

Multi-National Face Classification and Recognition Based on MobileNet Network

Yijun Guo^{1,2}, Alimujiang Abudiriyimu³, Nurbiya Yadikar^{1,2}, Yali Zhu^{1,2}, Kurban Ubul^{1,2*}

¹College of Information Science and Engineering, Xinjiang University, Urumqi Xinjiang

²Key Laboratory of Signal and Information Processing, Xinjiang University, Urumqi Xinjiang

³Project Service Department, Xinjiang Uygur Autonomous Region Science and Technology Project Service Center, Urumqi Xinjiang

Email: *kurbanu@xju.edu.cn

Received: May 20th, 2020; accepted: Jun. 12th, 2020; published: Jun. 19th, 2020

Abstract

With the increasingly frequent economic, trade and cultural exchanges between countries, quickly and effectively distinguishing the identity of people in various countries is an important research in the field of face recognition. This paper focuses on the research of face classification and recognition in five Asian countries (Chinese, Japanese, Korean, Thailand and Indian). In this paper, face classification and recognition in five Asian countries are based on MobileNet. Because the faces of these five countries are similar, to reduce redundancy in this paper, octave convolution is inserted into the network to reduce redundancy and improve accuracy; and a method using a combination of center loss function and cross-entropy loss function is proposed to improve accuracy. After experimental verification, both the octave convolution and the center loss function proposed in this paper can improve the accuracy rate, and the highest accuracy rate can reach 87.84%, its Error top 1 is at least 0.120%.

Keywords

Multi-National Face Classification, Octave Convolution, Center Loss, Width Multiplier, Inverted Residuals

基于MobileNet网络多国人脸分类识别

郭奕君^{1,2}, 阿里木江·阿布迪日依木³, 努尔毕亚·亚地卡尔^{1,2}, 朱亚俐^{1,2}, 库尔班·吾布力^{1,2*}

¹新疆大学信息科学与工程学院, 新疆 乌鲁木齐

²新疆大学信号与信息处理重点实验室, 新疆 乌鲁木齐

³新疆维吾尔自治区科技项目服务中心项目服务部, 新疆 乌鲁木齐

*通讯作者。

文章引用: 郭奕君, 阿里木江·阿布迪日依木, 努尔毕亚·亚地卡尔, 朱亚俐, 库尔班·吾布力. 基于 MobileNet 网络多国人脸分类识别[J]. 图像与信号处理, 2020, 9(3): 146-155. DOI: 10.12677/jisp.2020.93018

Email: *kurbanu@xju.edu.cn

收稿日期: 2020年5月20日; 录用日期: 2020年6月12日; 发布日期: 2020年6月19日

摘要

随着各国经济贸易、文化交流往来的日益频繁, 快速、有效地区分各国人员身份是当前人脸识别领域的一项重要研究。本文针对亚洲区域五个国家(中国、日本、韩国、泰国、印度)进行人脸分类识别的研究, 本文基于MobileNet进行五国人脸分类识别, 因为这五国人脸较为相似, 为能有效降低冗余, 本文将八度卷积插入该网络中减少冗余, 提升精度; 并提出使用中心损失函数和交叉熵损失函数相结合的方法来提提升准确率。经过实验验证, 本文提出的在网络中插入八度卷积和中心损失函数两种改进方法均可以提升准确率, 其最高准确率可达87.84%, 其Error top 1最低达到0.120%。

关键词

多国人脸分类, 八度卷积, 中心损失函数, 宽度乘子, 倒残差模块

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着时代变迁与人类生活领域的交互, 为推动我国现代化体制建设进程, 促进世界经济全球化, 从改革开放时期以来, 我国始终坚持对外开放力度, 逐步扩大对外开放范围, 尤其是外贸方面, 使我国在对外贸易发展中取得长足的进步, 让我国同世界各国人民共同走向人类命运共同体。在国家间往来日益频繁的今天, 各国之间有经济、文化、繁育以及习俗之间的交融, 种族特征在随着时间的推移与人类间互相演变出现新的特征, 结合当前大数据知识应用体系留存不同国家人脸特征数据, 以逐步丰富人脸识别分析基础资源数据库。人脸识别就显得尤为重要, 它能有效对各个国家人员进行识别, 以便我国边检口岸等一系列安全系统能更有效、快速地识别各个国家人员, 构建更加有序的公共环境。也可以逐步丰富人脸识别分析基础资源数据库, 这对未来相关领域的研究将会是一份宝贵的财富。

近年来国内外也有不少研究者们对人脸种族识别进行了一定的研究。Sheerko Hma Salah [1]等人使用颜色和纹理的多层融合方案进行人脸种族特征的提取, 其样本数据是由多个数据库集合构成, 然后使用KNN分类器区分高加索人、东亚人、非洲人, 其平均准确率达96.5%。A. Rehman, G. Khan [2]等人使用Haralick + GLCM特征融合提取特征, 并使用预测随机数作为分类区分亚洲人、欧洲人和非洲人, 其准确率达到85.39%。黄惠[3]将PCA和2DPCA进行特征提取, 用多种不同的SVM作为分类器应用于自建中国少数民族(维吾尔族、柯尔克孜族、哈萨克族、回族、蒙古族)数据库进行民族识别, 实验表明, 使用RBF的SVM效果可达85%。王雅丽[4]使用CNN提取深度虹膜纹理特征作为底层向量, 并结合Fisher向量相融合, 使用SVM分类器区分亚洲人和非亚洲人, 最高准确率可达99.93%。邱盛[5]在深度学习的研究中, 利用了几种不同网络框架对自建的中国部分少数民族(汉族、蒙古族、藏族、维吾尔族)数据集进行民族识别分类, 其中包括卷积神经网络和集成神经网络, 并利用卷积神经网络搭建了一个基于这五个民族的人脸识别系统。

从当前国内外研究现状来看, 当前研究中基本都在区分人种, 而对同一区域不同国家的人脸面部研究几乎没有。因此本文提出针对同一区域(亚洲)上的五个国家(中国、日本、韩国、泰国、印度)进行人脸识别分类的研究。本文通过基于 MobileNet 网络上对本文自建的五国人脸数据集进行识别分类, 通过对原始网络、插入八度卷积以及加入中心损失函数后的实验数据对比, 验证本文提出方法的有效性。

2. 自建五国人脸数据集

2.1. 人脸数据集的采集

由于当前没有公开的这五国的人脸数据集, 且考虑到实地采集的困难性和限制性, 本文通过网络爬虫获取这五国(中国、日本、韩国、泰国、印度)的人脸数据样本。由于网络爬虫图像样本数据存在大量由光照、姿势、遮挡各种因素引起的大量差异, 本文通过使用, Multi-task convolutional neural network (MTCNN, 多任务卷积神经网络)对本文通过网络爬虫采集的人脸数据样本进行“清洗”, 即对人脸定位和对齐, 本文将原始数据集裁剪成了 128×128 尺寸大小。本文经过“清洗”后所建立的原始数据集样本共 10,266 张(中国 2881 张、日本 1747 张、韩国 2381 张、泰国 1971 张、印度 1286 张), 由于网络爬虫的限制导致得到的初始样本数据量不一致, 得到的部分样本数据如图 1 所示, 其样本顺序由左到右五列依次为中国、日本、韩国、泰国、印度。



Figure 1. Face samples of five countries

图 1. 部分五国人脸样本

2.2. 人脸数据集的扩充

由于使用深度学习, 需要大量人脸图像进行学习。故本文采取数据增强方法对样本数量进行扩充, 通过对原始照片旋转一定的小角度(角度旋转范围在 15 度以内)、改变饱和度、对比度、锐度以及亮度来进行样本数据的扩充。

3. 基于 Mobilenet 网络模型核心

3.1. MobileNet 网络

MobileNetV1 网络结构模型是 2017 年谷歌团队提出的[6], 该网络主要针对便捷式或嵌入式设备使用, 属于轻量级卷积神经网络, 在 2018~2019 年谷歌团队在该网络基础上又先后提出了 MobileNetV2 网络[7]、

MobileNetV3 网络[8]。本文的所有实验都是基于 MobileNetV1 和 MobileNetV2 两个模型上进行的。该网络核心思想是深度可分离卷积，其原理是把普通卷积拆分成一个深度卷积和一个点卷积的组合以减少运算量。其原理如图 2 所示。

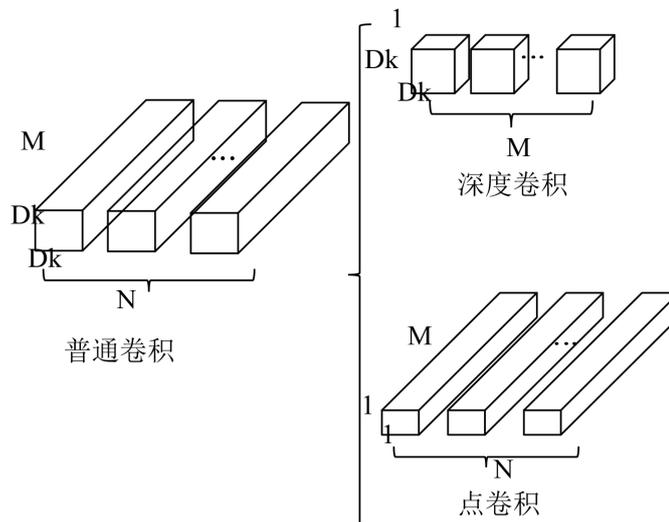


Figure 2. Deep separable convolution principle
图 2. 深度可分离卷积原理

MobileNet 网络的第二个核心是提出了两个超参数：宽度因子 α 、分辨率因子 β ，这两个超参数的目的也是减少运算量。理论上来说，宽度因子 α 的取值越大，其训练精度越高，相应地训练时间也和参数量也增加。同理，对于分辨率因子 β ，其取值越大，训练精度越高，相应地训练时间也和参数量也增加。宽度因子 α 取值范围在 $[0, 1]$ ，通常定义为四个值：0.25、0.5、0.75、1。分辨率因子 β 通常是 128×128 、 160×160 、 196×196 、 224×224 。

MobileNetV1 网络共 28 层，其网络结构中，除了第一层和最后一层使用普通卷积，其余层均采用深度可分离卷积代替普通卷积。每一层卷积后面均需使用 BN (Batch Normalization) 和 Relu6 激活函数，目的是为了能有效调整每一层的特征图分布，加速网络收敛，增强了网络的泛化能力。

MobileNetV2 网络共有 54 层，在 MobileNetV2 网络模型中，它是将 MobileNetV1 网络做了两处改进：1) 仿照 ResNet 网络的核心思想，加入了“倒残差模块”，即对原始特征图做“扩张 - 卷积 - 压缩”的操作；2) 去除最后层的 Relu6 激活函数，直接替换为线性输出。

3.2. 八度卷积

八度卷积[9]是 2019 年提出的一种新卷积，八度卷积的目的是为了提高精度的同时，减少冗余。它的核心思想就是通过空间尺度变化将图像拆分为高频、低频两块，此操作可以大大降低参数量。完成此步骤是通过使用网络定义的超参数 α 将图像分解成高频低频两部分，通过高斯滤波器使得低频的空间是高频空间的一半，将原图像可以拆分成图 3 所示，通过该方法可以有效减少空间冗余。

八度卷积的另一个优点是可以很好的嵌入到任一神经网络中，即可以在不改变原网络结构及参数的前提下，通过对超参数 α 的设定即可用八度卷积代替普通卷积，较好的解决低频信息的冗余问题。八度卷积的使用中，除第一层和最后一层外对超参数设定为： $\alpha_{in} = \alpha_{out} = \alpha$ ，第一层的输入和最后一层的输出均为 0，即可完成对八度卷积的即插即用。

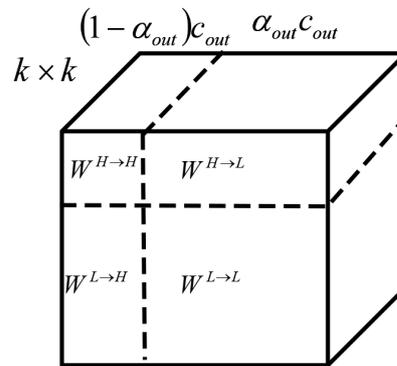


Figure 3. Principle of octave convolution
图 3. 八度卷积原理图

4. 损失函数

由于亚洲这五国人脸面部极为相似，尤其是中日韩三国，为了能更好地将他们进行分类研究，本文提出将中心损失函数[10]与交叉熵损失函数相结合的方法来改进实验精度。中心损失函数的作用是在分类识别中，减小类内的差异，也就是说使同一类的样本相似性更大，让同一类能尽可能向样本中心靠拢，而常用的交叉熵损失函数是为了增大各类别间的距离，减少类间的交叉问题。两者结合可以更好的区分开这五国人员。在两者结合后，整个网络的损失函数如公式(1)所示。

$$L = L_s + \lambda L_c$$

$$= -\sum_{i=1}^m \log \frac{e^{\omega_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^n e^{\omega_{y_j}^T x_i + b_{y_j}}} + \frac{\lambda}{2} \sum_{i=1}^m \|x_i - c_{y_i}\|_2^2 \quad (1)$$

其中， L_s 、 L_c 分别代表交叉熵损失函数和中心损失函数， λ 表示两个损失函数之间的比重。 x_i 表示在全连接层前的特征， c_{y_i} 代表第 y_i 个类的特征中心， m 代表的是每个 batch 的样本数量， n 是类别数量，本文中的 n 是 5。

5. 基于 MobileNet 网络实验分析

5.1. 评价标准

本文实验中通过采用网络爬虫自建的五国人脸数据集，分别是中国、日本、韩国、泰国、印度五国。实验共用 68,495 张样本数据，各国分别 13,699 张，训练测试按照 7:3 比例随机抽取，即训练共 47,945 张，测试 20,550 张。在下面三个网络模型中，模型迭代次数均设置为 1000 次。本文中所有实验均在 64 位 Windows 7 的环境，内存为 16 GB，CUDA 10.1，Python 3.6 配置 mxnet 和 python 环境搭载 GPU 的主机上进行编程调试。

在 MobileNetV1 网络结构中，为了验证宽度因子 α 对实验结果的影响，分别选取了 $\alpha = 0.75$ (MobileNetV1_075)、 $\alpha = 1$ (MobileNetV1_100) 两个模型进行实验，然后又对比了相同参数下，MobileNetV1_100 和 MobileNetV2_100 两个模型，验证 MobileNetV2 网络相较于 MobileNetV1 网络在本实验中效果是否得到提升。在本文整个实验评价指标均采用 Accuracy 和 Error top 1，其中 Accuracy 是针对训练集进行评价，Error top 1 是针对测试集进行评价的。其公式如下：

$$Accuracy = \frac{\text{正确预测样本数}}{\text{总样本数}} \quad (2)$$

$$Error-top1 = \frac{\text{正确标记与输出不同的样本数}}{\text{总样本数}} \quad (3)$$

5.2. 实验结果分析

本文基于 MobileNetV1 和 MobileNetV2 两个网络上进行学习，首先是将五国人脸数据集应用在原网络上进行分类识别，其次是在两个网络模型上插入八度卷积后进行分类学习，最后是在此基础上改进损失函数的实验结果。图 4 是在 MobileNetV1_075 模型上的实验结果，图 5 是在 MobileNetV1_100 模型上的实验结果，图 6 是在 MobileNetV2_100 模型上的实验结果。

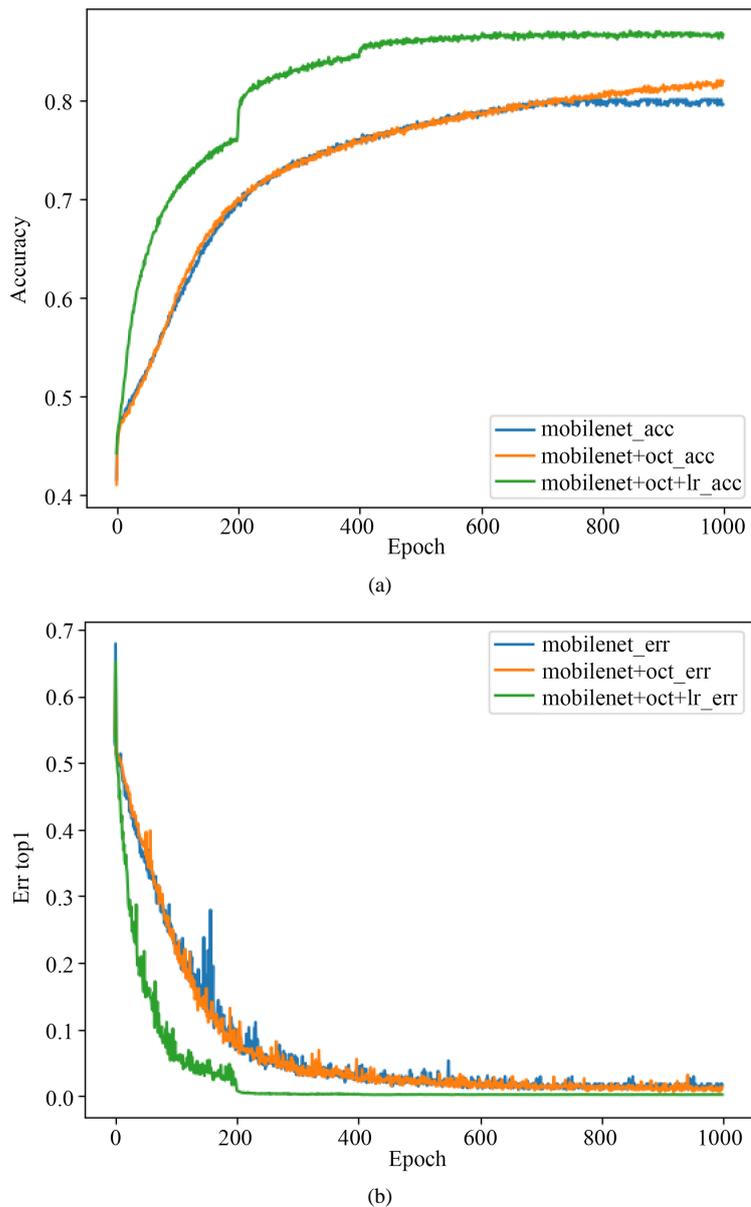
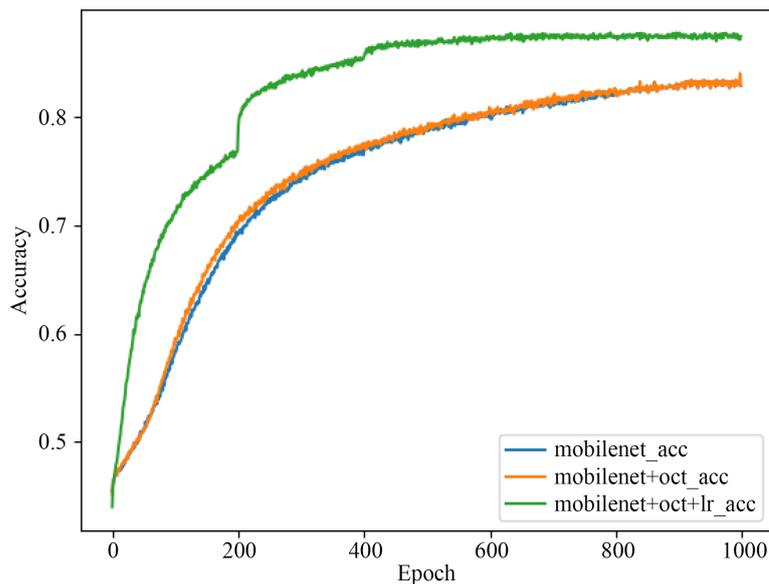


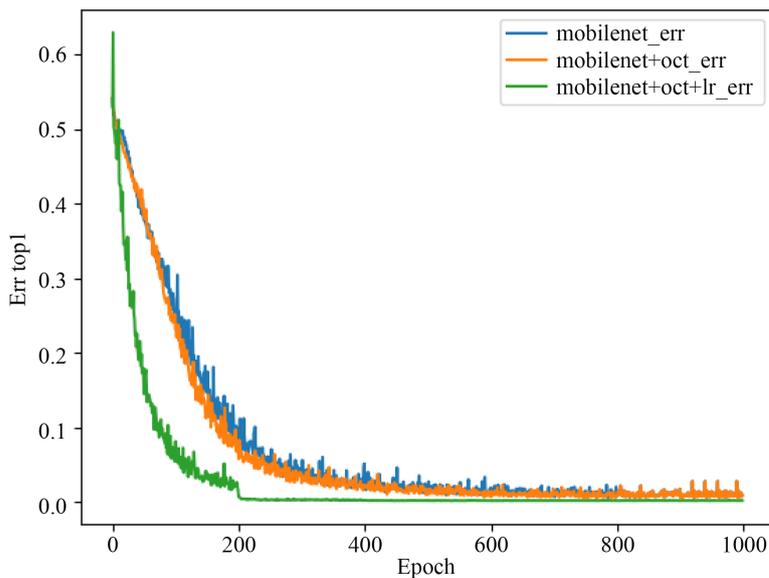
Figure 4. (a) Based on MobilenetV1_075 training result graph; (b) Based on MobilenetV1_075 test result graph

图 4. (a) 基于 MobilenetV1_075 训练结果图; (b) 基于 MobilenetV1_075 测试结果图

由 MobilenetV1_075 两个实验结果图可以看出，在加入八度卷积后，其训练精度得到了一定的提升，然后在加入八度卷积基础上改进其损失函数，我们可以看到不管是在训练集上的 Accuracy 还是测试集上的 Error top 1 的得到了显著改善。



(a)



(b)

Figure 5. (a) Based on MobilenetV1_100 training result graph; (b) Based on MobilenetV1_100 test result graph

图 5. (a) 基于 MobilenetV1_100 训练结果图; (b) 基于 MobilenetV1_100 测试结果图

由 MobilenetV1_100 两个实验结果图可以看出，在加入八度卷积后，其训练精度曲线与原模型基本保持吻合，只得到了较不明显的提升，然后在加入八度卷积基础上改进其损失函数，我们可以看到不管是在训练集上的 Accuracy 还是测试集上的 Error top 1 的得到了显著改善。

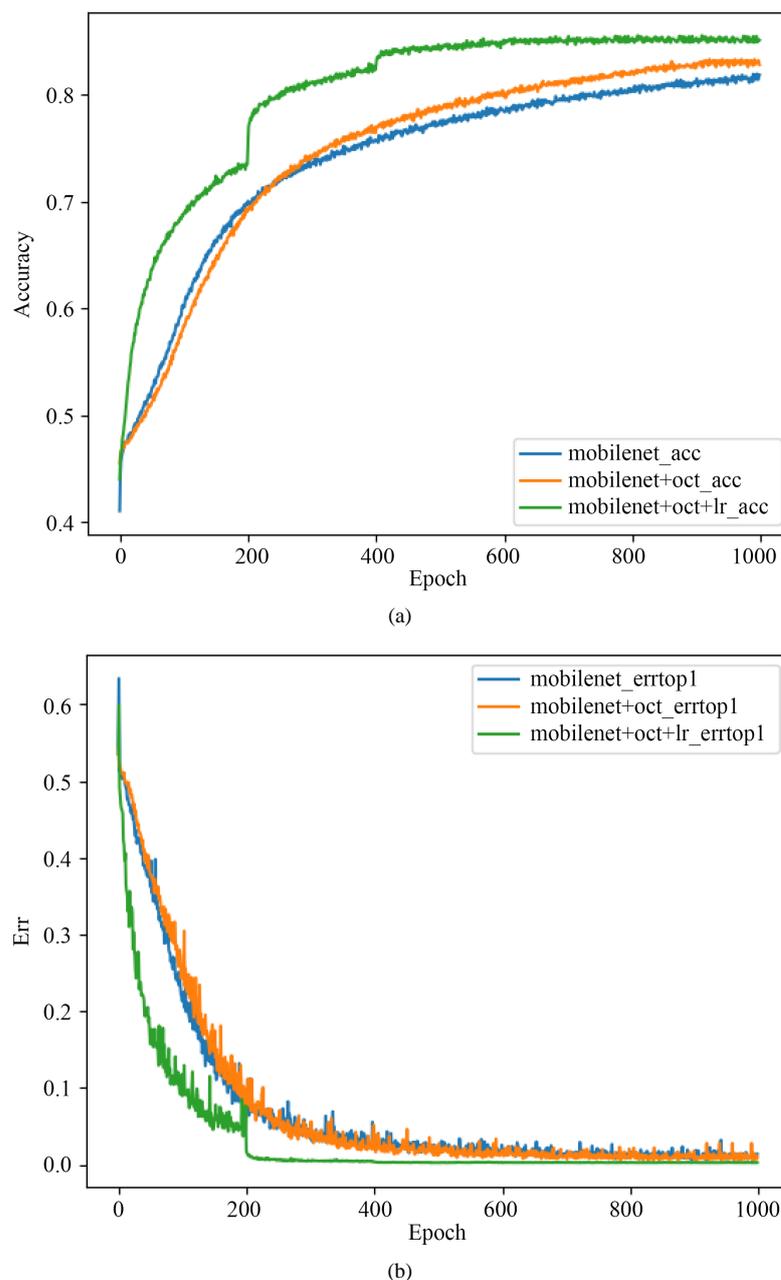


Figure 6. (a) Based on MobilenetV2_100 training result graph; (b) Based on MobilenetV2_100 test result graph

图 6. (a) 基于 MobilenetV2_100 训练结果图; (b) 基于 MobilenetV2_100 测试结果图

由 MobilenetV2_100 两个实验结果图可以看出, 在加入八度卷积后, 其训练精度得到了较为显著的提升, 而后在加入八度卷积基础上改进其损失函数, 我们可以看到不管是在训练集上的 Accuracy 还是测试集上的 Error top 1 的得到了显著改善。具体实验中我们还要考虑训练时间等其他因素来判别三个网络模型的优越性。其实验结果统计如表 1 所示。表中 p1、p2、p3 分别代表上述三个实验, 即 p1 是原网络上得到的结果, p2 是加入八度卷积后的结果, p3 是在 p2 实验基础上加入中心损失函数与交叉熵损失函数之后得到的实验结果。

Table 1. Based on the results of facial classification experiments in the five countries**表 1.** 基于五国人脸分类实验结果

网络模型	方法	Accuracy	Error-top 1	训练时间
MobileNetV1_075	p1	80.04%	1.051%	185,241 s
	p2	81.96%	0.715%	184,491 s
	p3	86.82%	0.185%	215,000 s
MobileNetV1_100	p1	83.21%	0.671%	208,791 s
	p2	83.98%	0.613%	205,794 s
	p3	87.84%	0.120%	210,789 s
MobileNetV2_100	p1	81.98%	0.725%	319,546 s
	p2	83.14%	0.694%	320,456 s
	p3	85.57%	0.212%	322,600 s

由上述实验结果图和实验总结表可知，在原始模型上加入八度卷积后，MobileNetV1_075、MobileNetV1_100、MobileNetV2_100 的三个模型上 Accuracy 分别提升了 1.92%、0.77%、1.16%。Error-top 1 分别下降了 0.336%、0.058%、0.031%。在此基础上加入中心损失函数后，三个网络模型的精度均达到了最高，分别是 86.82%、87.84%、85.57%，相较原网络分别提升了 6.78%、4.63%、3.59%。Error-top 1 分别下降了 0.866%、0.551%、0.513%，均在该方法上达到了最低。对比 MobileNetV1_075 和 MobileNetV1_100 两组实验结果，我们可以看出在不改变其他参数仅改变宽度因子 α 时，其实验结果得到了提升，证明了宽度乘子对实验结果影响的有效性。对比 MobileNetV1_100 和 MobileNetV2_100 实验结果可以看出，在本实验中 MobileNetV2 网络效果低于 MobileNetV1 网络，且训练时间是 MobileNetV1 网络上的 1.5 倍。

6. 总结与展望

本文提出将八度卷积应用到五国人脸(中国、日本、韩国、泰国、印度)分类识别中，有效减少了冗余提升了精度，且在使用了中心损失函数和交叉熵损失函数相结合后，其精度达到了最高，整个实验结果中，精度最高可达到 87.84%，其 Error-top 1 达到了 0.120%，实验结果验证了本文提出改进方案的可行性，能有效提升准确率。本文的实验也说明了 MobileNet 网络中超参数 α 不同时实验效果的差异性，有效区分各国人脸信息在未来是一个重要的研究领域，为构建和谐安全幸福的良好社会环境有着不可替代的作用。在后续的学习研究中时，主要着力于解决以下问题：

1) 人脸数据样本对于区分各国人员起着至关重要的一步，因此如何能构建更好的各国人脸数据集应更进一步的研究。且本文所构建的人脸数据集仅限于这五国的明星，并没有很好地包含各个国家不同年龄等因素，这需要在后续研究中不断改进。

2) 由于本文选取的网络模型是“轻便型”网络模型，因此可以针对此构建一个可适用于移动端口等的各国人员分类识别系统。

基金项目

国家自然科学基金(61862061, 61563052)。

参考文献

- [1] Du, H.B., Salah, S.H. and Ahmed, H.O. (2014) A Color and Texture Based Multi-Level Fusion Scheme for Ethnicity Identification. *Proceedings of SPIE—The International Society for Optical Engineering*, Baltimore, 22 May 2014, 91200B. <https://doi.org/10.1117/12.2057722>
- [2] Rehman, A., Khan, G., Siddiqi, A., *et al.* (2018) Modified Texture Features from Histogram and Gray Level Co-Occurrence Matrix of Facial Data for Ethnicity Detection. *5th International Multi-Topic ICT Conference (IMTIC)*, Jamshoro, 25-27 April 2018, 1-6. <https://doi.org/10.1109/IMTIC.2018.8467231>
- [3] 黄慧. 基于 PCA 与 2DPCA 的少数民族人脸识别比较[D]: [硕士学位论文]. 新疆: 伊犁师范学院电子与信息系, 2016.
- [4] 王雅丽, 马静, 李海青, 等. 基于虹膜纹理深度特征和 Fisher 向量的人种分类[J]. *中国图象图形学报*, 2018, 23(1): 28-38.
- [5] 邱盛. 基于深度学习的人脸民族识别[D]: [硕士学位论文]. 广州: 华南理工大学计算机工程与科学系, 2016.
- [6] Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., *et al.* (2017) Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 1-9.
- [7] Sandler, M., Howard, A.G., Zhu, M., Chen, L.C., *et al.* (2018) MobileNetV2: Inverted Residuals and Linear Bottlenecks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Salt Lake City, 18-23 June 2018, 4510-4520. <https://doi.org/10.1109/CVPR.2018.00474>
- [8] Howard, A.G, Sandler, M., Chu, G., Chen, L.C., Chen, B., Tan, M., *et al.* (2019) Searching for MobilenetV3. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seoul, 27 October-2 November 2019, 1314-1324. <https://doi.org/10.1109/ICCV.2019.00140>
- [9] Chen, Y.P., Fang, H.Q., Xu, B., *et al.* (2019). Drop an Octave: Reducing Spatial Redundancy in Convolutional Neural Networks with Octave Convolution. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, 27 October-2 November 2019, 3434-3443. <https://doi.org/10.1109/ICCV.2019.00353>
- [10] 魏哲, 王小华. 淋巴结转移检测的八度卷积方法[J]. *计算机应用*, 2020, 40(3): 723-727.