

基于改进残差网络的驾驶员异常情绪识别方法

潘东杰¹, 秦晓¹, 崔玉霞^{1†}, 王宪伦^{1,2}

(1. 青岛科技大学机电工程学院, 山东 青岛 266061; 2. 青岛安捷医疗科技有限公司, 山东 青岛 266100)

摘要:为了保障交通安全,避免负面情绪影响驾驶员状态导致交通惨剧的发生,提出了一种结合注意力算法的改进残差网络来监测车辆驾驶场景中驾驶员的异常情绪。首先,使用自适应直方图等一系列手段对图像进行预处理,突出表情特征;然后,采用ECA块自优化通道的权重值,实现局部跨信道交互,提高精度,增强信道特性;最后,将注意力机制与特征提取算法融合实现表情识别。结果表明,与改进前的网络相比,该表情识别方法参数更少,速度更快,分类效果更好,在CK+、FER2013和RAF-DB数据集上的准确率分别为96.38%、72.25%和87.97%。该算法能够更有效地识别驾驶员的异常情绪,对预防交通事故与保障行车安全具有积极意义。

关键词:深度学习;计算机视觉;表情识别;行为检测

中图分类号:TP242

文献标识码:A

Driver Abnormal Emotion Recognition Based on Improved RepVGG

PAN Dongjie¹, QIN Xiao¹, CUI Yuxia^{1†}, WANG Xianlun^{1,2}

(1. College of Electromechanical Engineering, Qingdao University of Science and Technology, Qingdao, Shandong 266061, China;

2. Qingdao Anjie Medical Technology Co., Ltd., Qingdao, Shandong 266100, China)

Abstract: To ensure traffic safety and avoid negative emotions affecting the driver's state leading to traffic accidents, this paper proposes an improved residual network combined with attention algorithms to monitor the abnormal emotions of drivers in vehicle driving scenarios. Firstly, a series of techniques such as adaptive histograms are used to preprocess the image and highlight facial features; then, using ECA blocks to self-optimize the weight values of channels, achieving local cross channel interaction, improving accuracy, and enhancing channel characteristics; finally, the attention mechanism is integrated with the feature extraction algorithm to achieve expression recognition. The results show that compared with the pre-improved network, this expression recognition method has fewer parameters, faster speed, and better classification performance. The accuracy on the CK+, FER2013, and RAF-DB datasets is 96.38%, 72.25%, and 87.97%, respectively. This algorithm can more effectively identify abnormal emotions of drivers, which has a positive significance for preventing traffic accidents and ensuring driving safety.

Key words: deep learning; computer vision; facial emotion recognition; behavior monitoring

作为人类感知系统的重要组成部分,情绪对交通安全的影响越来越受到人们的关注,研究表明愤怒和沮丧等负面情绪会影响交通安全并增加事故风险^[1-2]。Winter 和 Dodou 的一项研究发现,愤怒和沮丧等负面情绪会影响交通安全并增加事故风险^[1-2]。Winter 和 Dodou 的一项研究发现,愤怒和沮丧等负面情绪会影响交通安全并增加事故风险^[1-2]。

Winter 和 Dodou 的一项研究发现,愤怒和沮丧等负面情绪会影响交通安全并增加事故风险^[1-2]。

收稿日期:2023-11-30

基金项目:国家自然科学基金资助项目(51105213)

作者简介:潘东杰(1999—),男,山东东营人,硕士研究生,研究方向:机器人及智能制造技术。

†通信联系人, E-mail: yxcui@126.com

怒情绪下的司机更有可能做出攻击性的驾驶行为，如尾随和超速，这会增加交通事故的风险^[3]。Dula 和 Geller 的一项研究发现，驾驶员的压力和攻击性驾驶行为之间存在相关性，这会增加交通事故的可能性^[4]。Lajunen 和 Summala 的一项研究表明，司机在愤怒和沮丧状态下的车祸率要高得多^[5]。

使用车载摄像头和智能系统来检测 and 解决驾驶员疲劳和其他危险行为是提高驾驶安全的一种有效方法。Kanade 等开发了一种智能驾驶员监控系统，该系统使用人脸识别和眼动追踪技术来检测驾驶员的注意力是否集中^[6]。

在驾驶员情绪识别领域，面部表情识别算法普遍采用深度学习方法，例如，Xie 等开发了一种从面部表情中识别情绪的深度学习方法。该研究取得了显著的成功，证明了识别和分类情绪的准确性^[7]。此外，Li 等的一项研究提出了一种用于面部表情识别的深度卷积神经网络(CNN)模型，在基准数据集上获得了良好的结果^[8]。

为了解决残差网络训练过程中出现的

特征丢失问题，更好地捕捉人脸微表情等特征，优化表情的分类，本研究将 RepVGG 与 ECAnet 融合，在实现跨通道交互的同时，向网络添加多条梯度流通通道。与传统的卷积神经网络相比，本研究提出的融合网络 REN-RepVGG，推理更快，参数更少，鲁棒性更强，能够更好地识别出有害的情绪。

1 基础算法

本文所提出的算法以 RepVGG 表情识别网络为基础。RepVGG 以 VGG 网络为基础，添加了残差神经网络中的 Identity 和残差结构，并提出了结构重参数化技术。RepVGG 网络由大小为 3×3 的卷积核组成，内嵌有 ReLU 激活函数。

RepVGG 通过结构重参数化，将多分支上的卷积层融合相加合并为一个卷积层，实现从多分支模型到单路模型的转换。因此，RepVGG 网络的内存占用率更小，可以获得更快的推理速度。RepVGG 的多分支结构如图 1 所示。

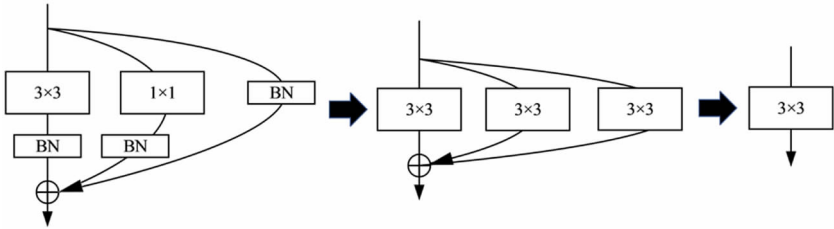


图 1 RepVGG 多分支结构

残差网络存在多个分支，训练一个网络相当于训练了多个网络。模型的加深不仅可以使模型的深层获得较好的鲁棒性，还可以更好地处理深层网络的梯度消失问题。

2 注意力引入与损失函数的优化

2.1 通道注意力的引入

与传统的简单特征融合方法相比，通道注意力机制可以更精细地控制每个通道的权重，更好地捕捉特征并抑制噪声。ECAnet 由 SENet 的激励部分改进而来。ECA 模块在不降维的情况下使用 GAP 方法对特征图进行聚类，通过自适应函数确定卷积核的大小 k ，执行一维卷积运算，最后使用 Sigmoid 函数对其进行归一化，并将其与原始特征图的相应通道相乘，以此获得加权特征图。ECAnet 结构原理如图 2 所示^[10]。

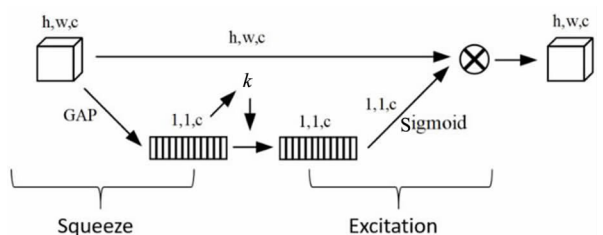


图 2 ECA 模块

ECA 网络模块使用大小为 k 的一维卷积在原始信道数据之间进行卷积运算，实现本地信道信息的融合，实现局部跨信道交互，从而有效提高模型精度，增强信道特性，避免特征丢失。变量“ k ”表示跨通道交互的覆盖范围。其中，“通道”是输入特征的通道数量，即：

$$k = \left\lceil \frac{\log_2 C}{\gamma} + \frac{b}{\gamma} \right\rceil \tag{1}$$

其中， $\gamma = 2, b = 1$ 。

增加 ECA 模块可以避免在信道保持学习过程中丢失信道重构的有用信息，减少模型中的参数

数量,实现局部跨信道交互,有效提高模型精度,增强信道特性,从而提高模型性能。

2.2 损失函数的优化

中心损失函数可以使类内的特征接近中心点,并使类内变化最小化;然而,它本身并不能完全实现类间分类。其次,Softmax 学习到的类特征不具有足够的区分性,为了增强其区分能力,本研究将两个损失函数组合使用。

假设特征图尺寸是 $(x \times y \times M)$,将输入特征图扁平化为 $[N \times 1]$ 个向量 $(N = x \times y \times M)$,之后将扁平化的向量送进全连接层,全连接层参数权重为 $W(T \times N)$, T 表示分类的类别数,全连接层处理后会得到 $[T \times 1]$ 个的向量。之后将其输入到 Softmax 层并归一到 $[0, 1]$ 之间,得到每个类别的概率,最后再通过交叉熵计算损失。

Softmax 损失函数如式(2)所示:

$$L_{\text{softmax}} = -\frac{1}{T} \sum_{j=1}^T \lg \frac{e^{a_j}}{\sum_{k=1}^T e^{a_k}} \quad (2)$$

其中, a_j 表示这 $[T \times 1]$ 个向量中的第 j 个值,分母表示所有值的求和。

设输入的人脸图像为 x_i ,其人脸对应的类别为 y_i ,给每个类别规定中心点的位置 c_{y_i} ,使每个人脸图像对应的特征 x_i 尽量地靠近每个对应类别的中心点 c_{y_i} 。在理想状态下 c_{y_i} 应该随着特征的改变而调整,这就是说将整个训练样本进行计算,然后根据每个类别对应所有图片特征的平均值计算结果来确定类别中心点的位置 c_{y_i} 。

$$L_{\text{center}} = \frac{1}{2} \sum_{i=1}^m \|x_i - c_{y_i}\|_2^2 \quad (3)$$

通过参数 λ 调整两个函数的比值:

$$L = L_{\text{softmax}} + \lambda L_{\text{center}} = -\frac{1}{T} \sum_{j=1}^T \lg \frac{e^{a_j}}{\sum_{k=1}^T e^{a_k}} + \frac{\lambda}{2} \sum_{i=1}^m \|x_i - c_{y_i}\|_2^2 \quad (4)$$

当 $\lambda = 0$ 时,式(4)可以看成是只有 Softmax 损失函数的特例。

2.3 表情识别算法的网络

在车辆行驶过程中,驾驶员的头部和面部通常固定在同一位置。使用车载摄像头捕捉驾驶员的面部图像后,图像将被裁剪成合适的大小。

驾驶员异常情绪的识别过程中,使用摄像头捕捉到图像,进行人脸检测,并对采集到的图像信息进行预处理;之后,送入识别网络进行特征提取与

特征识别,实现对面脸表情的识别与分类。

图像预处理的主要内容包括调整样本图像的对比度,以便于模型学习面部表情特征,从而使后续识别更加准确。

表情识别的主要内容包括特征提取与特征分类。预处理后的图片被送入到四个串联的 RepVGG 和 ECA 融合而成的 REN 块结构。该结构如图 3 所示。

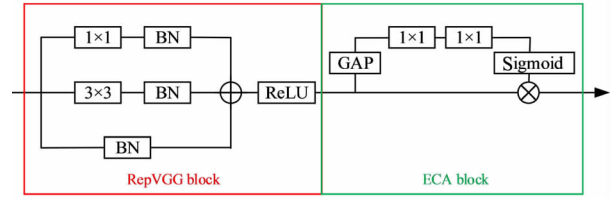


图 3 REN 块结构

提取特征后,图片经全连接层送入分类网络。使用 Softmax 与中心损失(Center loss)函数实现分类过程,并判断人员当前的情绪状态是否异常。表情识别网络 REN-RepVGG 的总体结构如图 4 所示。

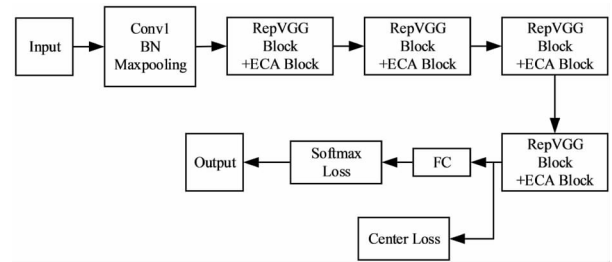


图 4 REN-RepVGG 框架示意图

3 实验设置与数据分析

实验环境包括 Windows 10 操作系统、AMD 5600X 处理器、NVIDIA GeForce RTX 3060Ti 显卡、CUDA 12.0 加速库、Python 3.6 编程语言和 Pytorch 1.9 深度学习框架。

3.1 实验数据集

为了全面评估本文提出的 REN-RepVGG 面部表情识别模型的性能,选择了三个常用的公共数据集,即 CK+、FER2013 和 RAF-DB 作为实验数据集。

CK+(扩展 Cohn-Kanade 数据集)是 CK 数据集的扩展版本,是 123 名志愿者在受控实验室环境中拍摄的面部表情图像的集合。FER2013 数据集在 2013 年的国际机器学习会议上推出,数据集由 35886 张尺寸为 48×48 像素的灰度面部表情图像

组成。RAF-DB 是一个大规模的面部表情识别数据集,包含从互联网下载的 30000 多张面部表情图像。

3.2 图像预处理与参数设置

本文中使用的对比度受限自适应直方图均衡 (CLAHE)^[11]可以在直方图中均匀分布样本亮度,

从而在不影响整体对比度效果的情况下增强局部对比度。原始图像、直方图均衡和自适应直方图均衡效果的比较如图 5 所示。从图中可以看出,自适应直方图均衡后,图像的整体对比度增加,面部信息不丢失,面部特征更加突出。

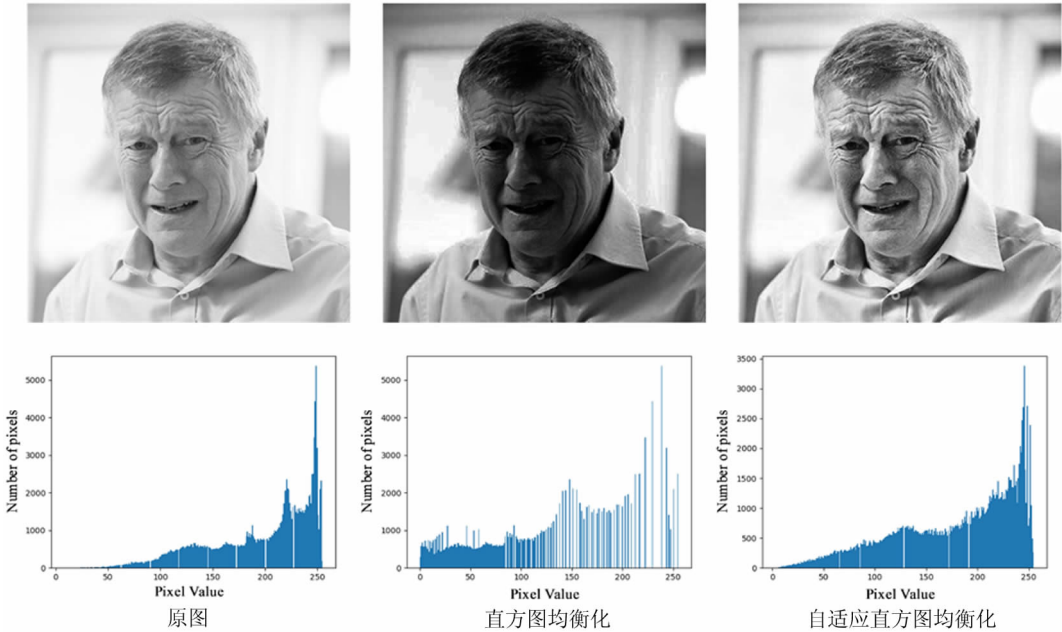


图 5 图像预处理效果对比图

网络模型的参数设置对模型的训练和模型性能有很大影响,选择合适的参数可以得到优秀的模型性能。本次实验对于三个数据集设置的迭代次数为 120, batch size 设置为 64, 初始学习率为 0.001。本文选用 Softmax 和中心损失函数相结合的损失函数求解误差,使用 Nesterov 动量的 SGD 和 Adam 优化器进行联合优化进行权重修整,SGD 的初始动量参数设置为 0.5,权重衰减为 0.00035;使用回调函数 RLRP(ReduceLROnPlateau) 作为学习率调度器对学习率进行调节,提升模型性能。

3.2 实验结果分析

在实验开始时,REN-RepVGG 网络被用作骨干网络。设置和改进所有模型参数后,在预处理的数据集上训练原始模型。CK +、FER2013 和 RAF-DB 数据集上,原始网络的准确率分别为 90.91%、68.35%和 83.93%。

对数据集进行预处理后,对 REN-RepVGG 模型进行训练,以获得最终的识别精度。在 CK+数据集上的准确率达到 96.38%,比原始模型高 5.47%。同时,在 FER2013 上的准确率达到 72.25%,比原始模型提高了 3.9%,在 RAF-DB 上的准确率达到 87.97%,提高了 3.04%。

改进前后每个数据集上改进模型的准确性比

较如图 6~图 8 所示。迭代结束时精度达到最大值并保持不变。结果表明,该模型已经实现了收敛,并且进一步的迭代不会显著提高精度。

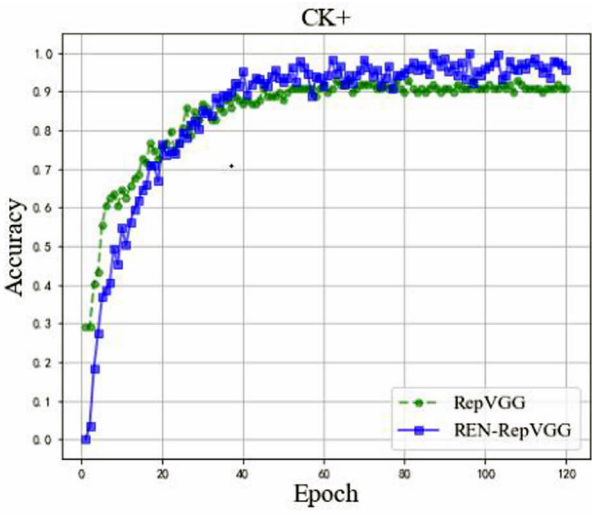


图 6 RepVGG 与改进网络准确率对比(CK+)

表 1、表 2 和表 3 将 REN-RepVGG 面部识别网络在 CK+、FER2013 和 RAF-DB 数据集上的实验结果与一些现有的改进面部识别算法进行了比较。REN-RepVGG 网络在 CK+数据集上的性

能略弱于其他算法,但在 FER2013 和 RAF-DB 数据集上性能较高,进一步证明了该算法在人脸识别分类网络中的优越性能。

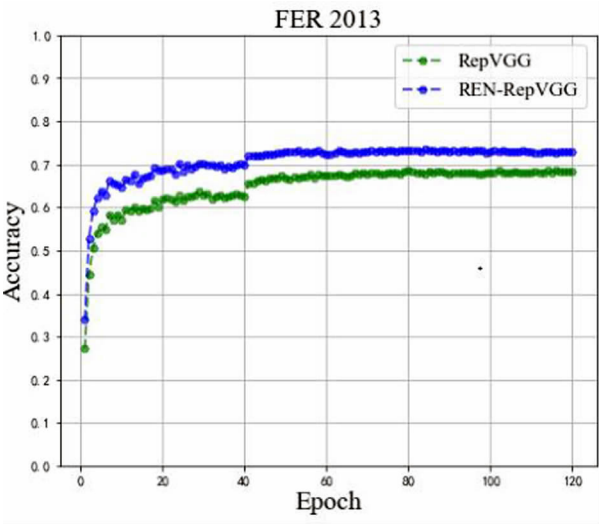


图 7 RepVGG 与改进网络准确率对比 (FER2013)

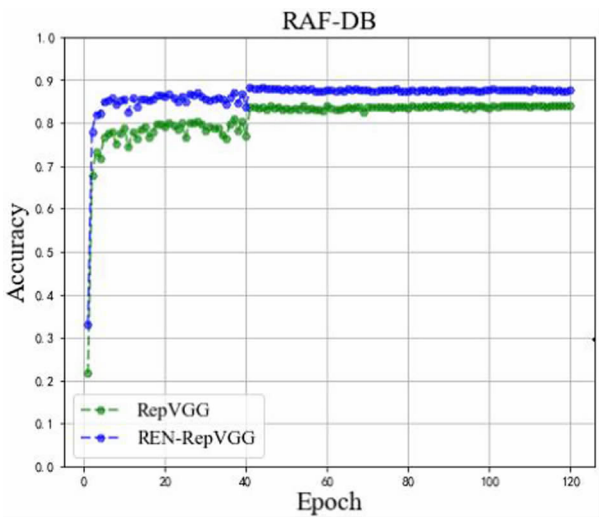


图 8 RepVGG 与改进网络准确率对比 (RAF-DB)

表 1 在 CK+数据集上的表情识别准确率对比		
作者	模型	准确率/%
Naveen et al. [12]	HOG-LBP	94.20
Pu et al. [13]	VGG	95.74
Shahzad et al. [14]	VGG-FCNN	98.40
本研究	REN-RepVGG	96.38

表 2 在 FER2013数据集上的表情识别准确率对比		
作者	模型	准确率/%
Fard et al. [15]	Resnet-50	71.48
Shen et al. [16]	CNN	71.89
Vulpe-Grigorași et al. [17]	Sobel-Resnet	72.16
本研究	REN-RepVGG	72.25

表 3 在 RAF-DB数据集上的表情识别准确率对比		
作者	模型	准确率/%
Fard et al. [15]	Resnet-50	85.93
Zhu et al. [18]	ResNeXt-50	86.80
Farzaneh et al. [19]	Resnet-18	87.78
本研究	REN-RepVGG	87.97

使用原始 RepVGG 网络训练三个数据集获得的混淆矩阵如图 9~图 11 所示。

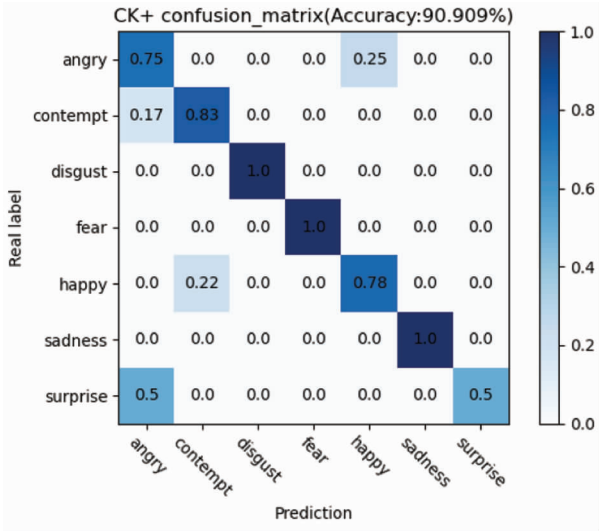


图 9 RepVGG 的混淆矩阵(CK+)

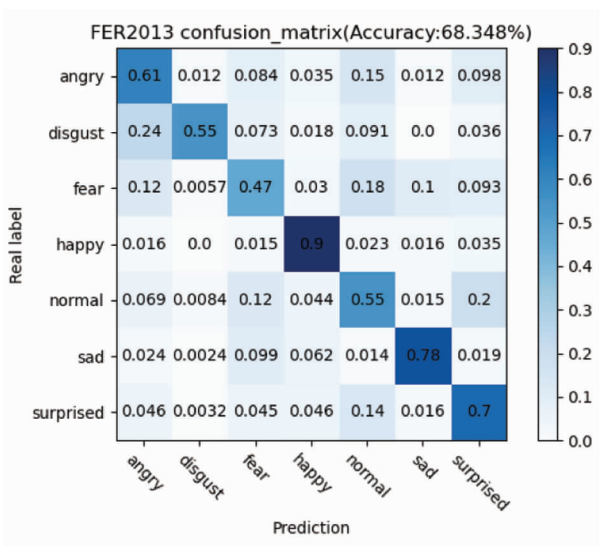


图 10 RepVGG 的混淆矩阵(FER2013)

图 12~图 14 是改进网络的混淆矩阵结果。可以看出,修改后的网络在各个数据集中对愤怒、轻蔑、幸福和惊讶等情绪的识别与分类精度显著提高。

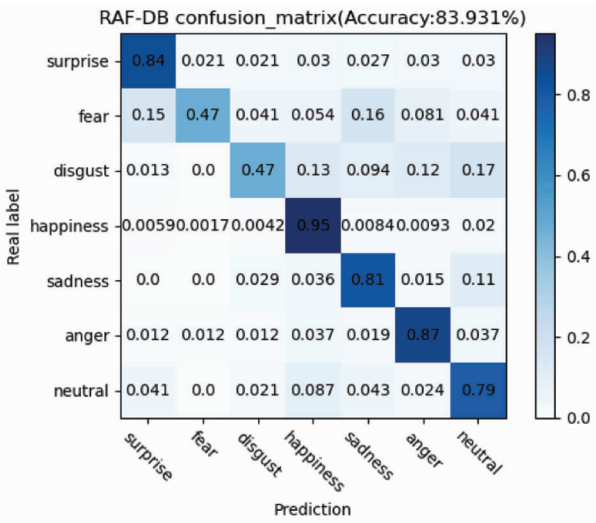


图 11 RepVGG 的混淆矩阵(RAF-DB)

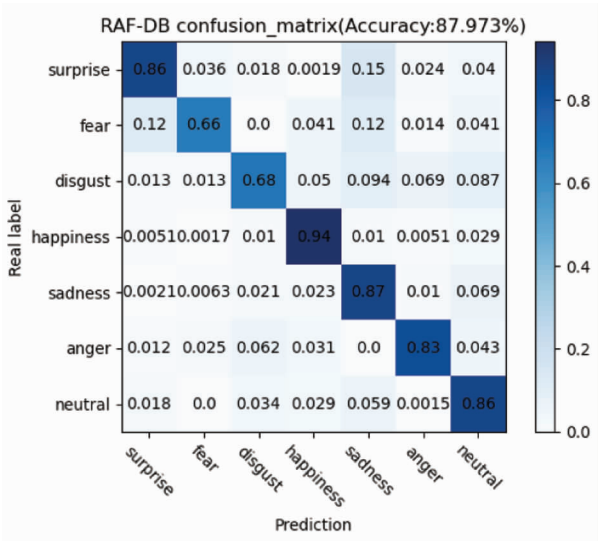


图 14 REN-RepVGG 的混淆矩阵(RAF-DB)

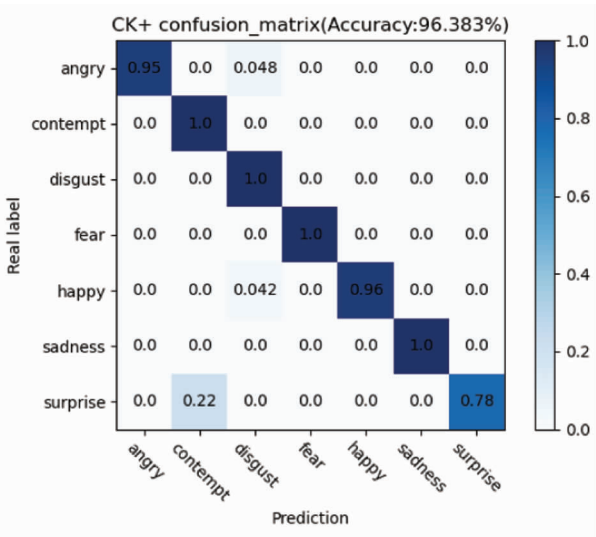


图 12 REN-RepVGG 的混淆矩阵(CK+)

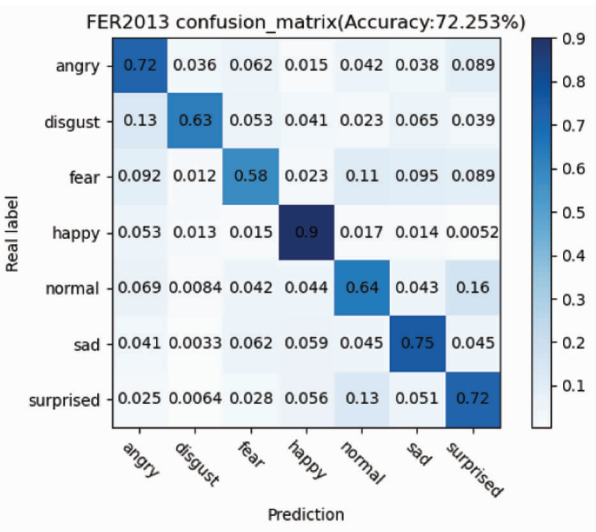


图 13 REN-RepVGG 的混淆矩阵(FER2013)

在 CK+数据集上,愤怒和轻蔑的识别准确度从 75% 和 83% 提升到了 95% 和 100%。在 FER2013 数据集上则分别从 61% 和 55% 提升到了 72% 和 63%。在 RAF-DB 数据集上则分别从 84% 和 47% 提升到了 86% 和 66%。证明 REN-RepVGG 相较于原始网络,对车辆驾驶过程中驾驶员的负面情绪识别准确度有了明显提高。

4 结 论

本文在 RepVGG 的基础上,对现有表情识别算法进行了改进,引入了 ECA 等注意力机制。REN-RepVGG 网络参数更少,占用内存少而且能实现多通道的并行训练。实验结果表明,REN-RepVGG 在数据集上的准确率相较于原始模型至少提升了 3%。本文将该算法与现有最先进的情感识别算法进行了比较验证,进一步验证了算法的优越性。最后,混淆矩阵的结果验证了 REN-RepVGG 针对愤怒等负面情绪的识别准确率相比于传统 RepVGG 提升了 12% 左右。

REN-RepVGG 较好地解决了传统 RepVGG 网络针对愤怒、轻蔑等驾驶员异常情绪识别精度低下、误分类较多的问题。该算法能更精准地实现对驾驶员异常情绪的识别,对预防交通事故与保障行车安全具有积极意义。

参考文献

[1] 苗馨宁. 基于面部表情和语音信息的驾驶人情绪识别研究[D]. 长春: 吉林大学, 2023.

- [2] 肖华飞. 驾驶员情绪识别、风险映射与预警方法研究[D]. 重庆:重庆大学,2022.
- [3] WINTER D J ,DODOU D . Response to second commentary on “the driver behaviour questionnaire as a predictor of accidents: a meta-analysis” [J]. *Journal of Safety Research*, 2012,43(1):94—98.
- [4] DULA S C ,GELLER E . Risky, aggressive, or emotional driving: addressing the need for consistent communication in research[J]. *Journal of Safety Research*,2003,34(5):559—566.
- [5] LAJUNEN T ,SUMMALA H . Can we trust self-reports of driving? Effects of impression management on driver behaviour questionnaire responses [J]. *Transportation Research Part F: Psychology and Behaviour*,2003,6(2):97—107.
- [6] KANADE P, DAVID F, KANADE S. Convolutional neural networks (CNN) based eye-gaze tracking system using machine learning algorithm[J]. *European Journal of Electrical Engineering and Computer Science*, 2021, 5(2): 36—40.
- [7] XIE S F,SHAN S G,CHEN X L, et al. Fusing local patterns of Gabor magnitude and phase for face recognition. [J]. *IEEE Transactions on Image Processing : a Publication of the IEEE Signal Processing Society*,2010,19(5):1349—1361.
- [8] LI J ,WANG Y D ,SEE J , et al. Micro-expression recognition based on 3D flow convolutional neural network[J]. *Pattern Analysis and Applications*,2019,22(4):1331—1339.
- [9] ZHANG Z X, WU Y X, ZHANG J, et al. Efficient channel attention for deep convolutional neural networks[C]. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2021.
- [10] LIU W Y, WEN Y D, YU Z D, et al. Large-margin softmax loss for convolutional neural networks[J]. *CoRR*,2016,abs/1612.02295.
- [11] REZA A M. Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement[J]. *Journal of VLSI Signal Processing Systems for Signal, Image and Video Technology*, 2004, 38: 35—44.
- [12] NAVEEN N H K ,SURESH A K ,GURU S M P , et al. Automatic facial expression recognition combining texture and shape features from prominent facial regions[J]. *IET Image Processing*,2022,17(4):1111—1125.
- [13] PU L P, ZHU L Y. Differential residual learning for facial expression recognition[C]//2021 The 5th International Conference on Machine Learning and Soft Computing. 2021: 103—108.
- [14] SHAHZAD T, IQBAL K, KHAN M A, et al. Role of zoning in facial expression using deep learning [J]. *IEEE*, 2023, 11: 16493—16508.
- [15] FARD A P, MAHOOR M H. Ad-corre: adaptive correlation-based loss for facial expression recognition in the wild [J]. *IEEE*, 2022, 10: 26756—26768.
- [16] SHEN X Y, XU X M, ZHUANG Y. Facial Emotion Recognition Based On Sobel-Resnet[C]//2021 IEEE 5th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). *IEEE*, 2021, 5: 484—488.
- [17] VULPE-GRIGORAȘ I A, GRIGORE O. Convolutional neural network hyperparameters optimization for facial emotion recognition [C]//International Symposium on Advanced Topics in Electrical Engineering (ATEE). *IEEE*, 2021: 1—5.
- [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[J]. *Sensors*, 2022, 22(4): 1350.
- [19] FARZANEH A H, QI X J. Facial expression recognition in the wild via deep attentive center loss[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2021: 2402—2411.