

基于深度残差网络的人脸表情识别

卢官明 朱海锐 郝强 闫静杰

(南京邮电大学通信与信息工程学院, 南京, 210003)

摘要: 针对深度卷积神经网络随着卷积层数增加而导致网络模型难以训练和性能退化等问题, 提出了一种基于深度残差网络的人脸表情识别方法。该方法利用残差学习单元来改善深度卷积神经网络模型训练寻优的过程, 减少模型收敛的时间开销。此外, 为了提高网络模型的泛化能力, 从KDEF和CK+两种表情数据集上选取表情图像样本组成混合数据集用以训练网络。在混合数据集上采用十折(10-fold)交叉验证方法进行了实验, 比较了不同深度的带有残差学习单元的残差网络与不带残差学习单元的常规卷积神经网络的表情识别准确率。当采用74层的深度残差网络时, 可以获得90.79%的平均识别准确率。实验结果表明采用残差学习单元构建的深度残差网络可以解决网络深度和模型收敛性之间的矛盾, 并能提升表情识别的准确率。

关键词: 人脸表情识别; 卷积神经网络; 深度残差网络; 残差学习; 深度学习

中图分类号: TP391 **文献标志码:** A

Facial Expression Recognition Based on Deep Residual Network

Lu Guanming, Zhu Hairui, Hao Qiang, Yan Jingjie

(College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing, 210003, China)

Abstract: The training of deep convolutional neural networks becomes more and more difficult and its performance is degraded with the increase of the number of convolution layers to solve the problem. A facial expression recognition method is presented based on deep residual network. The method uses building blocks for residual learning to improve the training and optimization process of the deep convolutional neural network model and reduce the time cost of the model convergence. In addition, to improve the generalization ability of the network model, a hybrid dataset for training network model is made up of the expression image samples which are selected from the KDEF and CK+ expression datasets. The comparative experiment was conducted with 10-fold cross validation method on the hybrid dataset. In term of expression recognition accuracy, we compared the residual networks with residual learning and the conventional convolution neural networks without residual learning and demonstrated the effect of network depth on the recognition accuracy. The average recognition accuracy of 90.79% is achieved as a 74-layer deep residual network is adopted. The experimental results show that the deep convolutional neural network constructed with building blocks for residual learning can solve the contradiction between the network depth and the model convergence, and can improve the accuracy of expression recognition.

基金项目: 国家自然科学基金(61501249)资助项目; 江苏省重点研发计划(BE2016775)资助项目; 江苏省自然科学基金(BK20150855)资助项目。

收稿日期: 2018-02-27; **修订日期:** 2018-03-14

Key words: facial expression recognition; convolutional neural networks; deep residual network; residual learning; deep learning

引言

在人类交流过程中,人脸表情不仅可以呈现交流者的情绪状态,更能传递交流者的情感信息。如何让机器像人类一样能够理解和表达情感,从而实现人和机器的智能交互,是当前人工智能领域的重要研究课题。人脸表情识别作为情感计算的关键技术,在患者情绪状态分析、疲劳检测、抑郁症的严重程度和疼痛程度评估^[1]等方面有着广阔的应用前景。

在基于传统机器学习的表情识别系统中,面部表情的特征提取与分类器设计是分离的,要想获得好的识别性能,须要提取出具有良好鉴别能力的特征,同时又要设计合适的分类器。国内外学者在这两方面都进行了大量的研究,提出了基于 Gabor 小波变换^[2]、方向梯度直方图(Histogram of oriented gradients, HOG)^[3]、局部二值模式(Local binary pattern, LBP)^[4-5]、流形学习^[6]的特征提取方法,以及基于支持向量机(Support vector machine, SVM)^[7]的分类器等。由于人脸表情具有高维、非刚性、多尺度变化、易受光照和角度影响等特点,基于传统机器学习的表情识别算法的性能依赖于人工提取特征的优劣,人为干扰因素较大,算法的鲁棒性和识别精度有待提高。

近年来,深度卷积神经网络(Deep convolutional neural network, DCNN)在图像分类和识别等领域取得了突破性的进展^[8]。2014年,牛津大学视觉几何组(Visual Geometry group, VGG)提出的 VGGNet 模型^[9]在 ILSVRC 比赛中将 Top-5 错误率降低到了 7.32%。Google 团队提出的 GoogLeNet 模型(又称 Inception 模型)及其后续版本^[10-13],将 ImageNet 数据集上的 Top-5 错误率降低至 4.8%。然而,随着网络深度的增加,训练网络时由于会出现梯度消失或梯度爆炸的问题而使训练变得困难。为了解决上述问题,何恺明等人提出了深度残差网络 ResNet^[14]。该网络采用残差学习单元作为基本组成部分,在原始卷积层外部加入捷径连接(Shortcut connections)支路构成基本残差学习单元,通过顺序叠加残差学习单元成功地缓解了 DCNN 难以训练和性能退化问题,在 2015 年的 ILSVRC 比赛中夺得冠军,将 Top-5 错误率降低至 3.57%。在 ResNet 推出后不久,Google 团队就借鉴了 ResNet 的思想,提出了 Inception V4 和 Inception-ResNet-V2 两种网络模型^[15],并通过融合这两种网络模型,在 ILSVRC 比赛中将 Top-5 错误率降低到了 3.08%。可见,ResNet 及其思想具有很强的推广性。

DCNN 在图像分类和识别领域取得的突破性进展,也为研究表情识别提供了基础与借鉴。卢官明等人^[16]设计了一种用于人脸表情识别的卷积神经网络(CNN)模型,其中包括 3 层卷积层、2 层池化层、1 层全连接层和 1 层 Softmax 层,将表情特征提取和表情分类两个步骤结合在一起,同时采取 Dropout 和数据集扩增策略缓解了网络的过拟合问题,在扩充的 Cohn-Kanade 人脸表情数据集(The Extended Cohn-Kanade database, CK+)上进行了实验,取得了 84.55% 的平均识别准确率。李勇等人^[17]提出了一种跨连接 LeNet-5 卷积神经网络用于人脸表情识别,在 CK+ 表情数据集上取得了 83.74% 的平均识别准确率。然而,目前大多数用于人脸表情识别的 DCNN 模型都是在 CK+ 等表情数据集上训练得到的,由于 CK+ 表情数据集上的图像都是正面拍摄的图像,与实际应用场景中采集的人脸图像有较大的差别,而且每类表情图像的样本数量有限,所以,在 CK+ 表情数据集上训练得到的模型容易产生过拟合现象,会导致模型的泛化能力很弱。为了提高网络模型的泛化能力,本文从 KDEF(Karolinska directed emotional faces)和 CK+ 两种表情数据集上选取表情图像样本组成混合数据集以用于训练网络,提出了一种基于残差学习单元的深度残差网络,探索提高人脸表情识别准确率和鲁棒性的新途径。

1 残差学习的基本思想

从在 ILSVRC 比赛中取得的成绩来看,卷积神经网络在图像分类识别任务上的性能在很大程度上

得益于较深的网络模型。那么,是否可以简单地通过增加卷积层数就能提高卷积神经网络的特征学习能力和分类识别性能?文献[18]指出如果单纯地通过堆叠卷积层来增加网络深度,那么当卷积层数增加到某个值后,分类识别的准确率反而会下降。为了解决随卷积层数的增加而引起的性能退化问题,文献[14]提出一种深度残差学习框架,其基本思想是在构建卷积神经网络时通过加入捷径连接(Shortcut connections)支路构成基本残差学习单元,利用堆叠的非线性卷积层来拟合一个残差映射。假设期望的网络层映射为 $H(x)$,残差映射 $R(x):=H(x)-x$,其中, x 为输入信号,则 $H(x)=R(x)+x$,即期望得到的实际映射 $H(x)$ 不是传统神经网络中的输入的映射,而是输入的映射(即残差映射 $R(x)$)和输入 x 的相加。通过残差学习单元将卷积层对 $H(x)$ 的学习转化为对 $R(x)$ 的学习。这里假设对残差映射 $R(x)$ 进行学习可能会比直接对期望的实际映射 $H(x)$ 进行学习更加容易。

残差学习单元的结构示意图如图1所示。图1中的捷径连接执行一个简单的恒等映射(Identity mapping),既没有引入新的参数,也没有增加计算复杂度。将残差学习单元应用于深度卷积神经网络可以有效缓解网络模型训练时反向传播中的梯度消失问题,进而解决了深层网络难以训练和性能退化的问题。

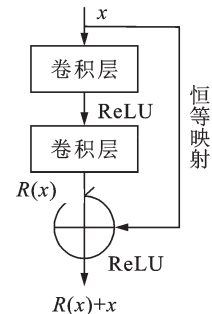


图1 残差学习单元的结构示意图

Fig.1 Building block for residual learning

2 深度残差网络的结构

本文提出的深度残差网络的结构如图2所示,包括依次相连的卷积层、批归一化(Batch normalization)层、修正线性单元(Rectified linear unit, ReLU)激活函数层、残差学习模块1、残差学习模块2、残差学习模块3、全局池化层、全连接层和Softmax分类层。输入图像首先经过4个 3×3 的卷积核进行卷积操作,步长为1,填充为1。全局池化层在 32×32 窗口内进行平均池化操作,步长为1。全连接层的输出神经元数设为7,对应7种不同的表情类别,输出1个7维向量。最后通过Softmax分类器输出分类结果。

残差学习模块1的结构如图3所示,包括若干个依次相连的具有相同结构的残差学习单元A。

残差学习模块2的结构如图4所示,包括1个残差学习单元B和若干个依次相连的具有相同结构的残差学习单元A。

残差学习模块3的结构如图5所示,包括1个残差学习单元B和若干个依次相连的具有相同结构的残差学习单元A。

3个不同残差学习模块中的残差学习单元A的结构都相同,如图6所示,其区别在于残差学习单元A中卷积层所用的卷积核的数量有所不同。

同样,残差学习模块2和残差学习模块3中的残差

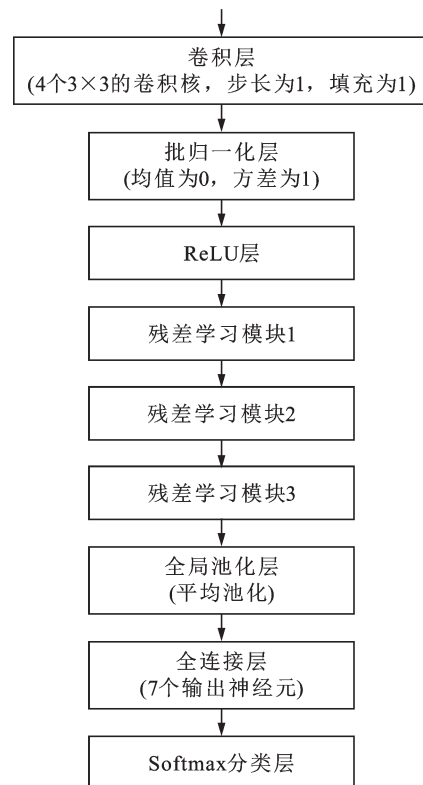


图2 深度残差网络的结构示意图

Fig.2 A block diagram of the deep residual network

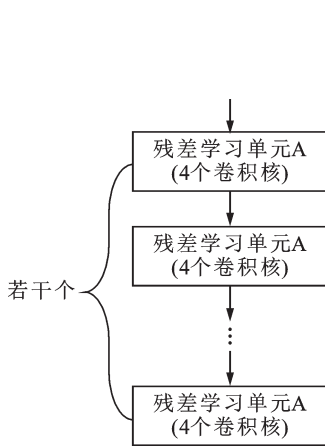


图3 残差学习模块1的结构示意图

Fig.3 A block diagram of the module-1 for residual learning

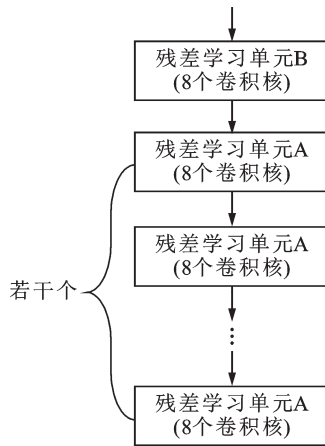


图4 残差学习模块2的结构示意图

Fig.4 A block diagram of the module-2 for residual learning

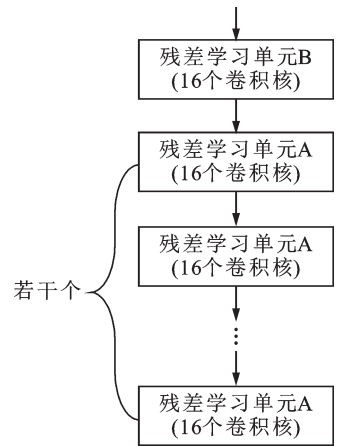


图5 残差学习模块3的结构示意图

Fig.5 A block diagram of the module-3 for residual learning

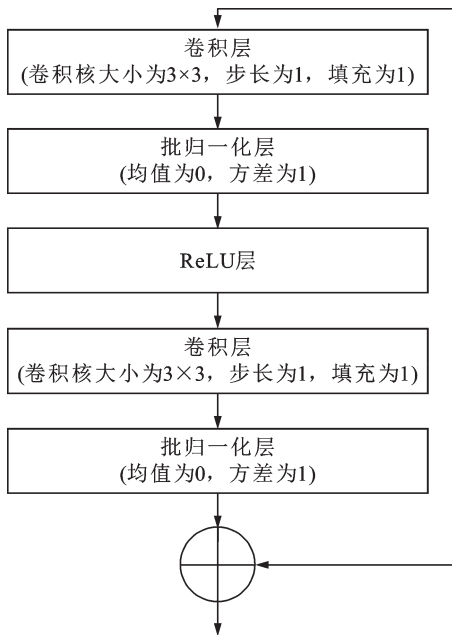


图6 残差学习单元A的结构示意图

Fig.6 A building block-A for residual learning

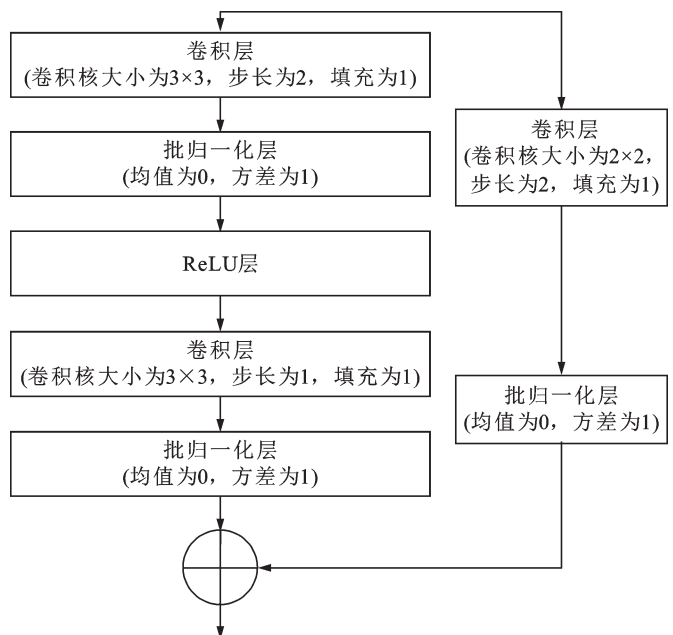


图7 残差学习单元B的结构示意图

Fig.7 A building block-B for residual learning

学习单元B的结构是相同的,如图7所示,其区别在于残差学习单元B中卷积层所用的卷积核的数量有所不同。

例如,当残差学习模块1、残差学习模块2和残差学习模块3分别包含6个残差学习单元时,残差网络的深度为38层;当残差学习模块1、残差学习模块2和残差学习模块3分别包含12个残差学习单元时,残差网络的深度为74层。这里所指的残差网络的深度是指卷积神经网络中不考虑捷径连接支路所包含的卷积层和全连接层的总层数。

3 深度残差网络模型的参数初始化和训练

3.1 网络模型参数的初始化

网络模型参数的初始化是一件很重要的事情。传统的固定方差的高斯分布初始化,是使权重初始值满足给定均值和标准差的高斯分布,但网络层数增加使模型很难收敛。Xavier初始化方法^[19]是为了让网络前向传播和反向传播时每一层方差尽量一致的一种方法,在推导的时候假设激活函数是线性的。而本文的深度残差网络采用ReLU非线性激活函数,因此,优先考虑采用基于ReLU设计的msra初始化方法^[20]。msra初始化方法让每个卷积层的权重初始值满足均值为0、标准差为 $\sqrt{2/N_l}$ 高斯分布,其中 N_l 是第 l 层输入的个数。全连接层中所有权重初始值为0。

3.2 网络模型参数的训练和优化

CNN本质上实现一种输入到输出的映射关系,能够学习大量输入与输出之间的映射关系,不需要任何输入和输出之间的精确数学表达式,只要用已知的模式对CNN加以训练,就可以使网络具有输入输出之间的映射能力。CNN执行的是有监督训练,训练过程分为前向传播和反向传播两个阶段,其中前向传播阶段就是利用隐层提取特征的过程,主要是卷积和池化操作;反向传播阶段采用反向传播(BP)算法传递误差,更新权重参数。

(1) 前向传播阶段。从训练集中随机选取24个小批量(Mini-batch)样本,同时输入到网络,逐层传播下去,最后Softmax分类层输出24个7维向量,向量中的每个元素分别代表输入样本被分到对应表情类别的概率。

(2) 反向传播阶段,也称为误差传播阶段。首先根据交叉熵(Cross-entropy)损失函数计算出损失值,然后使用文献^[21]的算法更新网络中每个可训练的权重参数使得损失值最小化,以优化整个网络。

4 实验结果与分析

本文实验是在开源深度学习框架Caffe上完成的。实验硬件平台为Intel(R) Core(TM) i5-6500 CPU,主频为3.2 GHz,内存为8 GB,操作系统为Ubuntu16.04 LTS版本,搭载Linux内核,同时借助NVIDIA GTX 1080TI GPU进行加速处理。

4.1 表情数据集

本文采用KDEF(The Karolinska Directed Emotional Faces)^[22]和CK+(The extended Cohn-Kanade Dataset)^[23]两个表情数据集。KDEF表情数据集包含了20岁至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,其中有表情类别标签的图像有4 898幅,每类表情图像的数量如表1所示。拍摄角度包括左侧面(Full left profile)、半左侧面(Half left profile)、正面(Straight)、半右侧面(Half right profile)、右侧面(Full right profile)。

CK+表情数据集包括了123位受试者的593个视频序列,其中有327个视频序列的最后一帧标记了表情的标签,每类表情图像的数量如表2所示。

由于CK+和KDEF表情数据集中的表情类别不完全一致,实验中我们从CK+表情数据集中舍弃了类别标签为轻蔑的18帧表情图像,从KDEF和CK+表情数据集中选取了7类表情共5170幅表情图像,其中每类表情图像的数量如表3所示。

表1 KDEF表情数据集中每类表情图像的数量

Tab. 1 The number of images for each expression classification on the KDEF dataset

中性	愤怒	厌恶	害怕	高兴	悲伤	惊讶
Neutral	Angry	Disgusted	Afraid	Happy	Sad	Surprised
700	700	700	700	700	699	699

表2 CK+表情数据集中每类表情图像的数量

Tab. 2 The number of images for each expression classification on the CK+ dataset

愤怒	厌恶	害怕	高兴	悲伤	惊讶	轻蔑
Angry	Disgust	Fear	Happy	Sadness	Surprise	Contempt
45	59	25	69	28	83	18

表3 实验中使用的每类表情图像的数量

Tab. 3 The number of used images for each expression classification in experiment

中性	愤怒	厌恶	害怕	高兴	悲伤	惊讶
700	740	750	720	760	720	780

实验中先将所有图像剪裁成128像素×128像素大小,如图8所示,并对像素值进行归一化操作。为了提高识别结果的可靠性,实验中采用十折(10-fold)交叉验证方法,即将每类表情图像平均分成10份,每次实验使用其中的1份用作测试样本集,其余的9份用作训练样本集,并采用镜像翻转的方式对训练样本进行扩充,将训练样本数扩充1倍,即共有9306幅图像。这样的识别实验重复10次,最后取10次的平均识别率。



图8 实验中使用的7类表情图像的样例

Fig. 8 Samples of seven classifications of facial expression images used in the experiment

4.2 实验结果与分析

为了比较带有残差学习单元的深度残差网络(ResNet)与不带残差学习单元的DCNN以及不同网络深度所具有的表情识别性能,实验对比了20,38,56,74层带有残差学习单元的ResNet与相同深度但不带残差学习单元的DCNN的识别准确率,其结果如图9所示。其中,74层ResNet的表情识别混淆矩阵如表4所示。在对比实验中,所谓的不带残差学习单元的DCNN,是指仅仅把带有残差学习单元的深度残差网络中的捷径连接支路删除后的堆叠DCNN,其他各层的操作都是一样的。

由图9可知,当采用不带残差学习单元的普通堆叠DCNN时,随着网络深度从20层增加到38层,表情识别准确率相应地从87.02%增加到87.20%,提升不明显;但随着网络深度从38层进一步增加到56,74层时,表情识别准确率并不会随着层数的增加而增加,反而从87.20%降低到86.71%、86.23%,出现了识别性能的退化现象,收敛性较差。反之,当采用带有残差学习单元的ResNet时,随着网络深度

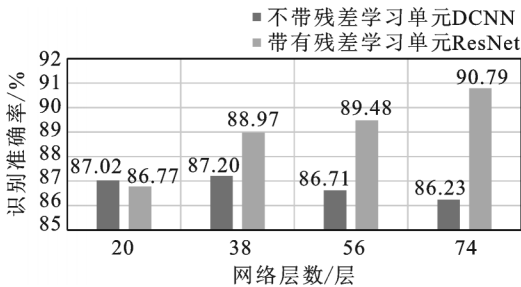


图9 不同结构神经网络的表情识别准确率比较

Fig. 9 Comparison among different neural networks in term of expression recognition accuracy

表4 74层ResNet的表情识别混淆矩阵

Tab. 4 Confusion matrix for ResNet with 74 layers %

表情	中性	愤怒	厌恶	害怕	高兴	悲伤	惊讶
中性	96.14	0.43	0.57	0.71	0.57	1.43	0.14
愤怒	1.08	88.92	5.00	1.89	0.95	1.76	0.41
厌恶	0.80	2.67	90.93	1.60	1.20	2.40	0.40
害怕	1.11	2.78	2.22	82.22	2.64	3.47	5.56
高兴	0.39	0.13	0.26	0.39	97.89	0.39	0.53
悲伤	3.89	3.19	2.92	2.78	0.28	86.39	0.56
惊讶	0.26	0.38	0.26	5.77	0.13	0.51	92.69

从20层增加到38层、56层、74层,表情识别准确率相应地从86.77%增加到88.97%、89.48%、90.79%,没有出现识别性能的退化现象,收敛性较好。实验结果表明采用残差单元构建的深度卷积神经网络可以解决网络深度和模型收敛性之间的矛盾,随着网络深度的加深,能够学习到鉴别能力更强的表情特征,可提升表情识别的准确率和鲁棒性。

由表4可知,当采用74层的ResNet时,中性和高兴这两类表情的识别准确率分别达到96.14%和97.89%,愤怒、厌恶、悲伤和惊讶4类表情的识别准确率也分别达到了88.92%、90.93%、86.39%和92.69%。害怕和惊讶两种表情容易被混淆,惊讶表情被误判为害怕表情的概率为5.77%,而害怕表情被误判为惊讶表情的概率为5.56%。另外,愤怒表情也容易被误判为厌恶表情,误判概率为5.00%。

5 结束语

深度卷积神经网络通过组合多层神经元,提取不同层次的特征,不断迭代组合成更高层次的抽象特征,相比传统的人工设计特征具有更强的特征表达能力和泛化能力。但是,随着卷积神经网络深度的加深,一方面由于在训练网络时会出现梯度消失或梯度爆炸的问题而使训练变得困难,另一方面由于网络模型结构越来越复杂,随之带来的不利因素就是更大的计算量和更高的内存需求,这会极大地影响模型对输入图像的处理速度。针对深度卷积神经网络随着卷积层数增加而导致网络模型难以训练和性能退化等问题,本文提出了一种基于深度残差网络的人脸表情识别方法。实验结果验证了所提出的方法的有效性,但仍然存在有待进一步研究之处。一方面,深度神经网络是一种数据驱动的监督学习模型,需要从海量样本数据中学习数据特征,只有通过大量的有标签样本数据训练后才会获得理想的识别效果;而目前本文所用的表情数据集还不够充足,当网络深度进一步加深时,需要学习的权重参数过多会导致过拟合等问题。另一方面,本文对所提出的深度残差网络采用基于端到端的训练,训练过程复杂,在计算能力有限的实验平台条件下,再增加网络深度有一定难度。然而,有研究表明,基于迁移学习(Transfer learning)的方法可以将从某一任务中学习到的特征迁移到新任务中,比如在ImageNet等大规模数据集上预先训练学习到的特征,在其他很多图像识别任务中也取得了很好的结果。因此,下一步的研究工作将尝试使用预先训练的深度残差网络ResNet-101、ResNet-152作为特征提取器,以解决因训练样本数量不足而导致网络模型出现过拟合问题。

参考文献:

- [1] 彭进业,杨瑞靖,冯晓毅,等.人脸疼痛表情识别综述[J].数据采集与处理,2016,31(1):43-55.
Peng Jinye, Yang Ruijing, Feng Xiaoyi, et al. Survey on facial expression recognition of pain[J]. Journal of Data Acquisition and Processing, 2016, 31(1):43-55.
- [2] Gu W F, Xiang C, Venkatesh Y V, et al. Facial expression recognition using radial encoding of local Gabor features and classifier synthesis[J]. Pattern Recognition, 2012, 45(1):80-91.
- [3] Wang X H, Jin C, Liu W, et al. Feature fusion of HOG and WLD for facial expression recognition[C]// Proceeding of the 6th IEEE/SICE International Symposium on System Integration. [S.l.]:IEEE, 2013: 227-232.
- [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] 卢官明,石婉婉,李旭,等.基于LBP特征和稀疏表示的新生儿疼痛表情识别[J].南京邮电大学学报(自然科学版),2015,35(1):19-25.
Lu Guanming, Shi Wanwan, Li Xu, et al. Recognition method for neonatal pain expression based on LBP feature and sparse representation[J]. Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition), 2015, 35(1):19-25.
- [6] 卢官明,左加阔.基于二维局部保持鉴别分析的特征提取算法[J].南京邮电大学学报(自然科学版),2014,34(5):1-8.
Lu Guanming, Zuo Jiakuo. Feature extraction based on two-dimensional locality preserving discriminant analysis [J]. Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition), 2014, 34(5):1-8.
- [7] 卢官明,郭旻,李晓南,等.基于SVM的新生儿疼痛表情识别[J].南京邮电大学学报(自然科学版),2008,28(6):7-11.
Lu Guanming, Guo Min, Li Xiaonan, et al. Recognition for expression of pain in neonate using support vector machine[J]. Jour-

- nal of Nanjing University of Posts and Telecommunications (Natural Science Edition), 2008, 28(6):7-11.
- [8] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]// 26th Annual Conference on Neural Information Processing Systems 2012. Lake Tahoe, NV, USA: [s. n.], 2012: 1097-1105.
- [9] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL].(2015-04-10) [2018-04-01]. <https://arxiv.org/abs/1409.1556>.
- [10] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]// Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: IEEE, 2015: 1-9.
- [11] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]// Proceedings of the 32nd International Conference on Machine Learning. Lille, France: [s. n.], 2015: 448-456.
- [12] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the Inception architecture for computer vision[C]// Proceedings of the 29th IEEE Computer Society Conference on Computer Vision and Pattern Recognition. USA: [s. n.], 2016: 2818-2826.
- [13] Chollet F. Xception: Deep learning with depthwise separable convolutions[C]// Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: [s. n.], 2017: 1800-1807.
- [14] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]// Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016: 770-778.
- [15] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-ResNet and the impact of residual connections on learning [C]// Proceedings of the 31st AAAI Conference on Artificial Intelligence. [S.l.]: AAAI, 2017: 4278-4284.
- [16] 卢官明, 何嘉利, 闫静杰, 等. 一种用于人脸表情识别的卷积神经网络[J]. 南京邮电大学学报(自然科学版), 2016, 36(1):16-22.
Lu Guanming, He Jiali, Yan Jingjie, et al. Convolutional neural network for facial expression recognition[J]. Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition), 2016, 36(1):16-22.
- [17] 李勇, 林小竹, 蒋梦莹. 基于跨连接 LeNet-5 网络的面部表情识别[J]. 自动化学报, 2018, 44(1):176-182.
Li Yong, Lin Xiaozhu, Jiang Mengying. Facial expression recognition with cross-connect LeNet-5 network[J]. Acta Automatica Sinica, 2018, 44(1):176-182.
- [18] He K M, Sun J. Convolutional neural networks at constrained time cost[C]// Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: IEEE, 2015: 5353-5360.
- [19] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks[C]// Proceedings of the 13th International Conference on Artificial Intelligence and Statistics. Sardinia, Italy: [s. n.], 2010: 249-256.
- [20] He K M, Zhang X Y, Ren S Q, et al. Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification [C]// Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015: 1026-1034.
- [21] Sutskever I, Martens J, Dahl G, et al. On the importance of initialization and momentum in deep learning[C]// Proceedings of the 30th International Conference on Machine Learning. Atlanta, GA, USA: [s. n.], 2013: 2176-2184.
- [22] Lundqvist D, Flykt A, Öhman A. The Karolinska directed emotional faces—KDEF, CD ROM from department of clinical neuroscience, psychology section[M]. Stockholm: Karolinska Institutet, 1991.
- [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[C]// Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. San Francisco, CA, USA: IEEE, 2010: 94-101.

作者简介:



卢官明(1965-),男,博士,教授,博士生导师,研究方向:图像处理与多媒体通信、机器学习、计算机视觉、智能视频分析, E-mail: lugm@njupt.edu.cn.



朱海锐(1993-),男,硕士研究生,研究方向:图像处理、模式识别、机器学习。



郝强(1993-),男,硕士研究生,研究方向:图像处理、模式识别、机器学习。



闫静杰(1983-),男,博士,讲师,研究方向:图像处理、模式识别、情感计算、计算机视觉、智能视频分析。