基于混合注意力机制的表情识别研究*

高　健,林志贤,郭太良
(福州大学 物理与信息工程学院,福建 福州 350116)

摘　要:针对目前传统人脸表情识别算法存在特征提取复杂、表情识别率低等问题,提出一种基于混合注意力机制的ResNet人脸表情识别方法。该方法把通道注意力模块和空间注意力模块组成混合注意力模块,将混合注意力模块嵌入ResNet残差学习分支中。针对CK+人脸表情数据集过小问题,采用数据增强策略扩充数据集。实验结果表明,改进后的ResNet在CK+数据集上表情识别准确率为97.04%,有效提高了表情识别准确率。

关键词:表情识别;ResNet;混合注意力;数据增强
中图分类号:TP18　　**文献标识码**:A　　DOI:10.19358/j.issn.2096-5133.2020.01.011
引用格式:高健,林志贤,郭太良.基于混合注意力机制的表情识别研究[J].信息技术与网络安全,2020,39(1):59-62.

Research on expression recognition based on hybrid attention mechanism

Gao Jian,Lin Zhixian,Guo Tailiang
(College of Physics and Information Engineering,Fuzhou University,Fuzhou 350116,China)

Abstract:The traditional facial expression recognition algorithm has the problems of complex feature extraction,low expression recognition rate and so on, in order to solve these problems,a facial expression recognition method based on the combination of hybrid attention mechanism and ResNet is proposed. The method combines the channel attention module and the spatial attention module into the hybrid attention module,and then embeds the hybrid attention module into the ResNet residual learning branch. For the CK+ facial expression data set is small,the data augmentation strategy is adopted to expand the data set. The experimental results show that the expression recognition accuracy of the improved ResNet is 97.04% on the CK+ data set. This method improves the accuracy of expression recognition.

Key words: expression recognition;ResNet;hybrid attention;data augmentation

0　引言

人脸表情识别作为人机交互的重要组成部分,一直是计算机视觉的研究热点,被广泛应用于公共安全、在线教育、医疗等领域。目前,表情识别的研究工作主要分为传统的人工特征提取和基于深度学习两个方向。人工特征常被用于提取图像的外观特征,包括Gabor、HOG以及局部二进制LBP[1-3]等。但由于人工特征受限于算法的设计,计算复杂,在表情识别中效果不佳,正逐渐被基于深度学习的卷积神经网络所取代。

利用深度学习进行图像识别任务时,通常选择增加卷积神经网络的深度、宽度以及丰富网络感受野的方式来提升网络性能和容量。而在网络中引入注意力机制相比以上三种方式可以使网络重点关注图像细节特征,将原先的平均分配资源变成根据关注对象的重要程度进行重新分配,对模型中不同部分赋予权重,从中提取关键特征信息。文献[4]提出SENet网络结构,采用压缩和激励模块(Squeeze-and-Excitation block,SE),对重要通道特征进行强化从而提升识别率。文献[5]提出瓶颈注意力模块,可与任何前向传播神经网络结合。文献[6]提出一种卷积注意力模块(Convolutional Block Attention Module,CBAM),结合了空间注意力和通道注意力,相比SENet[4]只包含通道注意力识别效果更佳。

本文提出一种基于混合注意力机制的人脸表情识别方法,设计两个轻量级的通道注意力模块和空间注意力模块,以顺序排列的方式组成混合注意力模块,并将其嵌入ResNet中,使网络模型提取更

* 基金项目:国家重点研发计划课题(2016YFB0401503);福建省科技重大专项(2014HZ0003-1);广东省科技重大专项(2016B090906001);广东省光信息材料与技术重点实验室开放基金资助项目(2017B030301007)

多表情细节特征,以提升表情识别率。

1 残差网络

2015 年,何凯明等人提出的 ResNet[7] 在 ImageNet 竞赛中刷新了多项纪录,ResNet 因其简单有效的结构在深度学习发展史中占据重要位置。随着网络深度越深,CNN 能提取到不同层次的特征越丰富,但简单地加深网络深度,将会导致训练过程中出现网络退化等问题。

残差网络能够解决深度神经网络退化问题是因为其提出了一种短路连接(shortcut connection),增加了恒等映射(identity mapping),如图 1 所示。当网络输入为 x 时,则所学习到的特征是 $F(x)+x$,即单元输入与输出直接相加。再用 ReLU 激活函数激活,不给网络增加额外参数,网络优化难度下降,提高了训练效率。

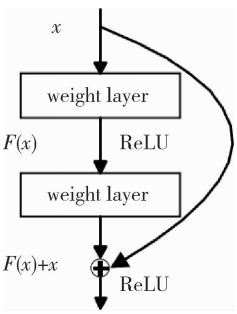

图 1 残差模块

2 基于注意力机制的网络模型

2.1 通道注意力机制

通道注意力机制就是关注不同特征通道重要性,在卷积神经网络中,对于二维图像经过卷积核后会产生图像特征矩阵(H,W,C),其中 H,W 代表图像空间尺度,即高度和宽度,C 代表图像特征通道。通过建模各个特征通道的重要性,给通道特征赋予权重,根据任务需求进行强化或者抑制不同通道。

本文设计的通道注意力模块(channel attention module,Mc)如图 2 所示,相比 SENet[4] 中所使用的平均池化层,本文所使用的最大池化层能保留更多的人脸纹理特征,池化层之后加入 1×1 的卷积层进行降维操作,降低通道数量,在输出处也放置一个 1×1 的卷积核,进行升维,利用降维和升维操作实现通道间信息交换。再经过 Sigmoid 激活函数得到通道注意力结果。

通道注意力模块定义如式(1)所示:

$$Mc(F) = \sigma(f(MaxPool(F))) \quad (1)$$

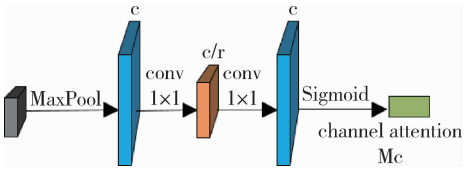

图 2 通道注意力模块

2.2 空间注意力机制

空间注意力机制就是关注特征空间位置的重要程度,为输出特征图生成空间注意力权重,依据特征权重进行强化或者抑制不同空间位置特征。本文设计的空间注意力模块(spatial attention module,Ms)如图 3 所示,将通道注意力模块输出的特征图作为空间注意力模块的输入,采用平均池化和最大池化对输入的特征图进行通道压缩,然后进行拼接操作并采用 3×3 的空洞卷积(dilation = 2)[8] 提取感受野,空洞卷积相比标准卷积在不引入额外参数的情况下,可以扩大卷积的感受野,捕获多尺度信息。最后经过 Sigmiod 激活函数生成空间注意力特征图。

空间注意力模块定义如式(2)所示:

$$Ms(F) = \sigma(f^{3\times3}(concat[MaxPool(F);AvgPool(F)])) \quad (2)$$

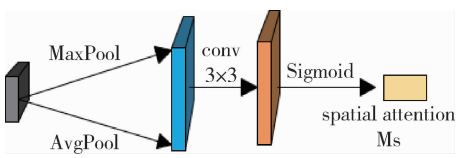

图 3 空间注意力模块

2.3 基于混合注意力机制的 ResNet

本文采用基于混合注意力机制的 ResNet34 模型进行表情识别研究,通过融合通道注意力和空间注意力来提取人脸表情特征,通道注意力模型关注各通道间特征信息,空间注意力模型关注通道内的局部位置信息。借鉴 CBAM[6] 卷积模块的排列方式将通道注意力模块和空间注意力模块顺序排列,设计成混合注意力模块。

本文采用的网络模型如图 4 所示,ResNet34 共包含 5 个卷积部分(conv1 ~ conv5),在其 conv3、conv4、conv5 卷积部分的残差单元后添加混合注意力模块,输入图像特征分别经过通道注意力模块和空间注意力模块,获得通道特征权重 Wc 和空间特征权重 Ws,对得到的重要特征进行强化,最后对经过混合注意力模块的图像特征进行短路连接,得到最终输出特征。

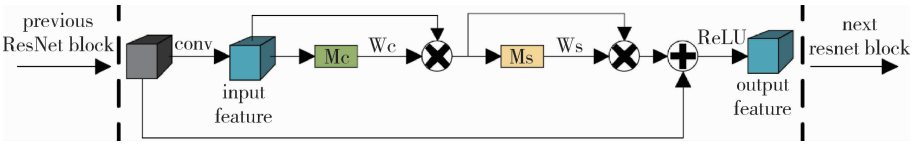

图 4　混合注意力模块的 ResNet

3　实验

3.1　表情数据集与预处理

本实验采用 CK + [9] 人脸表情数据集，CK + 数据集是在 Cohn-Kanade Dataset 的基础上扩展而来，包括 123 位志愿者的 593 个视频序列，本次实验按帧截取每个表情视频序列各 5 张图片作为训练数据，一共 1 635 张图片，分为 7 类表情，包括愤怒、蔑视、厌恶、恐惧、高兴、悲伤和惊讶，CK + 数据集中各表情样例如图 5 所示。作为在理想条件下拍摄的图片，CK + 数据集质量严谨可靠。

图 5　CK + 数据集 7 类表情样例

由于采集到的 CK + 人脸表情数量较少，本文实验采用数据增强策略。对 CK + 人脸表情图像进行随机裁剪、小角度偏转以及增强对比度等操作，扩充人脸表情数据库。CK + 数据集经过数据增强操作后共有 8 000 张表情图像，各类表情图像在 1 000 张左右，通过数据增强可以有效降低目标分类错误，防止训练过程出现过拟合现象。

3.2　实验参数

本文实验均在 PyTorch 深度学习框架上实现，实验硬件平台处理器为 Intel Core i7-8700，内存 16 GB，显卡为 NVIDIA GeForce RTX 2070 8G。

实验采用带有动量优化的随机梯度下降（Stochastic Gradient Descent，SGD）算法对模型参数进行更新，动量设为 0.9。一共训练 200 个周期，训练批次样本为 32，初始学习速率设为 0.01，采用动态衰减学习率，分类函数采用 Softmax。实验采用十折交叉验证法进行训练和测试，即将表情数据分成 10 份，其中 9 份为训练集，1 份为测试集。

3.3　实验结果与分析

为了研究通道注意力、空间注意力和混合注意力对于人脸表情识别的有效性，本文进行了 4 种对比实验，分别是无嵌入注意力模块的 ResNet；单独嵌入通道注意力模块的 ResNet，记作 Mc-ResNet；单独嵌入空间注意力模块的 ResNet，记作 Ms-ResNet；以及嵌入混合注意力模块的 ResNet，记作 Mcs-ResNet。表 1 是以上 4 种网络在 CK + 数据集中的实验结果。

表 1　CK + 数据集实验结果　（%）

	愤怒	蔑视	厌恶	恐惧	高兴	悲伤	惊讶	总计
ResNet	95.98	85.12	95.67	89.05	100	87.73	99.23	94.87
Mc-ResNet	97.94	87.08	97.33	91.16	100	89.52	100	96.35
Ms-ResNet	96.58	86.69	97.63	90.51	100	88.75	99.86	95.91
Mcs-ResNet	98.32	88.29	98.18	91.75	100	91.50	100	97.04

在 CK + 数据集上，ResNet 的表情识别准确率均低于其他三种嵌入注意力模块的神经网络模型，其识别准确率为 94.87%。将通道注意力模块和空间注意力模块分别嵌入 ResNet，表情识别准确率分别达到 96.35%、95.91%，相较于未嵌入注意力模块的 ResNet 提升了 1.48% 和 1.04%，说明在人脸表情识别任务中引入注意力机制有助于表情关键特征的提取，能够有效地提升表情识别准确率。本文提出的基于混合注意力机制的 ResNet 在 CK + 数据集上识别准确率为 97.04%。将通道注意力和空间注意力同时引入卷积神经网络中，相比单独引入一种注意力，混合注意力在表情识别任务中取得的识别准确率最高，与未嵌入注意力模块的网络相比提升了 2.17%。

图 6 是未嵌入注意力模块 ResNet 和嵌入混合注意力模块 Mcs-ResNet 在训练过程中表情识别准确率的变化曲线。由图可知，未嵌入注意力模块的 ResNet 在训练约 70 个 epoch 后开始收敛，而嵌入混

合注意力模块后的 Mcs-ResNet 于 50 个 epoch 后开始收敛。由图可知,本文采用混合注意力模块的 Mcs-ResNet 由于可以更快提取关键表情特征,所以训练效率更高。

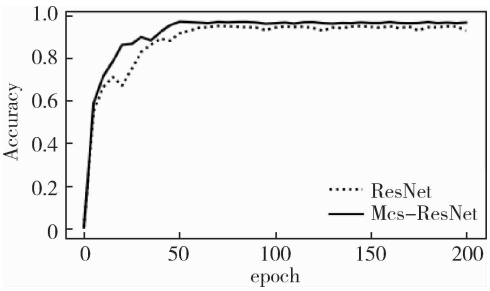

图 6 ResNet 和混合注意力 Mcs-ResNet 的准确率

以上实验结论验证了本文提出的基于混合注意力机制 ResNet 在人脸表情识别任务上的有效性。将本文实验结果与其他方法在 CK + 表情数据集上的实验结果进行对比,结果如表 2 所示,说明本文方法相比其他人脸表情识别方法拥有一定优势。

表 2 本文方法与其他方法准确率对比

方法	准确率/%
CFER[10]	94.87
Island loss[11]	95.48
Lopes[12]	95.75
CDMML[13]	96.60
本文方法	97.04

4 结论

本文提出一种基于混合注意力机制的 ResNet 用于表情识别研究,其中混合注意力模块由通道注意力模块和空间注意力模块组合而成。引入混合注意力机制后的卷积神经网络,将同时关注网络的通道维度信息和空间位置信息,使网络在训练过程中重点关注表情细节特征而忽略其他无关信息。本文方法在 CK + 表情数据集上识别准确率为 97.04%,相比未引入注意力机制的网络模型提高 2.17%,且训练速度更快。后续研究将继续优化注意力模块结构,提高表情识别准确率。

参考文献

[1] 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.

[2] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]. IEEE Computer Society Conference on Computer Vision & Pattern Recognition. IEEE Computer Society, 2005:886-893.

[3] HE J, CAI J F, FANG L Z, et al. A method of facial expression recognition based on LBP fusion of key expressions areas[C]. Control and Decision Conference. IEEE, 2015: 4200-4204.

[4] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks [C]. Computer Vision and Pattern Recognition. IEEE, 2018:7132-7141.

[5] PARK J, WOO S, LEE J Y, et al. Bam: bottleneck attention module[J]. arXiv preprint arXiv:1807.06514, 2018.

[6] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]. Proceedings of the European Conference on Computer Vision (ECCV). Springer, Cham, 2018:3-19.

[7] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. Computer Vision and Pattern Recognition. IEEE, 2016:770-778.

[8] YU F, KOLTUM V. Multi-scale context aggregation by dilated convolutions[C]. Proceedings of International Conference on Learning Representations. Puerto Rico: IEEE Press, 2016:397-410.

[9] 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]. Computer Vision and Pattern Recognition. IEEE, 2010:94-101.

[10] SUN Y, WEN G. Cognitive facial expression recognition with constrained dimensionality reduction[J]. Neurocomputing, 2017, 230:397-408.

[11] 曾逸琪,关胜晓. 一种基于隔离损失函数的人脸表情识别方法[J]. 信息技术与网络安全,2018,37(6):80-84.

[12] LOPES A T, AGUIAR E D, SOUZA A F D, et al. Facial expression recognition with convolutional neural networks: coping with few data and the training sample order[J]. Pattern Recognition, 2016, 61:610 -628.

[13] YAN H. Collaborative discriminative multi-metric learning for facial expression recognition in video [J]. Pattern Recognition, 2018, 75:33-40.

(收稿日期:2019-11-11)

作者简介:

高健(1994 -),男,硕士研究生,主要研究方向:深度学习、图像处理。

林志贤(1975 -),通信作者,男,博士,教授,博士生导师,主要研究方向:信息显示技术、平板显示驱动系统及图像处理。E-mail:lzx2005000@163.com。

郭太良(1963 -),男,硕士,研究员,博士生导师,主要研究方向:信息显示技术。

版权声明