基于改进 MobileNet 网络的人脸表情识别

王韦祥[1] 周 欣[1,2] 何小海[1*] 卿粼波[1] 王正勇[1]
[1](四川大学电子信息学院 四川 成都 610065)
[2](中国信息安全测评中心 北京 100085)

摘 要 针对轻量级卷积神经网络 MobileNet 应用于人脸表情识别实时性较差、最小输入尺寸较大、准确率不高等问题,提出一种改进的 MobileNet 网络模型——M-MobileNet(Modified MobileNet)。M-MobileNet 具有比原网络更好的轻量级特性。该网络模型基于一种改进的深度可分离卷积层,不仅具有 MobileNet 模型中深度可分离卷积减少卷积计算量的特点,还解决了在深度卷积层后可能会导致信息丢失的问题。在分类器选择上,M-MobileNet 使用线性支持向量机(SVM)进行人脸表情分类,参数量较 MobileNet 网络大大减少。在 CK+、KDEF 数据集及移动端上的实验证明,改进后的 MobileNet 网络模型具有更好的识别性能。

关键词 MobileNet 表情识别 深度可分离卷积 支持向量机

中图分类号 TP391.4 **文献标志码** A DOI: 10.3969/j.issn.1000-386x.2020.04.023

FACIAL EXPRESSION RECOGNITION BASED ON IMPROVED MOBILENET

Wang Weixiang[1] Zhou Xin[1,2] He Xiaohai[1*] Qing Linbo[1] Wang Zhengyong[1]
[1](*College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, Sichuan, China*)
[2](*China Information Technology Security Evaluation Center, Beijing 100085, China*)

Abstract In order to solve the problems of poor real-time performance, large minimum input size and low accuracy of lightweight convolutional neural network MobileNet in facial expression recognition, we propose an improved MobileNet network model called M-MobileNet(Modified MobileNet). It has better lightweight characteristics than the original network. M-MobileNet is based on an improved depthwise separable convolution layer. It has the characteristics of depthwise separable convolution in MobileNet model to reduce the amount of convolution calculations, and can the problem that information may be lost after depthwise convolution layer. The M-MobileNet network model used the linear support vector machine(SVM) to classify facial expressions, and the parameters of M-MobileNet were greatly reduced, compared with the MobileNet network. The experiments on CK+ and KDEF data sets and mobile terminals show that the improved MobileNet network model has better recognition performance.

Keywords MobileNet Facial expression recognition Depthwise separable convolution Support vector machine

0 引 言

人脸表情识别是计算机视觉领域一大热点[1]。人脸表情作为人类情绪的直接表达,是非语言交际的一种形式[2]。人脸表情识别技术目前主要的应用领域包括人机交互(HCI)、安保、机器人制造、医疗、通信、汽车等。在人机交互、在线远程教育、互动游戏、智能交通等新兴应用中,自动面部表情识别系统是必要的[3]。

人脸表情识别的重点在于人脸表情特征的提取。对于人脸表情的提取,目前已出现两类特征提取方法。一种是基于传统人工设计的表情特征提取方法,如局

收稿日期:2019-06-23。国家自然科学基金项目(61871278);四川省科技计划项目(2018HH0143);四川省教育厅项目(18ZB0355);成都市产业集群协同创新项目(2016-XT00-00015-GX)。王韦祥,硕士生,主研领域:机器视觉,图像处理。周欣,博士生。何小海,教授。卿粼波,副教授。王正勇,副教授。

部二值模式(Local Binary Pattern,LBP)[4]、定向梯度直方图(Histogram of Oriented Gradients,HOG)[5]、尺度不变特征变换(Scale Invariant Feature Transform,SIFT)[6]等,这些方法不仅设计困难,并且难以提取图像的高阶统计特征。另一种是基于深度学习的表情特征提取方法,目前深度神经网络已广泛应用在图像、语音、自然语言处理等各个领域。为了适应不同的应用场景,越来越多的深度神经网络模型被提出,例如 AlexNet[7]、VGG[8]、GoogleNet[9]和 ResNet[10],这些网络模型被广泛应用于各个领域,在人脸表情特征提取及分类上,也取得了不错的效果。

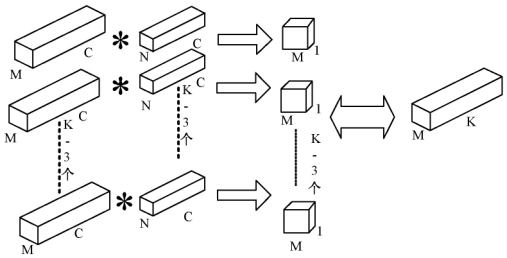
但随着深度神经网络模型的不断发展,其缺点也逐渐显现。网络模型的复杂化、模型参数的大量化等缺点,使得这些模型只能在一些特定的场合应用,移动端和嵌入式设备难以满足其需要的硬件要求。复杂网络模型对于硬件的高要求限制了其应用场景。基于此,Howard 等[11]在 2017 年 4 月提出了一个可以应用于移动端和嵌入式设备的 MobileNet 轻量化网络模型。文中提出的深度可分离卷积层,在保证精度损失不大的情况下,大大减少了网络的计算量,从而为计算设备"减负"。但是这个版本的 MobileNet 模型在深度卷积层后引入非线性激活函数 ReLU,而深度卷积没有改变通道数的能力,其提取的特征是单通道的,且 ReLU 激活函数在通道数较少的卷积层输出进行操作时,可能导致信息丢失。为了解决 MobileNet 第一版的问题,2018 年 1 月 Sandler 等[12]提出第二版的 MobileNet,即 MobileNetV2。MobileNetV2 使用了倒转的残差结构,即在采用当时流行的残差结构的同时,在进入深度卷积前先将输入送入 1×1 的点卷积,把特征图的通道数"压"下来,再经过深度卷积,最后经过一个 1×1 的点卷积层,将特征图通道数再"扩张"回去。即先"压缩",最后"扩张"回去。前两步的输出都采用 ReLU 激活函数处理,最后一步采用线性输出,可在一定程度上减少信息的丢失。然而此模型应用于实际人脸表情识别中识别效果依然不佳,且参数量和运算量很大,在安卓手机上测试时,实时性表现也不佳。除此之外,MobileNet 网络模型的最小输入尺寸为 96×96,而人脸表情识别允许更小的输入尺寸,因此 MobileNet 难以满足实际人脸表情识别的需要。

针对 MobileNet 的上述缺点,本文设计了一个基于输入尺寸为 $48\times48\times1$ 的单通道灰度图片的改进 MobileNet 模型——M-MobileNet,不仅大大减小了网络模型的参数量和运算量,使其更切合人脸表情识别的特点,还提升了其在人脸表情识别的实时性,除此之外,在 CK + 及 KDEF 人脸表情数据集上也取得了较高的识别率。由于 MobileNetV2 中采用"点卷积-深度卷积-点卷积"结构,其运算量和参数量较直接采用"深度卷积-点卷积"方式更多,MobileNetV2 使用的残差网络结构也较直接使用顺序级联方式更复杂,且在实验部分其在人脸表情识别率较其他优秀模型更低。因此为了更好地保留深度卷积后输出的特征,不同于 MobileNetV2 中"先压缩再扩张"的思想,M-MobileNet 在深度卷积层输出后,去掉用于提取非线性特征的激活层,采用线性输出,同时为了提高模型的非线性表达,依然在点卷积后采用 ReLU 激活函数,在各卷积层之间依然采用顺序级联方式,不使用残差连接方式,即采用本文提出的改进的深度可分离卷积层。另一方面,由于 MobileNetV1 和 MobileNetV2 使用 Softmax 分类器来进行分类,而由于人脸表情特征的特点,表情的类间区分本身就不高,所以 Softmax 在表情识别领域并不是很合适[13]。而 SVM 分类器作为一种具有较强泛化能力的通用学习算法,且对大数据高维特征的分类支持较好[14],其中 L2-SVM[15]具有较好的可微可导性,故本文网络使用 L2-SVM 代替 MobileNet 中 Softmax 分类器,训练深度神经网络进行分类。实验验证了 M-MobileNet 网络相比 MobileNet 网络可以有效提高人脸表情识别的准确率;同时为了验证使用 SVM 分类器的有效性,还增加了与"M-MobileNet + Softmax"网络模型的对比实验。

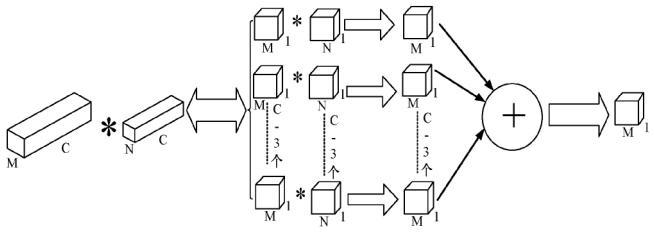
1 改进的深度可分离卷积层

传统标准卷积既过滤输入又将过滤后的输出进行组合,最终形成一组新的输出[11],如图 1(a)所示。假设输入特征图大小为 $M\times M$,通道数为 C,标准卷积的卷积核大小为 $N\times N$,个数为 K,并且假设输出与输入尺寸一致,则经过标准卷积后输出尺寸为 $M\times M$,输出通道数为 K。传统标准卷积过程,实际上包含了两步:特征过滤和将过滤后的结果组合,图 1(b)显示了输入特征图与第 $i(1\leqslant i\leqslant K)$ 个卷积核进行标准卷积的过程。在这个过程中,首先输入特征图中的每个通道与对应的卷积核的每个通道进行卷积,卷积的结果是形成了 C 个 $M\times M$ 的单通道特征图,然后将这 C 个结果合并,最终形成一个 $M\times M\times1$ 的单特征图。由于有 K 个卷积核,因此输入特征图与所有 K 个卷积核进行标准卷积后,共有 K 个 $M\times M\times1$ 结果,最终结果为 $M\times M\times K$ 的输出特征图,如图 1(a)所示。标准卷积的计算量为:

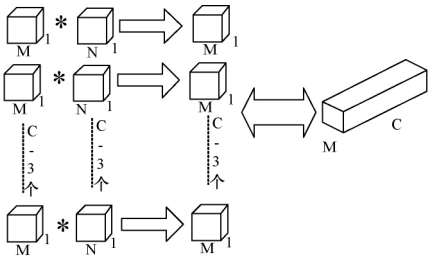
$$M \times M \times N \times N \times C \times K \tag{1}$$



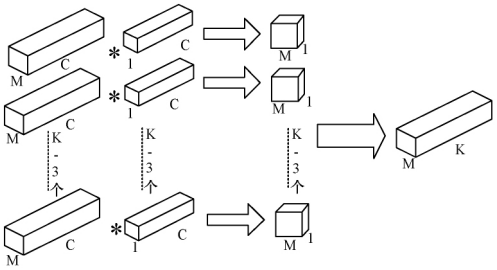
(a) 标准卷积



(b) 输入与第 $i(1 \leqslant i \leqslant K)$ 个卷积核标准卷积过程示意图



(c) 深度卷积



(d) 点卷积

图 1　标准卷积、深度卷积、点卷积过程示意图

深度可分离卷积则是将标准卷积分解为一个深度卷积和一个点卷积[11]，深度卷积过程实际上是将输入的每个通道各自与其对应的卷积核进行卷积，最后将得到的各个通道对应的卷积结果作为最终的深度卷积结果。实际上，深度卷积的过程完成了输入特征图的过滤，深度卷积过程如图 1(c) 所示，其计算量为:

$$M \times M \times N \times N \times C \tag{2}$$

这里的点卷积则是将深度卷积的结果作为输入，卷积核大小为 1×1，通道数与输入一致。点卷积过程类似标准卷积，实际上是对每个像素点在不同的通道上进行线性组合(信息整合)，且保留了图片的原有平面结构、调控深度。相比于深度卷积，点卷积具有改变通道数的能力，可以完成升维或降维的功能。点卷积过程如图 1(d) 所示，其计算量为:

$$M \times M \times 1 \times 1 \times C \times K = M \times M \times C \times K \tag{3}$$

因此深度可分离卷积总的计算量为:

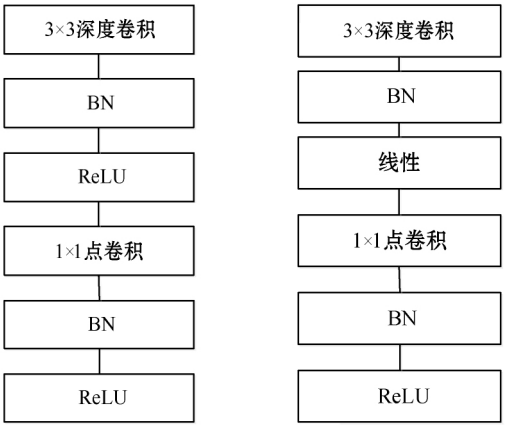
$$M \times M \times N \times N \times C + M \times M \times C \times K \tag{4}$$

深度可分离卷积与传统标准卷积计算量相比:

$$\frac{M \times M \times N \times N \times C + M \times M \times C \times K}{M \times M \times N \times N \times C \times K} = \frac{1}{K} + \frac{1}{N^2} \tag{5}$$

从式(5) 可以看出，深度可分离卷积可有效减少计算量，若网络使用卷积核大小为 3×3，则深度可分离卷积可减少 8 至 9 倍计算量。相比传统标准卷积，这种分解在有效提取特征的同时，精度损失也较小[11]。

在 MobileNet 网络中，为了更好地表现网络的非线性建模能力，同时为了防止梯度消失，减少了参数之间的依存关系，缓解过拟合发生，深度可分离卷积在深度卷积和点卷积后都使用了 ReLU 激活函数。同时，为了防止梯度爆炸，加快模型的收敛速度，提高模型精度，在 ReLU 激活函数前加入 BN 层，如图 2(a) 所示。



(a) MobileNet 中深度可分离卷积层　　(b) 改进的深度可分离卷积层

图 2　MobileNet 及本文深度可分离卷积模型

ReLU 定义如下:

$$f(x) = \max(0, x) = \begin{cases} x & x > 0 \\ 0 & \text{其他} \end{cases} \tag{6}$$

式中: x 表示输出，$f(x)$ 表示输出。

然而 ReLU 也有缺陷，它可能会使神经网络的一部分处于"死亡"的状态。假设网络在前向传导过程中如果有一个很大的梯度使得神经网络的权重更新很大，导致这个神经元对于所有的输入都给出了一个负值，然而这个负值经过 ReLU 后输出变为 0，这个时候流过这个神经元的梯度就永远会变成 0 形式，也就是说这个神经元不可逆转地"死去"了。神经元保持非激活状态，且在后向传导中"杀死"梯度。这样权重无法得到更新，网络无法学习，自然就丢失了信息。而在深度卷积的过程中，由于深度卷积没有改变通道数的能力，其提取的特征是单通道的。而 ReLU 激活函数

在通道数较少的卷积层输出进行操作时，如果出现这种情况就可能导致信息的丢失，所以深度卷积后进行非线性是有害的，甚至可能影响网络的建模能力。为此，文献[12]使用了线性的反转残差网络，然而此网络应用于实际人脸表情识别中效果依然不佳，参数量和运算量依然很大，在安卓手机上测试时，实时性表现也不佳。除此之外，文献[12]中的网络模型允许的输入尺寸与实际人脸表情图像的输入尺寸也不一致，因此 MobileNet 难以满足实际人脸表情识别的需要。为了避免这些现象的发生，本文提出了一种改进的深度可分离卷积层，即在深度卷积后去掉中 ReLU 激活函数而采用线性输出，其余与 MobileNet 中深度可分离卷积层一致，如图 2(b) 所示。线性输出表达如下：

$$f(x) = Wx + b \tag{7}$$

式中：W 表示权重，b 表示偏置，x 表示输入，$f(x)$ 表示输出。

改进后的深度可分离卷积层的计算量与改进前相同，即保留了 MobileNet 网络中深度可分离卷积可减少卷积计算量的优势。同时，改进后的深度可分离卷积层在深度卷积层后采用了线性输出，使得各通道的信息完全保留下来，从而为后续人脸表情识别提供可靠的人脸表情特征。为了验证使用改进后的深度可分离卷积层的有效性，本文在实验部分对使用未改进的深度可分离卷积层的模型进行了对比。

2 改进的 MobileNet 网络模型

本文受到 MobileNet 网络模型启发，结合人脸表情识别的特点，在尽可能在减小网络的计算量并且保持较高的识别率的原则下，设计了一个基于改进深度可分离卷积层输入尺寸为 48 × 48 × 1 的改进 MobileNet 网络模型 M-MobileNet，其网络结构如表 1 所示。

表 1 M-MobileNet 网络结构

Type/Stride		Filter Shape	Input Size
Conv/s2		3 × 3 × 1 × 32	48 × 48 × 1
Conv dw/s1		3 × 3 × 32dw	24 × 24 × 32
Conv/s1		1 × 1 × 32 × 64	24 × 24 × 32
Conv dw/s2		3 × 3 × 64dw	24 × 24 × 64
Conv/s1		1 × 1 × 64 × 128	12 × 12 × 64
Conv dw/s1		3 × 3 × 128dw	12 × 12 × 128
Conv/s1		1 × 1 × 128 × 128	12 × 12 × 128
5 ×	Conv dw/s1	3 × 3 × 128dw	12 × 12 × 128
	Conv/s1	1 × 1 × 128 × 128	12 × 12 × 128

续表 1

Type/Stride	Filter Shape	Input Size
Conv dw/s2	3 × 3 × 128dw	12 × 12 × 128
Conv/s1	1 × 1 × 128 × 256	6 × 6 × 128
Conv dw/s1	3 × 3 × 256dw	6 × 6 × 256
Conv/s1	1 × 1 × 256 × 256	6 × 6 × 256
AvgPool/s1	Pool 6 × 6	6 × 6 × 256
FC/s1	256 × 7	1 × 1 × 256
SVM/s1	Classifier	1 × 1 × 7

在 M-MobileNet 网络中，首先将输入特征图通过一个标准卷积层，然后在通过 10 个改进后的深度可分离卷积层，然后依次经过平均池化层、全连接层提取特征。为了更好地提取特征和使网络快速收敛，每经过一个点卷积层处理后的输出都要经过 BN 层和 ReLU 非线性激活函数处理从而增加非线性表达能力，而在深度卷积层为了尽可能保留信息，则只经过 BN 层，不加入非线性激活函数，采用线性输出，即原有 MobileNet 模型中所有的深度可分离卷积层采用本文改进后的深度可分离卷积层。在分类器设计方面，MobileNet 采用 Softmax 分类器。假设分 7 类，$p_i(i=1,2,\cdots,7)$ 表示 Softmax 层 7 个结点的离散概率，显然 $\sum_{i=1}^{7} p_i = 1$。设 h 为倒数第二层节点的激活，w 为倒数第二层与 SoftMax 层连接的权重，$a_i(i=1,2,\cdots,7)$ 表示对应结点输出，则有：

$$a_i = \sum_k h_k W_{ki} \tag{8}$$

p_i 的计算公式为：

$$p_i = \frac{\exp(a_i)}{\sum_{j=1}^{7} \exp(a_j)} \tag{9}$$

根据式(8)、式(9)求得所有 7 个可能概率，取最大概率对应的类别即为最终预测类别。由于人脸不同表情的类间区分度本身就不高，使用 Softmax 分类器很可能会产生误判，因此在人脸表情识别方面不宜采用 Softmax 分类器。针对此问题，M-MobileNet 采用 L2-SVM 作为分类器。在 L2-SVM 中，对于给定训练数据 (x_n, y_n)，$n=1,2,\cdots,N$，$x_n \in \mathbf{R}^D$，$t_n \in \{-1,1\}$，带有约束性的支持向量机：

$$\min_{w,\xi_n} \frac{1}{2} w^{\mathrm{T}} w + C \sum_{n=1}^{N} \xi_n \tag{10}$$

$$\text{s.t.} \quad w^{\mathrm{T}} x_n t_n \geqslant 1 - \xi_n \quad \forall n$$

$$\xi_n \geqslant 0 \quad \forall n$$

目标函数：

$$\min_{w,\xi_n} \frac{1}{2} w^{\mathrm{T}} w + C \sum_{n=1}^{N} \max(1 - w^{\mathrm{T}} x_n t_n, 0)^2 \quad (11)$$

预测类别为：

$$\arg\max_t (w^{\mathrm{T}} x)\, t \quad (12)$$

式中：w 表示最优超平面法向量，C 表示用来调节错分样本的错误比重，ξ_n 表示松弛因子。

由于 L2-SVM 挖掘不同类别数据点的最大边缘，具有较好的可微可导性，正则化项对错分数据惩罚力度更大[16]，且具有较强的泛化能力以及其对大数据高维特征的分类支持较好[14]，对于人脸表情特征区分较好，所以本文使用 SVM 分类器替代 MobileNet 网络模型中的 Softmax 分类器对目标进行分类。

改进后的网络模型参数及与 MobileNetV1、MobileNetV2 的参数对比如表 2 所示。

表 2　本文模型与 MobileNetV1、MobileNetV2 参数量对比

模型	参数量/M
MobileNetV1[11]	4.2
MobileNetV2[13]	3.4
M-MobileNet	0.4

可以看出，M-MobileNet 网络模型参数较 MobileNetV1 减少 90% 左右，较 MobileNetV2 减少 88% 左右，大大缩减了 MobileNet 网络模型参数。

3　实　验

3.1　实验环境

本文在 PC 端上的实验以 Keras 深度学习框架为基础，以 TensorFlow 框架作为其后端，编程语言使用 Python 3.5，在 Windows 7 64 位操作系统上进行实验。硬件平台为：Intel Core i5-7500 3.4 GHz CPU，8 GB 内存。数据集使用 CK + 数据集（extended Cohn Kanade dataset）[17] 和 KDEF 数据集（The Karolinska Directed Emotional Faces dataset）[18]。实验中采用 Adam 优化器优化损失，epoch 为 100，batch_size 为 32。

本文的移动端的实时性实验在小米 8 手机上进行，CPU 为骁龙 710，内存 6 GB，操作系统为 Andriod 9.0。编程语言为 Java。

3.2　数据集选取

本文在 CK + 数据集上分别进行 6 分类和 7 分类实验，在 KDEF 数据集上进行 7 分类实验。

CK + 数据集包括 123 名年龄在 18 ~ 30 岁之间的 593 个表情序列。其中，带标签的表情序列有 327 个，为了避免引起误会，余下不确定的表情序列不带标签。标签总共有 7 类，包括“快乐”，“悲伤”，“愤怒”，“惊讶”，“恐惧”，“厌恶”和“蔑视”。每个带标签的表情序列仅有一个标签。每个表情序列都是以中性表情开始，以对应峰值表情结束。在 7 分类实验中，为了比较本文模型和目前国际上在 CK + 数据集上识别率处于领先地位的模型的准确率，在本实验中，采用国际上比较通用的数据集选取和结果验证方式，即选用所有带标签的表情序列中的最后三帧，总共得到了 981 幅图像的实验数据集，其表情选取数量分布如表 3 所示。然后将数据集进行交叉验证，国际上常见的交叉验证策略包括 8 折交叉验证、10 折交叉验证等，本文采用 10 折交叉验证策略。同理，在 6 分类实验中，除了数据集选取不同外（6 分类实验中舍弃了蔑视表情），其他步骤与 7 分类实验一致。

表 3　CK + 数据集 7 分类实验样本选取数量分布

表情	愤怒	蔑视	厌恶	恐惧	高兴	伤心	惊讶
数量/幅	135	54	177	75	207	84	249

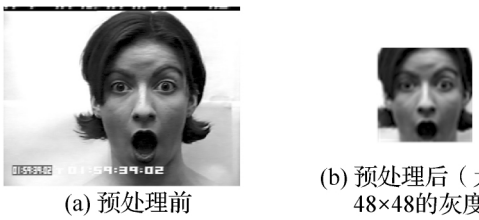
KDEF 数据集包括了 20 ~ 30 岁年龄段的 70 位业余演员（35 位女性和 35 位男性）的 7 类表情图像，共有 4 900 幅，拍摄角度包括正负 90 度、正负 45 度以及正面角度。本文选用正面角度图像进行实验，其数量分布如表 4 所示。同样采用 10 折交叉验证策略。

表 4　KDEF 数据集实验样本选取数量分布

表情	愤怒	蔑视	厌恶	恐惧	高兴	伤心	惊讶
数量/幅	140	140	140	140	140	140	140

3.3　人脸表情图像预处理

原始数据集的原始图像中包含了大量与人脸表情特征无关的冗余信息，且图像较大，因此不适合直接用于网络训练。因此在训练之前，对输入图片进行预处理是必要的。图 3 显示了图像预处理前和处理后的人脸表情图像示例。



(a) 预处理前　　(b) 预处理后（大小为 48×48的灰度图）

图 3　预处理前后的人脸表情图像示例

预处理过程如图 4 所示，首先根据输入图片类型判断是否转换成单通道灰度图，若图片已经是单通道

灰度图，则直接转到一下步，反之则进行转换。然后对上一步的输出图像进行人脸检测，确定人脸区域。最后根据人脸区域对图像进行裁剪，将其裁剪至大小为 48 ×48 的单通道灰度图。

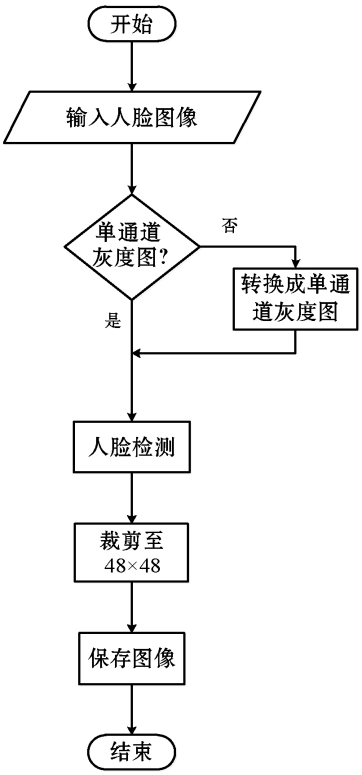


图 4 图像预处理流程图

3.4 实验结果及分析

3.4.1 CK + 数据集 7 分类

本文按照表 1 所设计的网络模型进行训练，将预处理后的图片输入网络，采用 10 折交叉验证策略对网络性能进行评估。表 5 显示了本文模型与其他国际上在 CK + 数据集(7 分类) 准确率上取得领先水平的算法模型的对比结果，同时也对比了 M-MobileNet + Softmax、MobileNet + Softmax、MobileNet + SVM、MobileNetV2 四种网络模型的准确率。

表 5 不同算法模型在 CK + 数据集(7 类) 上的准确率对比

模型	准确率/%
IACNN [19]	95.37
IL-CNN [20]	94.35
DeRF [21]	97.30
2B(N + M) Softmax [22]	97.10
MobileNet + Softmax	98.26
M-MobileNet + Softmax	99.19
MobileNet + SVM	98.88
MobileNetV2 [13]	98.16
M-MobileNet	99.29

可以看出，使用本文网络 M-MobileNet 提取特征后，无论是使用 Softmax 分类器还是 SVM 分类器，其准确率都高于其他模型，说明 M-MobileNet 网络具有良好的特征提取能力。而相比于传统未改进的 MobileNet 网络模型，M-MobileNet + Softmax 网络模型高于未改进前 0.93%，且本文最终模型 M-MobileNet 高于未改进前 0.41%，说明改进的深度可分离卷积层相比未改进的深度可分离卷积层可以有效提高网络的识别率。从 M-MobileNet、MobileNet + Softmax、MobileNetV2 的准确率来看，M-MobileNet 高于 MobileNetV2 网络 1.13%，高于 MobileNet + Softmax 网络 1.03%，说明 M-MobileNet 提高了 MobileNet 在人脸表情上的识别率。从 M-MobileNet + Softmax 网络及 M-MobileNet 网络的准确率来看，使用 SVM 分类器高于使用 Softmax 分类器 0.1%，说明使用 SVM 分类器可以提高模型识别的准确率。

3.4.2 CK + 数据集 6 分类

本文比较了目前国际上对 CK + 数据集作 6 分类准确率较为领先的模型，同样采用 10 折交叉验证策略对网络性能进行评估。表 6 显示了本文模型与其他国际上在 CK + 数据集 6 分类准确率上取得领先水平的算法模型的对比结果，同时也对比了 M-MobileNet + Softmax、MobileNet + Softmax、MobileNet + SVM、MobileNetV2 四种网络模型的准确率。

表 6 不同算法模型在 CK + 数据集(6 类) 上的准确率对比

模型	准确率/%
DLP-CNN [23]	95.78
FN2EN [24]	98.60
DCN + AP [25]	98.90
MobileNet + Softmax	98.38
M-MobileNet + Softmax	98.59
MobileNet + SVM	98.17
MobileNetV2 [13]	96.57
M-MobileNet	99.25

可以看出，无论是使用 Softmax 分类器还是使用 SVM 分类器，使用基于改进的深度可分离卷积的 M-MobileNet 网络的模型的准确率都高于使用未改进的深度可分离卷积的模型，再次证明模型中使用改进的深度可分离卷积层相比使用未改进的深度可分离卷积层的网络模型可以有效提高模型的识别率。从 M-MobileNet、MobileNet + Softmax、MobileNetV2 的准确率

来看，M-MobileNet 高于 MobileNetV2 网络 2.68%，高于 MobileNet + Softmax 网络 0.87%，进一步证明 M-MobileNet 提高了 MobileNet 在人脸表情上的识别率。而虽然在 7 分类实验中使用 SVM 分类器相比使用 Softmax 分类器模型识别的准确率仅提高 0.1%，但是在 6 分类实验中 M-MobileNet 网络与 M-MobileNet + Softmax 网络相比，其准确率提高了 0.66%，明显提高了模型识别的准确率，说明使用 SVM 分类器能提高网络模型对人脸表情的识别准确率。而本文最终网络模型 M-MobileNet 对表情分类的准确率高于表中其他模型。

3.4.3　KDEF 数据集

本文模型在 KDEF 数据集的实验结果及与其他算法模型对比如表 7 所示。

表 7　不同算法模型在 KDEF 数据集上的准确率对比

模型	准确率/%
SCAE [26]	92.52
ResNet-19 [27]	94.49
MobileNet + Softmax	95.39
M-MobileNet + Softmax	95.97
MobileNet + SVM	95.66
MobileNetV2 [13]	95.17
M-MobileNet	96.70

可以看出，本文模型 M-MobileNet 准确率最高，高于表中所有其他模型，说明本文模型具有较好的识别性能。M-MobileNet 网络模型准确率高于 M-MobileNet + Softmax 0.73%，MobileNet + SVM 高于 MobileNet + Softmax 0.27%，说明使用 SVM 分类器相比 Softmax 分类器可以有效提高准确率; M-MobileNet 网络模型的准确率高于 MobileNet + SVM 1.03%，M-MobileNet + Softmax 网络模型的准确率高于 MobileNet + Softmax 0.58%，说明使用改进后的深度可分离卷积模型可以提高深度可分离卷积层的网络模型的准确率，进一步证明使用改进后的深度可分离卷积可以尽可能防止信息丢失。而 M-MobileNet 准确率高于 MobileNetV2 模型 1.53%。

3.4.4　移动端实时性

为了验证本文模型在移动端的实时性能，本文还在移动端上对比了 M-MobileNet 与 MobileNetV1、MobileNetV2 模型在小米 8 手机上的实时性表现，在 CK + 数据集上选取 7 种表情各一幅典型表情的图像进行预测，表 8 显示了各个模型预测 1 000 次后的结果。

表 8　本文模型与 MobileNetV1、MobileNetV2 移动端表现对比

模型	平均预测时间/ms
MobileNetV1 [11]	243
MobileNetV2 [13]	347
M-MobileNet + Softmax	36
M-MobileNet	39

从表 8 可以看出，M-MobileNet 网络无论是使用 Softmax 分类器还是 SVM 分类器，其实时性都比 MobileNetV1、MobileNetV2 好很多，说明本文模型相比 MobileNetV1、MobileNetV2 模型具有更好的实时性，结合表 2 可知，本文模型不仅减少了网络参数，同时还提高了实时性性能。同时从预测时间可以看出，M-MobileNet + Softmax 及 M-MobileNet 二者预测时间都小于 40 ms，可以看出二者都具有较好的实时性，考虑到二者预测时间相差不大，而在准确性实验中 SVM 分类器具有更好的准确率，因此本文最终采用 SVM 分类器。

4　结　语

本文提出了一种改进的 MobileNet 模型 M-MobileNet 用于人脸表情特征提取及分类。在 M-MobileNet 网络模型中，通过使用改进的深度可分离卷积层保证了网络的轻量级特性，解决了深度卷积的输出使用非线性激活函数而可能导致信息丢失的问题，提高了网络的特征提取能力。同时为了有效对表情进行分类，使用 SVM 分类器对人脸表情进行分类，提高了网络对于人脸表情的识别准确率。实验结果表明，本文模型不仅提高了模型的准确率，还实现与现有其他人脸表情识别模型更好的识别性能。在安卓手机上的实验证明，本文模型具有相较于改进前具有更好的实时性。与其他当前优秀算法模型的比较，也看出本文网络模型能够获得更好的识别率，说明其具有良好的应用价值。

参　考　文　献

[1] 张海涛，李美霖，董帅含. 两层级联卷积神经网络的人脸检测 [J]. 中国图象图形学报，2019，24(2)：203－214.

[2] 汤春明，赵红波，张小玉. 基于流形学习 2D-LDLPA 的东亚人脸表情识别算法 [J]. 计算机工程与应用，2018，54(17)：146－150.

[3] 何秀玲，高倩，李洋洋，等. 基于深度学习模型的自发学习表情识别方法研究 [J]. 计算机应用与软件，2019，36(3)：180－186.

[4] Shan C F, Gong S G, Mcowan P W. Facial expression recognition based on Local Binary Patterns: A comprehensive study [J]. Image and Vision Computing, 2009, 27(6): 803 – 816.

[5] Hu Y X, Zeng Z H, Yin L J, et al. Multi-view facial expression recognition [C]//Proceedings of the 8th International Conference on Automatic Face & Gesture Recognition. Piscataway, NJ: IEEE Press, 2008: 1 – 6.

[6] Tariq U, Lin K H, Li Z, et al. Recognizing emotions from an ensemble of features [J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B(Cybernetics), 2012, 42(4): 1017 – 1026.

[7] Krizhevsky A, Sutskever I, Hinton G. ImageNet classification with deep convolutional neural networks [C]//Proceedings of the 25th International Conference on Neural Information Processing Systems, 2012: 1097 – 1105.

[8] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. Computer Science, 2014, 52(3): 1 – 14.

[9] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2015: 1 – 9.

[10] He K, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2016: 770 – 778.

[11] Howard A G, Zhu M L, Chen B, et al. MobileNets: efficientconvolutional networks for mobile vision applications [EB/OL]. (2017 – 04 – 14) [2019 – 06 – 23]. https://arxiv.org/abs/1704.04861.

[12] Sandler M, Howard A, Zhu M L, et al. MobileNetV2: Inverted residuals and linear bottlenecks [EB/OL]. (2019 – 3 – 21) [2019 – 06 – 23]. https://arxiv.org/abs/1801.04381.

[13] Li S, Deng W. Deep facial expression recognition: A survey [EB/OL]. (2018 – 04 – 23) [2019 – 06 – 23]. https://arxiv.org/abs/1804.08348.

[14] 何俊, 刘跃, 李倡洪, 等. 基于改进的深度残差网络的表情识别研究 [J/OL]. 计算机应用研究: 1 – 5 [2019 – 04 – 15]. https://doi.org/10.19734/j.issn.1001 – 3695.2018.10.0846.

[15] Tang Y. Deep learning using linear support vector machines [EB/OL]. (2015 – 06 – 02) [2019 – 04 – 26]. https://arxiv.org/abs/1306.0239.

[16] 鲁新新, 柴岩. L2-SVM 下的短文本情感分类动态 CNN 模型 [J]. 计算机应用与软件, 2018, 35(1): 298-303.

[17] 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 [C]//Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. Piscataway, NJ: IEEE Press, 2010: 94 – 101.

[18] Lundqvist D, Flykt A, Ohman A. The Karolinska directed emotional faces-KDEF, CD ROM from department of clinical neuroscience, psychology section [M]. Stockholm: Karolinska Institutet, 1991.

[19] Meng Z B, Liu P, Cai J, et al. Identity-aware convolutional neural network for facial expression recognition [C]//Proceedings of 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition(FG 2017). Piscataway, NJ: IEEE Press, 2017: 558 – 565.

[20] Cai J, Meng Z B, Khan A S, et al. Island loss for learning discriminative features in facial expression recognition [C]//Proceedings of 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition(FG 2018). Piscataway, NJ: IEEE Press, 2018: 302 – 309.

[21] Yang H Y, Ciftci U, Yin L J. Facial expression recognition by de-expression residue learning [J]. International Journal on Computer Science & Engineering, 2018, 2(5): 2220 – 2224.

[22] Liu X F, Kumar B V K V, You J, et al. Adaptive deep metric learning for identity-aware facial expression recognition [C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops(CVPRW). Piscataway, NJ: IEEE Press, 2017: 522 – 531.

[23] Li S, Deng W H, Du J P. Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild [C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Piscataway, NJ: IEEE Press, 2017: 2584 – 2593.

[24] Ding H, Zhou S K, Chellappa R. FaceNet2ExpNet: Regularizing a deep face recognition net for expression recognition [C]//Proceedings of 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition(FG 2017). Piscataway, NJ: IEEE Press, 2017: 118 – 126.

[25] Zhang Z P, Luo P, Loy C C, et al. From facial expression recognition to interpersonal relation prediction [J]. International Journal of Computer Vision, 2018, 126(5): 550 – 569.

[26] Ruiz A, Elshaw M, Altahhan A, et al. Stacked deep convolutional auto-encoders for emotion recognition from facial expressions [C]//Proceedings of International Joint Conference on Neural Networks. Piscataway, NJ: IEEE Press, 2017: 1586 – 1593.

[27] 杜进, 陈云华, 张灵, 等. 基于改进深度残差网络的低功耗表情识别 [J]. 计算机科学, 2018, 45(9): 303 – 307, 319.