护的优势,但是存在算力不足等缺点.边云协同利用边缘端和云端的优势,有效解决了以上问题.张娴静等人<sup>[20]</sup>提出了一种边缘云框架下结合双模特征的人脸表情识别方法,通过使用边缘端对数据进行采集和处理然后发送给云端进行识别,从而降低了延迟并保护了用户的隐私,但边缘端只负责处理数据并没有充分利用边缘端的算力.况朝青等人<sup>[21]</sup>在SqueezeNet网络的基础上进行改进,设计了轻量级的模型并部署在边缘端,解决了识别时间长、功耗高等问题.但由于训练数据规模有限,其在面对复杂多变的表情样本时泛化能力较弱,鲁棒性有待提升.

为了进一步提高人脸表情识别模型的精度与效率,本文提出一种采用边云协同架构的ECA轻量化人脸表情识别方法.该方法利用云端和边缘端的优势,同时引入ECA注意力机制和深度可分离卷积,从而在保证识别准确率的同时有效降低模型的参数量.本文工作概括如下.

(1) 利用云端充足的存储与计算资源部署通用模型,并使用大规模通用数据集进行训练.

(2) 采用迁移学习方法,将云端预训练模型的浅层特征提取部分迁移至边缘端进行共享.边缘端模型在此基础上仅对后续层进行微调,以适应本地数据的分布差异.同时,在边缘模型中添加ECA注意力机制以增强对关键区域特征的聚焦能力,并使用深度可分离卷积结构来降低模型参数数量,以获得低延迟、高精度的人脸表情识别效果.

(3) 在CK+和FER2013数据集上对本文方法进行实验验证,并与其他方法进行对比,结果表明本文方法在保证轻量化的同时,显著提升识别精度.

## 1 边云协同架构与表情识别模型

### 1.1 边云协同架构

边云协同架构如图1所示.该框架充分利用云端强大的算力和存储能力,以及边缘端低延时的优势,满足识别场景下对实时性和准确性的需求.整体架构主要由云服务器和边缘服务器两层组件组成.云端利用大规模数据集训练通用模型,边缘端则通过迁移学习<sup>[22]</sup>使用云端训练好的浅层模型,并结合本地特征进行微调,构建适用于本地任务的表情识别模型.

迁移学习旨在将从相关源域获得的知识应用于目标域,以提升新任务的学习效率和性能.相比传统从零开始训练模型的方法,该策略显著减少了对大量本地数据的依赖,加快了模型收敛速度,降低了训练时间和计算资源消耗,特别适合资源受限的边缘端设备,且能提升模型的鲁棒性和识别精度.

### 1.2 表情识别模型

#### 1.2.1 VGG16网络模型

云服务器拥有丰富的计算和存储资源,因此,在云端部署通用型模型,并使用大规模的ImageNet数据集进行训练,这里使用VGG16模型<sup>[23]</sup>.

VGG16由牛津大学的研究团队Visual Geometry Group提出,由16个卷积层和3个全连接层组成,网络

结构如图2所示.网络通过多个小尺寸卷积核提高模型的表达能力,并在卷积层之间插入多个最大池化层进行下采样,以减小特征图的尺寸并保留最显著的特征,然后连接3个具有4096个隐藏单元的全连接层,最后由全连接层输出模型的预测结果.所有卷积层中卷积核的大小为 $3\times 3$ 、步长为1,使用ReLU激活函数,最大池化层大小为 $2\times 2$ 、步长为2.网络在捕捉复杂特征的同时,又保持了网络的简洁性.VGG16的模型结构相对简单且易于理解,被广泛应用于计算机视觉领域的各种任务,如图像分类、目标检测等.因此,在云端服务器部署VGG16模型.

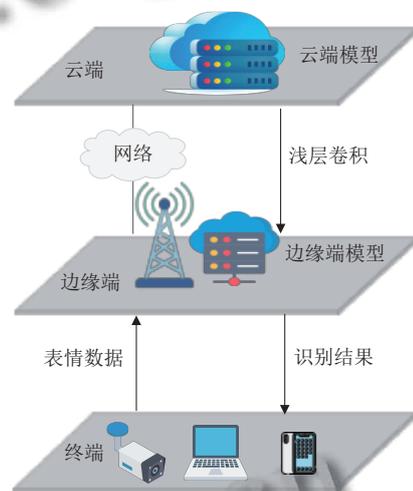


图1 边云协同架构

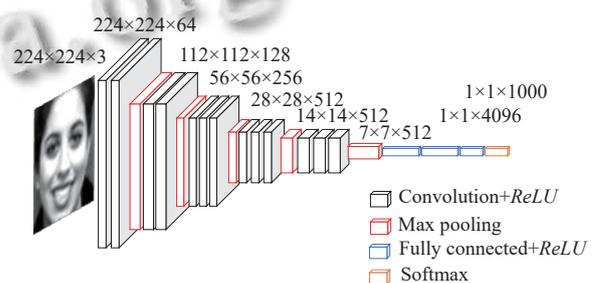


图2 VGG16网络结构

#### 1.2.2 ECA注意力机制

在表情识别中,背景冗余会干扰模型对关键表情特征的提取.为增强模型对表情区域的关注,本文在迁移学习的基础上引入了ECA-Net<sup>[24]</sup>通道注意力机制.ECA-Net通过全局平均池化和自适应一维卷积生成通道权重,从而突出关键区域并抑制冗余.其结构如图3所示.

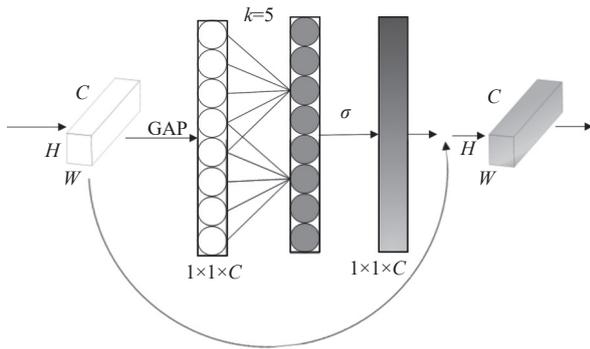


图3 ECA-Net结构

设输入特征图  $F \in R^{H \times W \times C}$ , 其中  $H$  和  $W$  分别表示特征图的高和宽,  $C$  表示通道数. ECA-Net 先对每个通道进行全局平均池化, 得到长度为  $C$  的通道描述向量, 随后在该向量上施加自适应一维卷积以建模通道间的局部依赖, 最后通过 Sigmoid 激活生成通道权重, 并与原始特征逐通道相乘, 实现注意力机制. 卷积核大小  $k$  由式 (1) 决定:

$$k = \left\lfloor \frac{\log_2(C) + b}{\gamma} \right\rfloor_{\text{odd}} \quad (1)$$

其中,  $C$  表示通道数;  $\lfloor t \rfloor_{\text{odd}}$  表示距离  $t$  最近的奇数,  $\gamma$  为缩放参数,  $b$  是偏移量. 式 (1) 仅依赖通道数来确定卷积核大小, 未考虑特征图空间分辨率和任务难度等因素. 本文在原始公式的基础上进行了优化, 提出自适应核大小计算方式:

$$k = \max \left( 3, \text{odd} \left( \frac{\log_2(C) + b}{\gamma} + \alpha \cdot \frac{\sqrt{H \cdot W}}{32} + \beta \cdot \tau \right) \right) \quad (2)$$

其中,  $\alpha \cdot \frac{\sqrt{H \cdot W}}{32}$  为空间分辨率对感受野范围的影响,  $\beta \cdot \tau$  根据任务复杂度动态调节卷积核大小. 同时,  $\text{odd}(\cdot)$  保证卷积核为奇数以维持卷积的对称性,  $\max(3, \cdot)$  则用于避免因卷积核过小而失效. 通过引入空间和任务因素, 改进后的  $k$  更具适应性, 可在不同图像尺寸和表情识别场景下更有效地提取关键特征, 提升模型性能与鲁棒性.

### 1.2.3 深度可分离卷积

为降低边缘端模型的计算量和参数量, 本文采用深度可分离卷积 (depthwise separable convolution, DSC)<sup>[25]</sup> 网络替代标准卷积, 从而实现模型的轻量化.

如图 4 所示, 深度可分离卷积分为两部分: 首先, 逐通道卷积 (depthwise convolution, DW) 使用一个卷积核仅作用于输入特征图的单一通道, 独立提取空间

特征; 随后, 逐点卷积 (pointwise convolution, PW) 采用  $1 \times 1$  卷积核对各通道进行线性组合, 融合通道信息生成输出特征图. 与标准卷积相比, DSC 显著减少了参数量和计算开销. 假设输入通道数为  $M$ 、输出通道数为  $N$ , 卷积核尺寸为  $k \times k$ , 输出特征空间尺寸为  $D_F \times D_F$ . 标准卷积的参数量和计算量为  $P_{\text{std}} = k^2 \times M \times N$  和  $C_{\text{std}} = k^2 \times N \times M \times D_F^2$ , 深度卷积的参数量和计算量为  $P_{\text{dw}} = k^2 \times M$  和  $C_{\text{dw}} = k^2 \times M \times D_F^2$ , 逐点卷积的参数量和计算量为  $P_{\text{pw}} = M \times N$  和  $C_{\text{pw}} = M \times N \times D_F^2$ , 深度可分离卷积与标准卷积的参数量和计算量之比为:  $\frac{1}{N} + \frac{1}{k^2}$ .

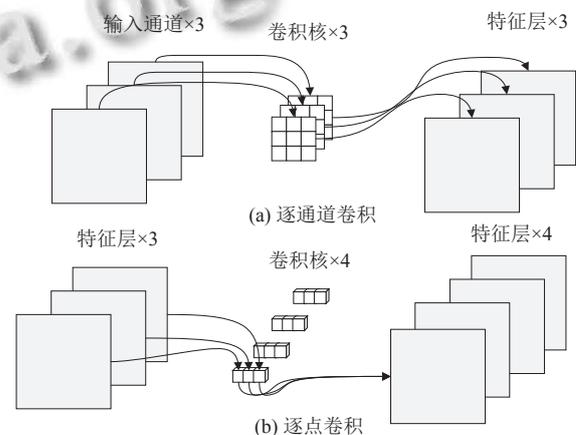


图4 深度可分离卷积结构图

深度可分离卷积使用的是  $ReLU$  激活函数, 其定义为:

$$f(x) = \max(0, x) = \begin{cases} x, & x > 0 \\ 0, & \text{其他} \end{cases} \quad (3)$$

$ReLU$  函数具有计算简单、收敛速度快等优点, 但在输入值接近或小于 0 时, 可能引发梯度消失问题, 影响模型的训练效果. 为此, 本文使用  $H$ -swish 激活函数, 其表达式为:

$$H\text{-swish}(x) = x \frac{ReLU6(x+3)}{6} \quad (4)$$

$H$ -swish 在负值区间保留部分梯度信息, 缓解了梯度消失问题, 并具备较强的非线性表达能力. 同时, 其运算形式简单高效, 比较适合边缘计算场景下的模型部署与加速.

## 2 边云协同 ECA 轻量化表情识别方法

本文提出了一种边云协同 ECA 轻量化人脸表情识别方法, 兼具轻量化、高精度和低时延特性, 整体流

程如图5所示. 首先, 云端利用大规模通用数据集训练特征提取模型; 随后, 边缘端模型迁移学习云端模型浅层卷积参数, 并结合本地表情数据进行微调; 最终, 边

缘端模型独立完成表情特征提取与分类. 整个过程中, 仅同步浅层模型参数, 用户表情图像始终在本地处理, 保障数据隐私.

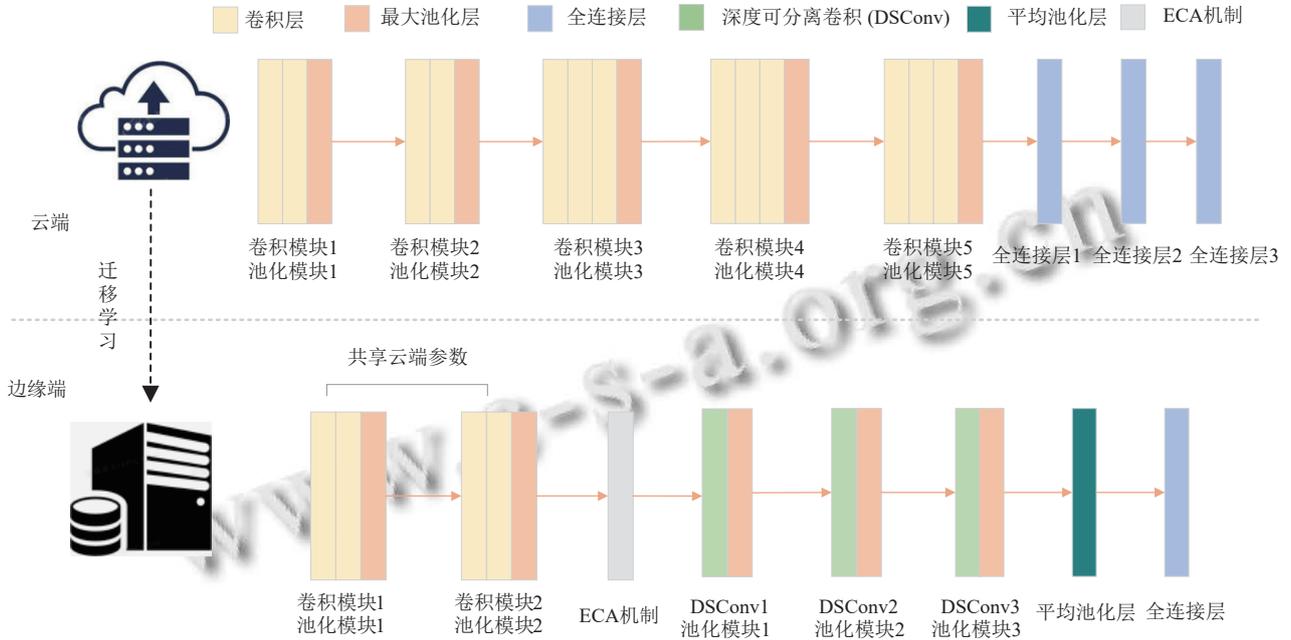


图5 边云协同表情识别流程

云端拥有足够的计算和存储资源, 使用大规模的通用数据集进行训练, 来获得强泛化能力的通用模型. 设数据集为 $\{(a_i, b_i)\}_{i=1}^U$ , 其中 $a_i \in \mathbb{R}^D$ 表示输入样本,  $b_i \in \mathbb{R}^K$ 为对应的类别标签,  $D$ 是输入样本维度,  $K$ 是类别数量,  $U$ 是数据集的数量. 在云端服务器上部署卷积神经网络模型, 学习参数 $\Theta_c$ , 优化目标为:

$$g_c(\Theta_c) = \arg \min_{\Theta_c} \frac{1}{U} \sum_{i=1}^U \mathcal{L}(b_i, s(g(a_i; \Theta_c))) \quad (5)$$

其中,  $g(a_i; \Theta_c)$ 表示云端模型对 $a_i$ 的输出,  $s(\cdot)$ 表示 Softmax 操作:

$$s(z)_k = \frac{e^{z_k}}{\sum_{j=1}^K e^{z_j}}, \quad k = 1, \dots, K \quad (6)$$

$\mathcal{L}$ 为 Focal loss 损失函数:

$$\mathcal{L}(b_i, \hat{b}_i) = - \sum_{k=1}^K \alpha_k (1 - \hat{b}_{i,k})^\gamma b_{i,k} \log(\hat{b}_{i,k}) \quad (7)$$

其中,  $b_i$ 为 $a_i$ 的真实标签,  $\hat{b}_i$ 为 $a_i$ 预测标签,  $b_{i,k} \in \{0, 1\}$ ,  $\hat{b}_{i,k}$ 表示预测为第  $k$  类的概率,  $\alpha_k$ 为类别平衡因子,  $\gamma$ 为

调制因子.

边缘端模型迁移学习并冻结云端模型的浅层卷积, 仅微调剩余参数 $\Theta_e^*$ 以适应边缘设备上的特定数据分布. 边缘端同样使用 Focal loss 损失函数进行训练, 优化目标为:

$$g_e(\Theta_e^*) = \arg \min_{\Theta_e^*} \frac{1}{V} \sum_{j=1}^V \mathcal{L}(b'_j, s(g(a'_j; \Theta_e^*))) \quad (8)$$

其中,  $\{(a'_j, b'_j)\}_{j=1}^V$ 为边缘端数据集,  $a'_j \in \mathbb{R}^L$ 表示输入样本,  $b'_j \in \mathbb{R}^N$ 为对应的类别标签,  $L$ 是输入样本维度,  $N$ 是类别数量,  $V$ 是边缘端样本数量.  $g(a'_j; \Theta_e^*)$ 表示边缘端模型对 $a'_j$ 的输出,  $s(\cdot)$ 和 $\mathcal{L}(\cdot, \cdot)$ 计算如下:

$$s(z)_n = \frac{e^{z_n}}{\sum_{m=1}^N e^{z_m}}, \quad n = 1, \dots, N \quad (9)$$

$$\mathcal{L}(b'_j, \hat{b}'_j) = - \sum_{n=1}^N \alpha_n (1 - \hat{b}'_{j,n})^\gamma b'_{j,n} \log(\hat{b}'_{j,n}) \quad (10)$$

其中,  $b'_j$ 是 $a'_j$ 的真实标签,  $\hat{b}'_j$ 是 $a'_j$ 的预测标签,  $b'_{j,n} \in \{0, 1\}$ ,  $\hat{b}'_{j,n}$ 表示预测为第  $n$  类的概率,  $\alpha_n$ 为类别平衡因

子,  $\gamma$  为调制因子。

考虑到人脸表情图像中的关键信息通常局部分布, 为增强模型对局部关键区域的感知能力, 本文在边缘模型中引入 ECA 注意力机制增强网络对关键区域的关注, 提升网络的识别性能。此外, 在模型中使用深度可分离卷积替代标准卷积, 将参数量和计算量大幅压缩的同时保留丰富的特征表达能力, 有效降低了边缘设备资源消耗并加速推理, 边缘协同训练表情识别模型的伪代码见算法 1 和算法 2。

算法 1. 训练云端模型

输入: 数据集  $\{(a_i, b_i)\}_{i=1}^u$ , 云端模型。

输出: 前 4 层浅层卷积层参数  $\Theta_c^4$ 。

```

1. for epoch=1 to epochs do
2.   for each batch  $(a_i, b_i)$  in  $\{(a_i, b_i)\}_{i=1}^u$  do
3.      $\hat{b} \leftarrow s(g(a; \Theta_c))$ 
4.      $loss \leftarrow \mathcal{L}(b, \hat{b})$ 
5.      $\Theta_c \leftarrow \Theta_c - lr \cdot \nabla_{\Theta_c} loss$ 
6.   end for
7.   在验证集上评估准确率, 若无提升则降低  $lr$ 
8. end for
9. 提取前 4 层浅层卷积参数  $\Theta_c^4 \leftarrow \Theta_c[1 \dots 4]$ 
10. 返回  $\Theta_c^4$ 

```

算法 2 训练边缘端模型

输入: 参数  $\Theta_c^4$ , 边缘端模型, 表情数据集  $\{(a'_j, b'_j)\}_{j=1}^v$ 。

输出: 边缘端网络模型。

```

1. for epoch=1 to epochs do
2.   for each batch  $(a', b')$  in  $\{(a'_j, b'_j)\}_{j=1}^v$  do
3.      $\hat{b}' \leftarrow s(g(a'; \Theta_c^4))$ 
4.      $loss \leftarrow \mathcal{L}(b', \hat{b}')$ 
5.      $\Theta_e^* \leftarrow \Theta_e^* - lr \cdot \nabla_{\Theta_e} loss$ 
6.   end for
7.   在验证集上评估准确率, 若无提升则降低  $lr$ 
8. end for
9. 边缘端模型参数:  $\Theta_e = \Theta_c^4 \cup \Theta_e^*$ 
10. 返回边缘端网络模型 (含参数  $\Theta_e$ )

```

边缘端表情识别模型结构及详细参数信息如表 1 所示。本模型共包含 7 个卷积层, 前 4 个卷积层 (层 1、2、4、5) 迁移自云端模型, 卷积核大小均为  $3 \times 3$ , 步长为 1, 每个卷积层均连接批归一化和  $ReLU$  激活函数。在层 3、6、9、11、13 中加入窗口大小为 2、步长为 2 的最大池化层进行下采样。为了提升通道特征表达能力, 在层 7 引入了 ECA 注意力模块。随后在层 8、10、12 采用深度可分离卷积, 其中深度卷积核大小为

$3 \times 3$ , 点卷积核大小为  $1 \times 1$ , 步长均为 1, 每个深度可分离卷积均连接批归一化和  $H-swish$  激活函数。最后通过层 14 的平均池化层获得最终特征表示。

表 1 边缘端模型结构及参数

层	输入类型	卷积核	步长	输出尺寸
1	Convolution	$3 \times 3$	1	$44 \times 44 \times 64$
2	Convolution	$3 \times 3$	1	$44 \times 44 \times 64$
3	Maxpool	$2 \times 2$	2	$22 \times 22 \times 64$
4	Convolution	$3 \times 3$	1	$22 \times 22 \times 128$
5	Convolution	$3 \times 3$	1	$22 \times 22 \times 128$
6	Maxpool	$2 \times 2$	2	$11 \times 11 \times 128$
7	ECA-NET	$k$ 动态	—	$11 \times 11 \times 128$
8	DepthwiseSeparableConv1	DW: $3 \times 3$ , PW: $1 \times 1$	1	$11 \times 11 \times 256$
9	MaxPool	$2 \times 2$	2	$5 \times 5 \times 256$
10	DepthwiseSeparableConv2	DW: $3 \times 3$ , PW: $1 \times 1$	1	$5 \times 5 \times 512$
11	MaxPool	$2 \times 2$	2	$2 \times 2 \times 512$
12	DepthwiseSeparableConv3	DW: $3 \times 3$ , PW: $1 \times 1$	1	$2 \times 2 \times 512$
13	MaxPool	$2 \times 2$	2	$1 \times 1 \times 512$
14	AvgPool	$1 \times 1$	1	$1 \times 1 \times 512$

## 3 实验

### 3.1 数据集

本文采用 Extended Cohn-Kanada (CK+) 数据集和 FER2013 数据集进行实验评估, 实验采用的样本数量分布如表 2 所示。

表 2 CK+, FER2013 数据集实验样本选取数量分布表

数据集	生气	自然	厌恶	害怕	开心	悲伤	惊讶	轻蔑
CK+	135	—	177	75	207	84	249	54
FER2013	4953	6198	547	5121	8989	6077	4002	—

CK+数据集是目前比较通用的人脸表情数据集, 适合于人脸表情识别任务, 包含 123 名参与者, 年龄为 18–50 岁不等, 有不同的性别和血统。共有 593 个  $48 \times 48$  的图像序列, 每个序列记录了从中性表情到峰值表情的动态转变过程, 其中仅有 327 个序列附有情感标签。表情类别包括生气 (anger)、轻蔑 (contempt)、厌恶 (disgust)、害怕 (fear)、开心 (happy)、悲伤 (sadness) 和惊讶 (surprise) 这 7 种基本表情, 如图 6 所示。

FER2013 数据集是面部表情识别领域的一个标准基准, 由 Kaggle 人脸表情识别挑战赛提供。该数据集包含 35 887 张  $48 \times 48$  像素的灰度人脸图像, 每张图像被标注为 7 种基本情感之一: 生气、厌恶、害怕、开心、悲伤、惊讶和自然。数据集由 28 709 张训练集、3 589 张公共测试集和 3 589 张私有测试集组成。但 FER2013 数据集中各情感类别的样本数量不均衡, 例如, 厌恶

(disgust) 类别的图像数量最少, 仅约 600 张, 而其他类别的样本数量接近或超过 5 000 张, 如图 7 所示. 总体而言, FER2013 数据集以其规模适中、标注详细、表情类别丰富等特点, 成为面部表情识别任务中的重要资源.



图 6 CK+数据集



图 7 FER2013 数据集

### 3.2 实验设置

本实验在 Windows 10 平台上, 使用 Python 3.8 和 PyTorch 框架完成. 云端采用阿里云服务器, 配备 2 张 Intel Xeon Platinum 8260 处理器 (每张 24 核, 主频 2.4 GHz, 最高 3.8 GHz), 共计 48 核, 256 GB 内存, 两块 1.5 TB NVMe SSD 及 2 块 NVIDIA Tesla V100 (32 GB) GPU 用于模型训练与推理加速; 边缘端使用一台搭载 Intel Core i5-4590@3.30 GHz 处理器和 12 GB 内存的计算机, 模拟资源受限设备的部署环境. 在网络设置方面, 5G 环境下平均下行速率为 2.72 MB/s, 上行速率为 1.93 MB/s; WiFi 环境下上下行速率均为 13.5 MB/s, 用于模拟边云通信过程并评估所提方法在不同网络条件下的响应性能.

### 3.3 预处理

为提升模型在光照变化和噪声干扰下的稳健性与泛化能力, 采用 CLAHE 与 Gamma 校正的组合增强策略. CLAHE 能有效增强图像的局部对比度, 使面部表情区域更加清晰, 但可能带来整体亮度偏移; 而 Gamma 校正则能够平衡图像的亮度分布, 使暗部与亮部信息更为均衡, 但对局部细节的提升有限. 将二者结合使用, 既增强局部对比度, 又优化整体亮度分布, 从而在复杂

环境下显著提升表情特征的可辨识性. 如图 8 所示, 经过 CLAHE 与 Gamma 组合增强处理后的图像在细节表现与亮度对比方面均得到改善, 关键表情区域更加突出. 图 9 显示了增强前后的灰度直方图变化, 直观反映了亮度分布的优化效果. 训练阶段采用随机裁剪至 44×44 尺寸和水平翻转进行数据增强, 以增加样本多样性并提升模型对遮挡、表情变化和视角差异的适应性; 测试阶段使用 TenCrop 生成 10 个裁剪样本, 并对预测结果取平均, 以充分利用局部信息并增强泛化能力.



图 8 CLAHE+Gamma 组合增强图像

### 3.4 实验结果与分析

本实验在 CK+数据集上训练时, 为了降低模型过拟合风险, 仅选取每个序列的最后一帧图像作为有效样本进行训练与测试. 模型的初始学习率设置为 0.01, 批量大小为 128, 训练共进行 60 个训练周期 (epoch), 从第 20 个 epoch 开始, 每隔 8 个 epoch 将学习率衰减为原来的 0.8 倍. 采用 10 折交叉验证, 将数据集划分为 10 个子集, 依次轮流选取 1 个子集作为测试集, 其余 9 个子集用于训练. 最终模型在测试集上的平均准确率达到 98.76%. 在每一折的测试阶段, 引入 TenCrop 策略, 对每张测试图像生成 10 个裁剪区域分别进行预测, 最后对 10 个预测结果取平均值, 以提升模型在多样输入下的鲁棒性与准确性. CK+数据集上的混淆矩阵如图 10 所示.

从图 10 中可以看出, 开心、惊讶等表情的识别准确率较高, 这是由于这些情绪通常伴随明显的面部肌肉变化, 模型能够较为容易地捕捉到关键特征. 相比之下, 生气、厌恶、轻蔑等表情的识别率较低, 主要是因为这些表情间的特征相似、容易混淆, 导致模型在区分时存在困难. 此外, 训练样本数量有限以及不同类别样本量不平衡, 也在一定程度上影响了模型在某些类别上的泛化能力.

在 FER2013 数据集上的训练过程中, 模型的 epoch 设置为 250, 初始学习率同样为 0.01, 批量大小为 128. 从第 80 个 epoch 开始, 每 5 个 epoch 将学习率按 0.9 倍进行衰减. 测试阶段同样采用 TenCrop 策略进行评估. 最终在 PrivateTest 测试集上的准确率达到 71.93%, FER2013 数据集的混淆矩阵如图 11 所示.

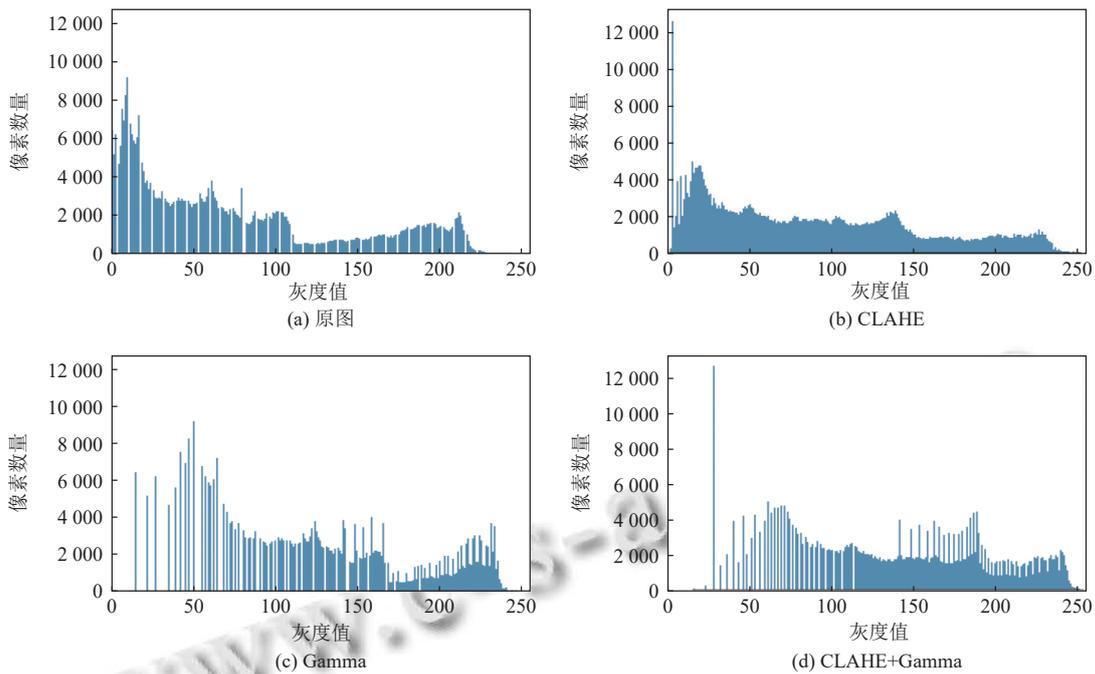


图9 灰度直方图

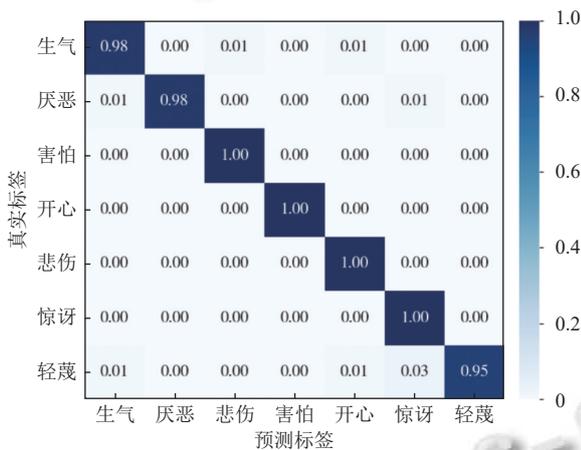


图10 CK+数据集混淆矩阵

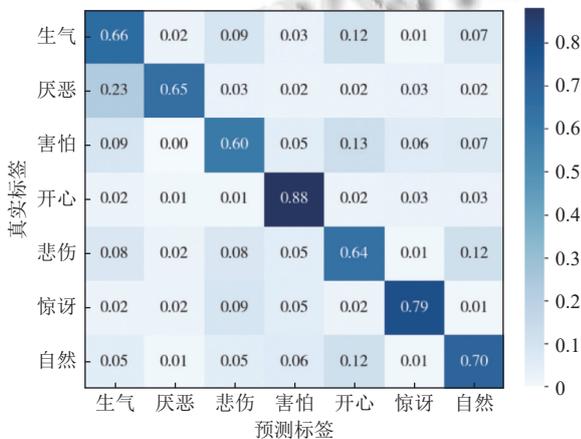


图11 FER2013数据集混淆矩阵

为了验证本文方法的有效性,在CK+和FER2013数据集上,将本方法与近几年国内外表情识别的经典算法进行对比,如图12和图13所示.从图12可见,在CK+数据集上,本方法达到98.76%的识别准确率,比较最近的方法SIFT+CNN(95.40%)提升了3.36个百分点;在FER2013数据集(图13)上,本方法取得71.93%的准确率,相比方法Deep-emotion(70.02%)提升了1.91个百分点.

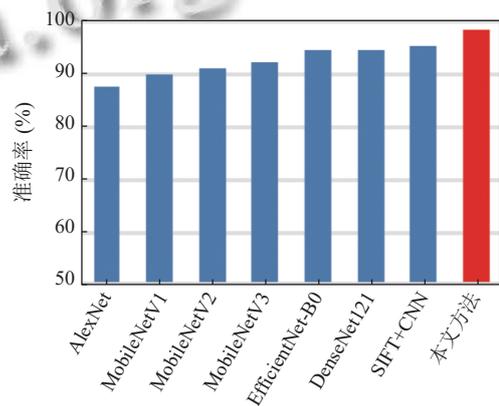


图12 不同算法在CK+数据集上的准确率

针对类别不平衡对少数类识别性能的影响,本文在训练过程中分别采用了传统交叉熵(cross-entropy, CE)损失函数和Focal loss损失函数进行对比实验.与CE相比,Focal loss在其基础上引入平衡因子和调制因

子,能够降低易分类样本的损失权重,从而提升难分类及少数类样本的贡献度,缓解类别不平衡问题.实验结果如表3所示,Focal loss在FER2013和CK+数据集上均有提升.

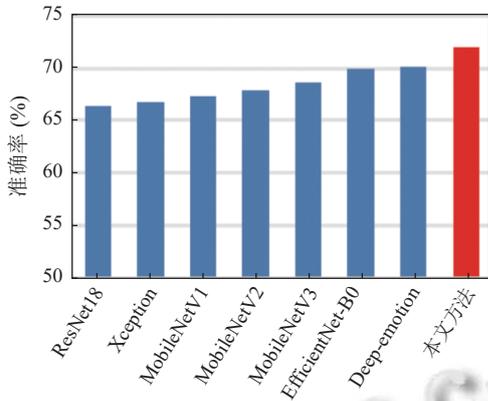


图 13 不同算法在 FER2013 数据集上的准确率

表 3 不同损失函数的准确率对比 (%)

数据集	损失函数	准确率
FER2013	CE	71.82
	Focal loss	71.93
CK+	CE	98.62
	Focal loss	98.76

为进一步验证本文方法在保证较高准确率的同时具备轻量化优势,选取图12和图13中部分典型的轻量级网络结构作为对比对象,并将其参数量与本文方法进行对比,结果如表4所示.由表4中数据可见,本文方法在FLOPs和参数量上均优于EfficientNet-B0与MobileNet系列模型,说明了本文方法在轻量化与性能之间的平衡优势.

表 4 模型复杂度参数量对比

方法	参数量 (M)	FLOPs
EfficientNet-B0 <sup>[26]</sup>	15.48	420
MobileNetV1 <sup>[27]</sup>	3.19	572.77
MobileNetV2 <sup>[28]</sup>	3.50	330
MobileNetV3 <sup>[29]</sup>	2.54	227.95
本文方法	0.703	192.043

为验证云边协同训练策略的有效性,本文在相同网络结构下开展了两组实验:一组采用迁移学习,另一组则在边缘数据集上从头训练.如表5所示,迁移学习显著提升了模型准确率:在CK+上从95.12%提升至98.76%,FER2013上由69.32%提升至71.93%.该结果说明,迁移学习有助于提升模型性能,验证了云边协同策略的必要性.

表 5 不同训练方式下模型准确率对比 (%)

数据集	训练方式	准确率
CK+	无迁移学习	95.12
	迁移学习	98.76
FER2013	无迁移学习	69.32
	迁移学习	71.93

为验证所提方法对时延的优化效果,我们将边云协同训练得到的表情识别模型分别部署到云端和边缘端,将4KB的表情图像经由5G和WiFi网络上传至云端和边缘端并接收结果的总时延.端到端总时延可表示为:

$$T = T_u + T_l + T_p + T_d \quad (11)$$

其中,  $T_u$  为上传时延,  $T_l$  为网络传输时延,  $T_p$  为模型推理时延,  $T_d$  为结果返回时延.实验结果如表6所示.在5G环境下,采用边缘端识别相比云端识别时延减少约38%;在WiFi环境下,时延减少约31%.

表 6 不同部署位置下的时延对比 (ms)

部署位置	5G场景时延	WiFi场景时延
云端	123	61
边缘端	76	42

### 3.5 消融实验

为验证所选迁移层数设置的合理性,本文使用VGG16模型在FER2013与CK+数据集上进行了不同迁移深度的对比实验.我们以VGG16的卷积块(conv block)为单位进行迁移:分别迁移了前1个Block(含2个卷积层)、前2个Block(含4个卷积层)、前3个Block(含7个卷积层)、前4个Block(含10个卷积层)以及全部5个Block(含13个卷积层).

实验结果如表7所示,随着迁移层数的增加,模型准确率先上升后略有下降,其中迁移前2个Block时准确率最高.进一步增加迁移深度不仅显著增加了模型参数量,还引入了与目标任务不匹配的高层特征,从而导致泛化性能下降.

表 7 不同迁移层数对模型性能的影响

迁移层数	CK+准确率 (%)	FER2013准确率 (%)	参数量 (M)
前1个Block	93.28	64.74	0.464
前2个Block	98.76	71.93	0.703
前3个Block	96.90	71.22	2.214
前4个Block	36.59	28.20	8.451
前5个Block	26.36	28.78	15.527

为了进一步验证边缘端模型各模块的有效性,本文开展了消融实验,如表8所示.详细步骤为:(1)完整

方法 (Ours): 边缘端模型使用云端预训练模型的前7层卷积层作为特征提取网络, 在此基础上引入ECA注意力机制以提升对关键特征的感知能力, 引入深度可分离卷积以减少模型参数量, 最终通过全局平均池化和全连接分类层完成分类任务; (2) Ours-ECA: 在完整方法基础上, 去除ECA注意力机制, 来验证ECA模块对模型识别精度的提升作用; (3) Ours-Separable: 在完整方法基础上, 去除深度可分离卷积模块, 验证Separable模块对模型参数量与效率优化的作用; (4) Ours-Separable-ECA: 在完整方法基础上同时去除深度可分离卷积和ECA注意力机制, 仅迁移云端预训练模型前7层特征提取网络, 并通过全局平均池化与分类层进行分类, 作为不引入任何结构优化模块的基础模型。

表8 消融实验

方法	参数量	FLOPs (M)	识别率 (%)	
			FER2013	CK+
Ours	702 919	192.058	71.93	98.76
Ours-ECA	702 904	192.058	70.62	95.39
Ours-Separable	261 831	183.286	68.46	92.38
Ours-Separable-ECA	261 816	183.270	64.26	90.34

#### 4 结束语

本文针对表情识别过程中模型参数复杂、容易受背景影响和高延迟等问题, 提出了一种边云协同ECA轻量化人脸表情识别的方法。首先, 在云端使用大规模数据集训练复杂模型, 然后通过迁移学习提升边缘端模型的特征提取能力; 在此基础上, 边缘端添加高效通道注意力机制使得模型更容易学习到有用信息; 最后, 使用深度可分离卷积替换标准卷积, 降低模型参数量, 得到适合在边缘端部署的轻量化表情识别模型。在CK+和FER2013人脸表情数据集上对边缘端模型进行评估, 验证了本文所提方法的有效性。在未来的工作中, 将考虑在更加真实复杂的场景下应用。在多种不同的边缘设备上部署模型, 并进行针对性的优化, 如量化和剪枝等来提升模型的效率。此外, 将面部表情识别方法应用在自动驾驶等不同领域的实际场景中。

#### 参考文献

- Mehrabian A. Silent Messages. Belmont: Wadsworth, 1971.
- Frank MG, Ekman P, Friesen WV. Behavioral markers and recognizability of the smile of enjoyment. *Journal of Personality and Social Psychology*, 1993, 64(1): 83–93. [doi: 10.1037/0022-3514.64.1.83]

- Cheng F, Yu JS, Xiong HL. Facial expression recognition in JAFFE dataset based on Gaussian process classification. *IEEE Transactions on Neural Networks*, 2010, 21(10): 1685–1690. [doi: 10.1109/TNN.2010.2064176]
- 裴巧娜. 基于光流法的运动目标检测与跟踪技术 [硕士学位论文]. 北京: 北方工业大学, 2009.
- Luo Y, Wu CM, Zhang Y. Facial expression recognition based on fusion feature of PCA and LBP with SVM. *Optik—International Journal for Light and Electron Optics*, 2013, 124(17): 2767–2770. [doi: 10.1016/j.ijleo.2012.08.040]
- Gu WF, Xiang C, Venkatesh YV, et al. Facial expression recognition using radial encoding of local Gabor features and classifier synthesis. *Pattern Recognition*, 2012, 45(1): 80–91. [doi: 10.1016/j.patcog.2011.05.006]
- Jia Q, Gao XK, Guo H, et al. Multi-layer sparse representation for weighted LBP-patches based facial expression recognition. *Sensors*, 2015, 15(3): 6719–6739. [doi: 10.3390/s150306719]
- LeCun Y, Boser B, Denker JS, et al. Handwritten digit recognition with a back-propagation network. *Proceedings of the 3rd International Conference on Neural Information Processing Systems*. Denver: MIT Press, 1989. 396–404.
- Szegedy C, Liu W, Jia YQ, et al. Going deeper with convolutions. *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition*. Boston: IEEE, 2015. 1–9.
- Targ S, Almeida D, Lyman K. ResNet in ResNet: Generalizing residual architectures. arXiv:1603.08029, 2016.
- Wang LM, Guo S, Huang WL, et al. Places205-VGGNet models for scene recognition. arXiv:1508.01667, 2015.
- Wang W, Li YT, Zou T, et al. A novel image classification approach via Dense-MobileNet models. *Mobile Information Systems*, 2020, 2020: 7602384.
- Zhang XY, Zhou XY, Lin MX, et al. ShuffleNet: An extremely efficient convolutional neural network for mobile devices. *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 6848–6856.
- 刘经纬, 许少伟. 基于深度学习的图像中人脸检测与表情识别. *信息系统工程*, 2024(7): 70–73.
- 董翠, 罗晓曙, 黄苑琴. 一种基于改进VGG16网络的人脸表情识别算法. *计算机与数字工程*, 2024, 52(1): 259–265.
- Nur AH, Talukder MSH, Adnan S, et al. A transfer learning approach with modified VGG16 for driving behavior detection in intelligent transportation systems. *Proceedings of*

- the 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT). Dhaka: IEEE, 2024. 1060–1065.
- 17 孙培源, 袁甲, 孙玉宝. 基于门控剪枝的轻量级表情识别方法. 中国电子科学研究院学报, 2025, 20(1): 75–82.
- 18 潘东杰, 秦晓, 崔玉霞, 等. 基于改进残差网络的驾驶员异常情绪识别方法. 计算技术与自动化, 2025, 44(2): 121–127. [doi: [10.16339/j.cnki.jsjsyzdh.202502022](https://doi.org/10.16339/j.cnki.jsjsyzdh.202502022)]
- 19 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): 199–206.
- 20 张娴静, 褚含冰, 刘鑫. 基于边缘云框架的高效安全人脸表情识别. 计算机工程与设计, 2021, 42(5): 1472–1478. [doi: [10.16208/j.issn1000-7024.2021.05.037](https://doi.org/10.16208/j.issn1000-7024.2021.05.037)]
- 21 况朝青, 贺超, 王均成, 等. 基于边缘计算中极端姿态和表情的人脸识别. 电子技术应用, 2021, 47(6): 30–34. [doi: [10.16157/j.issn.0258-7998.200968](https://doi.org/10.16157/j.issn.0258-7998.200968)]
- 22 吕军, 苒婉婷, 陈付龙, 等. 优化双线性 ResNet34 的人脸表情识别. 计算机系统应用, 2024, 33(11): 27–37. [doi: [10.15888/j.cnki.csa.009682](https://doi.org/10.15888/j.cnki.csa.009682)]
- 23 刘芳, 李俊吉. 融合 VGG 与注意力的学生微表情识别和情绪评估方法. 现代计算机, 2024, 30(18): 28–33.
- 24 王瑞绅, 宋公飞, 王明. 引入 ECA 注意力机制的 U-Net 语义分割. 电光与控制, 2023, 30(1): 92–96, 102.
- 25 于成成, 郭芝源. 基于轻量级神经网络的人脸表情识别研究. 物联网技术, 2024, 14(8): 49–52. [doi: [10.16667/j.issn.2095-1302.2024.08.011](https://doi.org/10.16667/j.issn.2095-1302.2024.08.011)]
- 26 Hoang VT, Jo KH. Practical analysis on architecture of EfficientNet. Proceedings of the 14th International Conference on Human System Interaction (HSI). Gdańsk: IEEE, 2021. 1–4.
- 27 Chen TT, Lin GJ, Zhang HJ, *et al.* Non-uniform illumination face recognition method based on improved MobileNetV1. Journal of Computers, 2023, 34(1): 145–155. [doi: [10.53106/199115992023023401011](https://doi.org/10.53106/199115992023023401011)]
- 28 严春满, 张翔, 王青朋. 基于改进 MobileNetV2 的人脸表情识别. 计算机工程与科学, 2023, 45(6): 1071–1078.
- 29 Liang XR, Liang JF, Yin T, *et al.* A lightweight method for face expression recognition based on improved MobileNetV3. IET Image Processing, 2023, 17(8): 2375–2384. [doi: [10.1049/ipr2.12798](https://doi.org/10.1049/ipr2.12798)]

(校对责编: 张重毅)