

基于特征融合的注意力双线性池细粒度表情识别

刘力源 张利锋 柳婵娟 贾世祥

(鲁东大学 信息与电气工程学院, 山东 烟台 264039)

摘要: 面部表情的细微差异可能传达出截然不同的情绪,使得表情识别成为一项颇具挑战的任务. 本文提出一种基于特征融合的注意力双线性池模型 FFABP,用于捕获、融合二阶局部特征和一阶全局特征,实现细粒度特征与粗粒度特征的融合. 该模型在双线性池模型的基础上引入注意力机制,利用双线性池模型的高维空间表示捕获表情之间的细微局部差别,借助注意力机制突出特征图中重要空间位置的作用;并使用自注意力模型来学习不同区域特征之间的依赖关系,从而获得图像的全局几何特征. 实验表明,与现有的其它模型相比,本文模型在数据集 FER2013 和 CK+ 上能够获得较好的性能表现和鲁棒性.

关键词: 特征融合; 注意力机制; 细粒度表情识别

中图分类号: TP391.41 **文献标志码:** A **文章编号:** 1673-8020(2020)02-0130-07

面部表情是反映情绪信号和心理活动的重要表现形式,由此衍生而来的面部表情识别技术在智能导学、学习状态检测和情感分析^[1]等方面有着广泛的应用,是计算机视觉、情感计算等领域的研究热点. 但近年来已有的研究集中在特定受控环境下具有给定显著面部特征和预先标记的表情分类场景,无法应对实际应用中复杂多变的表情识别需求,而且表情识别往往需要检测人脸表情细微的差异来揭示其内心活动.

一般来说,面部表情识别分为图像预处理、特征提取和分类 3 个步骤,特征提取是表情识别中的关键步骤,所提取特征的完整性和鲁棒性直接影响分类识别的准确性. 传统机器学习的特征提取方法主要包含几何特征提取和纹理特征提取. 几何特征提取的重点是在关键特征中定位关键点并建立关键点之间的几何关系. Ghimire 等^[2]提出从点、线和三角形中提取几何特征,并使用 Ada-Boost 算法选择最有区分度的几何特征. 纹理特征提取基于空间邻域的灰度分布. Gabor 小波变换和局部二进制模式(LBP)是纹理特征提取的经典方法,但是纹理特征提取主要用于局部区域,缺乏全局信息容易导致难以获得纹理尺度间像素的遗

传或相关性.

近年来,卷积神经网络在特征提取方面显示出显著优势,因为它可以保持几何位置的相对不变性;同时,随着网络深度的增加,该模型可以提取图像的高级语义信息,有效缩小底层特征与高级语义之间的差距. Li 等^[3]提出了一种新的深度保留卷积神经网络(DLP-CNN),其目的在于通过保持局部封闭性,同时最大化类间散度来增强深度特征的判别能力. Liu 等^[4]尝试通过学习网络选择对面部外观/形状变化有影响的特征,同时增强分类器以提高判别能力.

将传统机器学习特征提取方法与神经网络相结合以融合不同维度的多个特征已成为热门话题^[5]. Fan 等^[6]提出了一个框架,将使用卷积神经网络学习到的判别特征与包含形状和外观特征的手工特征相结合,以进一步提高识别的稳健性和准确性. Li 等^[7]将中央对称局部二进制模式(CS-LBP)与深度置信网络(DBN)融合,以利用局部和全局特征提取功能.

然而,多特征融合往往会导致维数灾难. 目前,研究人员较少关注浅层特征和深层特征的融合,细微的、小分辨率的面部表情需要深层特征以

收稿日期: 2019-12-31; 修回日期: 2020-02-27

基金项目: 山东省自然科学基金(ZR2016FM23); 山东省重点研究与发展计划(2017GHY215009)

第一作者简介: 刘力源(1994—),女,山东烟台人,硕士研究生,研究方向为非约束环境下人脸表情识别. E-mail: 2230588946@qq.com

通信作者简介: 张利锋(1977—),男,宁夏贺兰人,讲师,博士,研究方向为深度学习、分布式计算. E-mail: lifengzhang@ldu.edu.cn

增强识别能力. 基于上述认识, 本文利用引入了注意力机制的双线性池模型计算二阶局部统计信息, 用于捕获局部表情的细微特征; 并利用自注意力模型捕获全局特征之间的相关性, 将二阶局部特征和一阶全局几何特征相结合, 以提高细粒度表情识别的识别率.

1 表情识别模型

面部表情识别是细粒度分类领域的一个分支, 因此可以从细粒度分类研究中得到启发, 借助细粒度图像的相关研究思路进行面部微表情识别问题的研究. 考虑到特征空间的高阶统计量有助于局部特征的捕获, 可以用来捕捉人脸表情的细微变化. 因此, 在通过 CNN 提取特征之后, FFABP 尝试通过在每个位置使用外积操作, 将来自不同通道的元素相乘以捕获特征通道之间的成对相关性. 然后, 为了识别和建模特定位置的贡献的差异性, 使用注意力机制显式建模位置权重来自适应地重新校准特定位置的特征响应, 从而使网络能

够提高其对具有细微区分的区域特征的敏感性. 对于全局特征, 由于卷积运算仅考虑了位置的局部邻域, 缺乏对任意区域特征之间相关性的捕获, FFABP 采用自注意力模型在细粒度面部识别中计算局部面部区域的长期依赖性和相关性. FFABP 模型实现如图 1 所示.

FFABP 模型的数据处理流程如下: 首先, 对图像进行预处理, 通过 crop10 方法扩展数据集, 并将其输入残差网络(Resnet18)以生成潜在空间表示. 其次, 将 Resnet18 提取的特征, 分别送入两条支路: 上支路是借助双线性层对潜在空间的高阶统计量建模, 添加注意力机制来跟踪该信息, 以获取不同位置的重要性, 并对加权后的特征按空间维度进行求和池化操作, 获得二阶局部特征; 下支路借助自注意力机制, 来捕获不同局部区域特征之间的相关性, 得到一阶全局特征. 再次, 将这两部分特征采用连接方式融合, 经过 dropout 层以消除减弱神经元节点间的联合适应性, 增强模型泛化性. 最后, 送入全连接层, 得到分类为某一种表情的分类结果.

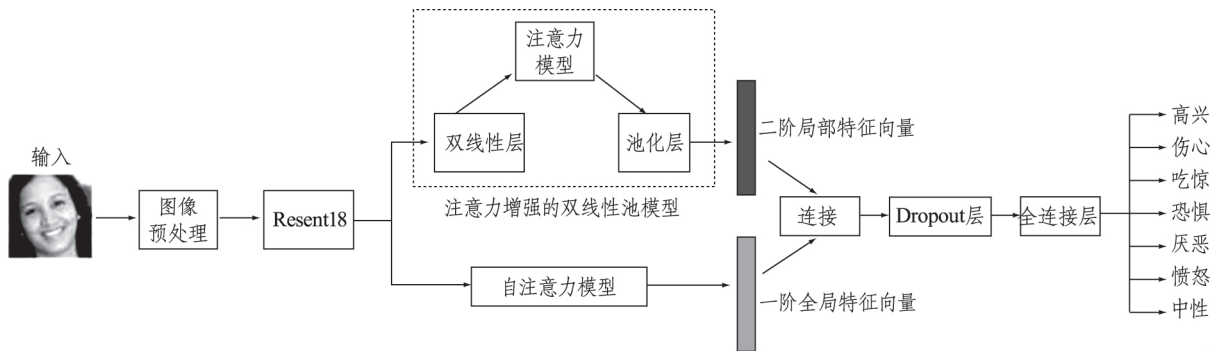


图 1 网络结构图

Fig. 1 Network pipeline structure

1.1 注意力增强的双线性池模型

双线性池可以在变换不变性模式下对局部成对特征交互进行建模, 这已被证明是解决细粒度识别任务的有力工具, 它适用于对成对因素的影响力建模, 例如样式和纹理的影响或脸部上的姿态和光流的影响. 双线性操作作为特征编码层, 在卷积层之后使用可提高多个视觉任务的性能.

双线性池模型由特征提取器组成, 它在特征空间的每个位置的不同通道上执行外积运算, 并将其合并以获取特征空间的描述符; 然而在双线性池操作中忽略了特征图不同空间位置贡献的信息量的差异. 为了缓解这一问题, FFABP 在双线性

池模型的基础上引入注意力机制, 对双线性特征对应特征图的各个空间位置进行加权, 突出图像中重要信息所在区域的作用. 注意力增强的双线性池模型的完整流程如图 2 所示.

为了计算面部表情的双线性特征, 将 Resnet18 最后一层的特征层作为双线性层的输入. 令 $\chi \in \mathbf{R}^{h \times w \times m}$ 表示特征图, 其中 h, w, m 分别表示特征通道的高度、宽度和数量. 描述向量 $x_i \in \mathbf{R}^m$ 表示特征图空间位置 i 的特征向量, 其中空间坐标索引为 $[1 \quad hw]$. 描述矩阵 $X \in \mathbf{R}^{m \times hw}$ 可定义为

$$X = [x_1 \quad x_2 \quad \dots \quad x_{hw}].$$

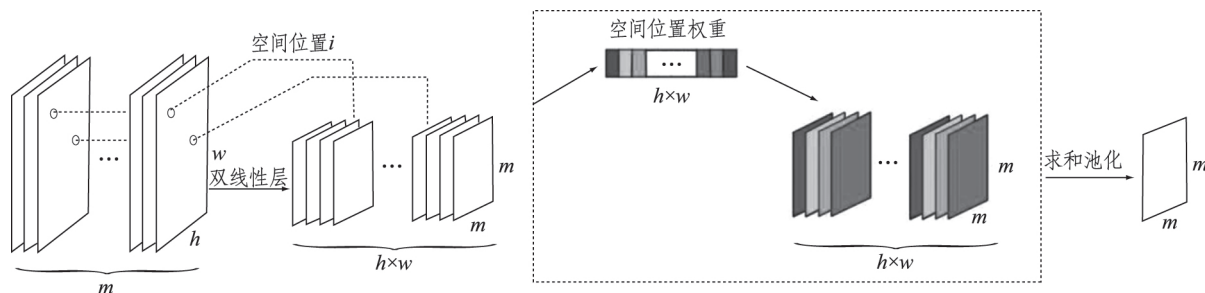


图2 注意力增强的双线性池模型

Fig. 2 Attention enhanced bilinear pooling model

对于每个特征空间内的点,计算其外积 $x_i x_i^T$, 位置 i 的特征矩阵计算公式如下:

$$Z_i = x_i x_i^T. \quad (1)$$

双线性池模型通过特征通道之间的成对相关性获取双线性池化操作, 可以通过 Gram 矩阵 $G \in \mathbf{R}^{m \times m}$ 来计算:

$$G = \frac{1}{hw} \sum_{i=1}^{hw} x_i x_i^T = \frac{1}{hw} X X^T. \quad (2)$$

G 是特征空间的整体表示, 用来捕获特征激活的二阶统计信息, 该信息也与协方差矩阵密切相关. 为了增强特征矩阵 G 的表示能力, FFABP 引入注意力机制来改进双线性池模型. 注意力机

制可以解释为一种将可用信用分配给输入信号中最有用的部分的机制. 具体来说, 就是通过学习的方式来自动获取到每个空间位置的重要程度, 然后依照这个重要程度去提升有用的特征并抑制对当前任务用处不大的特征. 直观上, 某些特定位置的特征(如眼角和嘴角等)表达了区分情感的极其重要的信号, 因此这些位置应比其它位置分配更多权重. 总结来说, 注意力机制是一种「特征重标定」策略, 显式地对位置重要性进行建模, 并通过网络学习以自动获取权重层. 注意力模型如图3所示.

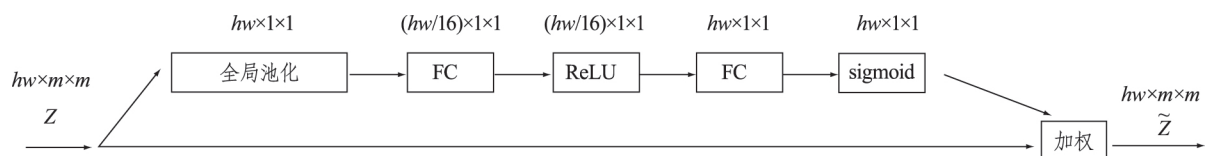


图3 注意力模型

Fig. 3 Attention model

首先将输入特征沿着空间维度进行特征压缩, 将特征压缩成一个一维描述符, 表征在空间位置上响应的分布. 计算公式如下:

$$Z_c' = f_{global}(Z) = \frac{1}{m \times m} \sum_{i=1}^m \sum_{j=1}^m Z_c(i, j). \quad (3)$$

为了计算各空间位置的重要程度, 将一维特征描述符 Z_c' 送入第一个全连接层(FC), 降维至 $1/16$ 以减少计算量; 之后是 ReLU 激活函数层; 再通过第二个全连接层升回到原来的维度; 最后使用 Sigmoid 函数获得每个空间位置的 $0 \sim 1$ 之间归一化的权重, 用来代表该空间位置的重要程度. 借助激励函数可以更好地拟合空间位置间复杂的相关性, 同时极大地减少参数数量和计算量. 在激活函数层前后设置两个 bottleneck 形式的全连接层以实现参数化空间位置选择机制, 可以有效限制模

型复杂性和增强泛化能力.

注意力模型获得的位置权重向量表示为 $\omega = [\omega_1 \ \omega_2 \ \dots \ \omega_{hw}]$, $\omega_i \in \mathbf{R}$, 然后将 Gram 矩阵 G 重写为

$$G' = \frac{1}{hw} \sum_{i=1}^{hw} \omega_i x_i x_i^T. \quad (4)$$

为了寻求压缩计算开销的方法, 本文通过均方根运算对 G' 进行归一化, $G' \rightarrow \sqrt{G'}$.

最后, 通过将 Gram 矩阵扩展为向量来获得双线性特征:

$$s: = \text{vec} G' \in \mathbf{R}^{m^2}.$$

通过注意力机制的作用, 可以让网络累积由注意力模型执行的位置重新校准. 2.4 节将通过实验进一步证明注意力模型的作用.

1.2 自注意力模型

自注意力机制是注意力机制的进一步改进, 可以更好地平衡模型的长期相关性和统计效率, 该机制以所有位置的特征加权和来计算响应, 捕获局部区域变化之间的相关性, 其中权重只以很小的运算成本来计算. 自注意力模型与传统卷积运算相比, 减少了对外部信息的依赖, 更擅长捕捉数据或特征的内部相关性.

自注意力模型如图 4 所示, 该模型将来自 Resnet18 最后一层的特征图 X 作为输入. 首先展开输入特征映射 X 为 $X \in \mathbf{R}^{m \times hw}$, 其中 h, w, m 分别

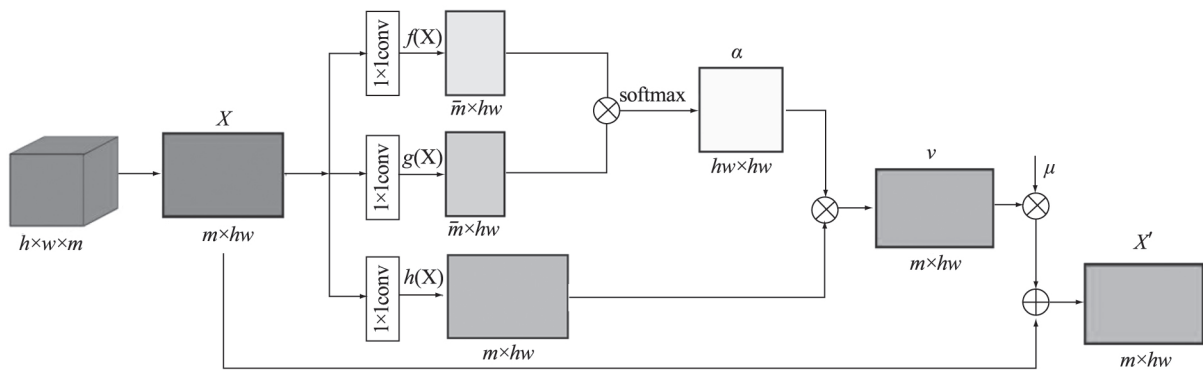


图 4 自注意力模型

Fig. 4 Self-attention model

另对特征图进行 1×1 卷积操作, 其中 $h(X) = W_h X = (h(x_1) \ h(x_2) \ \dots \ h(x_{hw})) \ h(x_i) \in \mathbf{R}^m$. 将此转换后的特征图与矩阵 α 相乘, 可得注意力层的输出为 $v = (v_1 \ v_2 \ \dots \ v_{hw}) \in \mathbf{R}^{m \times hw}$, 且

$$v_i = \sum_{i=1}^{hw} \alpha_i h(X_i). \quad (6)$$

实验中的学习权重矩阵 $W_f \in \mathbf{R}^{\bar{m} \times m}$, $W_g \in \mathbf{R}^{\bar{m} \times m}$ 和 $W_h \in \mathbf{R}^{m \times m}$ 是通过 3 个 1×1 卷积学习得到, 并且为了在计算矩阵 α 的过程中减少计算量, 设置超参数 $\bar{m} = m/8$.

将自注意力层的输出乘上一个比例参数 μ , 然后加回到输入特征图 X 上, 即

$$X' = \mu v + X. \quad (7)$$

最后, 将得到的自注意力特征图 $X' \in \mathbf{R}^{m \times hw}$ 作为全局相关特征.

1.3 线性分类器

本文采用全连接层作为分类器, 分类结果由下式决定:

$$y = (vec W)^T (vec \bar{X}) + b, \quad (8)$$

表示高度、宽度和通道数. 然后通过 2 个 1×1 卷积对特征图进行转换, 将特征图映射到两个嵌入空间 f 和 g 中以提取两个全局注意掩码, 其中 $f(X) = W_f X$ 和 $g(X) = W_g X$ 是线性函数. 然后计算 $f(X)$ 和 $g(X)$ 的点积, 并将其进一步送入 softmax 以提取相关性. 矩阵 α 的计算公式如下:

$$\alpha_{ji} = \frac{\exp(f(X_i)^T g(x_j))}{\sum_{i=1}^{hw} \exp(f(X_i)^T g(x_j))}, \quad (5)$$

α_{ji} 表示第 j 个位置对第 i 个位置的影响程度, 可以看作是第 j 个位置与第 i 个位置的相关程度.

式中: W 为全连接层权重, b 为偏置, \bar{X} 为融合后的特征. 为了减少计算开销, 在全连接层之前设置了 dropout 层(见图 5). 它迫使神经网络单元与其它随机选择的单元协同工作, 从而缓解神经元节点之间适应性的减弱.

2 实验结果分析

2.1 数据集

本文实验样本包括两个部分: FER2013 和 CK + 数据库. FER2013 数据集是 2013 年 Kaggle 比赛的数据, 由 35887 张人脸表情图片组成, 其中 28709 张训练图, 3589 张公开测试图和 3589 张私有测试图, 每一张图都是像素为 48×48 的灰度图. FER2013 数据库中一共包含 7 种表情: 愤怒、厌恶、恐惧、开心、难过、惊讶和中性, 分别对应于数字标签 0 ~ 6. FER2013 数据集的图像分辨率低, 并且不同类别的表情之间的差异非常细微.

CK + 数据库发布于 2010 年, 在 Cohn - Ka-

nade Dataset 的基础上扩展而来. CK + 数据集包含 123 个对象的 327 个被标记的表情图片序列, 共分为正常、生气、蔑视、厌恶、恐惧、开心和伤心 7 种表情. 对于每一个图片序列, 只有最后一帧被

提供了表情标签, 所以共有 327 个图像被标记. 该数据库是在实验室条件下获取的, 较为严谨可靠. CK + 是人脸表情识别中比较标准的一个数据库.

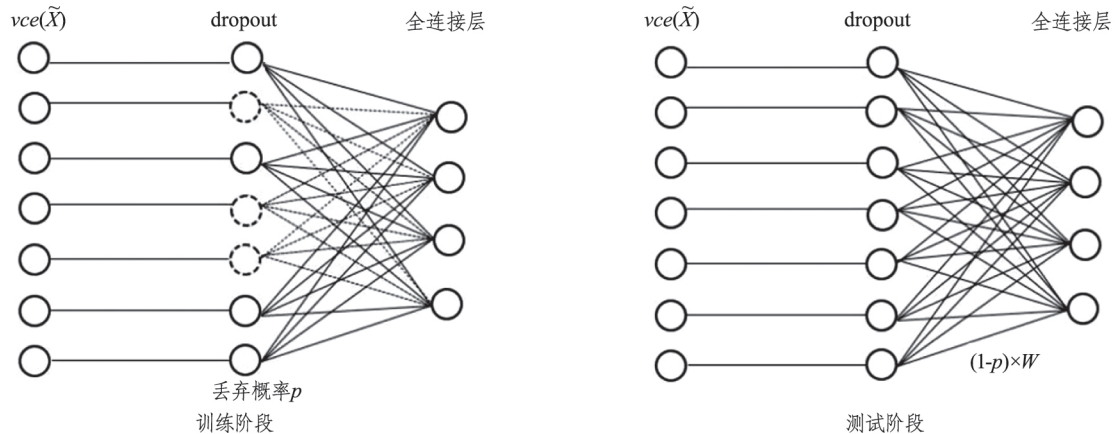


图5 Dropout 层

Fig. 5 Dropout layer

2.2 实验设置

实验主要验证不同数据集上不同模型的细粒度表情识别能力. FFABP 使用 Resnet18 实现潜在特征空间的生成, 并使用 crop10 操作扩充输入数据集, dropout 层的丢弃概率设置为 0.6. 选择 FER2013 和 CK + 作为实验数据集. 其中, 输入图像的尺寸送入网络前被裁剪为 44×44 .

2.3 模型准确性

第一组试验主要验证在 FER2013 数据集上不同模型的网络规模和识别准确率. 表 1 给出了 FFABP 模型与对照方法的模型网络大小和识别准确率对比. FER2013 数据集上人为识别的准确率为 $65\% \pm 5\%$, 从表 1 可以看出与绝大部分的模型相比较人类具有更高的识别准确性, 与其它方法相比, FFABP 模型能在较小的网络尺寸上实现更高的准确率.

表1 FER2013 数据集的网络大小和准确率

Tab.1 Network size and accuracy of the FER2013 dataset

模型	网络大小/Mb	准确率/%
Tang 等 ^[8]	12.0	71.20
Zhang 等 ^[9]	21.3	75.10
Devries 等 ^[10]	12.0	67.21
Guo 等 ^[11]	2.6	71.33
FFABP	11.2	73.63

第二组试验主要验证在 CK + 数据集上不同模型的识别准确率. 表 2 中给出了 CK + 数据集上不同方法的识别精度, 目前已有的模型都取得了 95% 以上的准确率. 如表 2 所示, FFABP 模型与其它模型相比, 准确性有 2% - 5% 的提升, 这说明 FFABP 模型具有更强的表征学习能力, 能有效地提取表情面部图像的深层有效特征.

表2 CK + 数据集的识别准确率

Tab.2 Comparison of the recognition accuracy of CK + dataset

模型	准确率/%
Meng 等 ^[12]	95.37
Liu 等 ^[13]	97.10
Yang 等 ^[14]	97.30
FFABP	99.12

2.4 注意力模型的有效性

注意力模型具有开箱即用的特点, 可与不同的模型协同工作. 为了验证注意力机制在面部表情细粒度识别任务中的实际作用, 本组实验尝试将注意力模型与不同对照算法结合. 表 3 的实验结果表明, 双线性池模型本身的性能优于 Resnet18, 引入注意力模型后, 性能得到了进一步的提升. 在 FER2013 数据集上, 注意力模型将准确率提高了约 10%. 在 CK + 数据集上也有 2% 左右的提升. 这证实了注意力模型的特征融合能力以及对特征建模的有效性.

表 3 FER2013 ,CK + 的识别准确率对比
Tab. 3 Comparison of the recognition accuracy
of FER2013 and CK +

模型	FER2013 准确率/%	CK + 准确率/%
双线性池模型	61.10	94.24
注意力增强的双线性池模型	70.97	96.68
Resnet18	58.57	87.17
Resnet18 + 注意力模型	60.20	90.20

3 结语

考虑到细粒度图像的局部微小变化特征及不同位置的相关性,引入双线性池并结合注意力机制的思想,设计了特征融合的网络,使用自注意力机制获取全局特征空间.通过设计多组对照实验,在 FER2013 和 CK + 数据集上验证了 FFABP 模型的准确性和有效性.结果表明,FFABP 模型能够捕获具有区分能力的特征,显著提高了识别准确性.

参考文献:

- [1] CHEN J Y ,LUO N ,LIU Y Y ,et al. A hybrid intelligence-aided approach to affect-sensitive e-learning [J]. Computing 2016 ,98(1) : 215 - 233.
- [2] GHIMIRE D ,LEE J ,LI Z ,et al. Recognition of facial expressions based on salient geometric features and support vector machines [J]. Multimedia Tools and Applications 2017 ,76(6) : 7921 - 7946.
- [3] LI S ,DENG W. Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition [J]. IEEE Transactions on Image Processing 2019 ,28(1) : 356 - 370.
- [4] LIU K ,ZHANG M M ,PAN Z G ,et al. Facial expression recognition with CNN ensemble [C] // International Conference on Cyberworlds 2016: 163 - 166.
- [5] ZHANG S H ,HE H ,KONG L F. Fusing multi-feature for video occlusion region detection based on graph cut [J]. Acta Optica Sinica 2015 ,35(4) : 0415001.
- [6] FAN X ,TJAHJAD T. Fusing dynamic deep learned features and handcrafted features for facial expression recognition [J]. Journal of Visual Communication and Image Representation 2019 ,65: 102659.
- [7] LI C ,WEI W ,WANG J Z ,et al. Face recognition based on deep belief network combined with center-symmetric local binary pattern [J]. Advanced Multimedia and Ubiquitous Engineering. Singapore Springer 2016 ,393: 277 - 283 .
- [8] TANG Y C. Deep learning using linear support vector machines [C] // International Conference on Machine Learning 2013.
- [9] ZHANG Z P ,LUO P ,LOY C C ,et al. Learning social relation traits from face images [C] // International Conference on Computer Vision 2015: 3631 - 3639.
- [10] DEVRIES T ,BISWARANJAN K ,TAYLOR G W ,et al. Multi-task learning of facial landmarks and expression [C] // Canadian Conference on Computer and Robot Vision 2014: 98 - 103.
- [11] GUO Y N ,TAO D P ,YU J ,et al. Deep neural networks with relativity learning for facial expression recognition [C] // IEEE International Conference on Multimedia and Expo Workshops 2016: 1 - 6.
- [12] MENG Z B ,LIU P ,CAI J ,et al. Identity-aware convolutional neural network for facial expression recognition [C] // IEEE International Conference on Automatic Face and Gesture Recognition 2017: 558 - 565.
- [13] LIU X F ,KUMAR B V ,YOU J ,et al. Adaptive deep metric learning for identity-aware facial expression recognition [C] // IEEE Conference on Computer Vision and Pattern Recognition Workshops 2017: 522 - 531.
- [14] YANG H Y ,CFTCI U ,YIN L J ,et al. Facial expression recognition by de-expression residue learning [C] // IEEE Conference on Computer Vision and Pattern Recognition 2018: 2168 - 2177.
- [15] LIN T Y ,ROYCHOWDHURY A ,MAJI S. Bilinear CNN models for fine-grained visual recognition [C] // International Conference on Computer Vision ,2015: 1449 - 1457.
- [16] KONG S ,FOWLKES C. Low-rank bilinear pooling for fine-grained classification [EB/OL]. [2019 - 11 - 20]. <https://arxiv.org/pdf/1611.05109.pdf>.
- [17] KOBAYASHI T. Low-rank bilinear classification: efficient convex optimization and extensions [J]. International Journal of Computer Vision 2014 ,110(3) : 308 - 327.
- [18] LI K ,JIN Y ,AKRAM M W ,et al. Facial expression recognition with convolutional neural networks via a new face cropping and rotation strategy [J]. The Visual Computer 2020 ,36(2) : 391 - 404.
- [19] 王怡文. 基于卷积神经网络及特征提取的面部表情识别算法 [J]. 电脑知识与技术 (学术版) ,2019 (6) : 203 - 205.
- [20] 王琳琳,刘敬浩,付晓梅. 融合局部特征与深度置信网络的人脸表情识别 [J]. 激光与光电子学进展 , 2018 ,55(1) : 204 - 212.
- [21] GAO Y ,BEIJBOM O ,ZHANG N ,et al. Compact bilinear pooling [C] // IEEE Conference on Computer Vision and Pattern Recognition 2016: 317 - 326.

Fine-grained Expression Recognition of Attention Bilinear Pooling Based on Feature Fusion

LIU Liyuan , ZHANG Lifeng , LIU Chanjuan , JIA Shixiang

(School of Information and Electrical Engineering , Ludong University , Yantai 264039 , China)

Abstract: The subtle differences in facial expressions can convey very different emotions , making facial recognition a challenging task. In this paper , a feature fusion attention bilinear pooling model (FFABP) has been proposed to capture and fuse second-order local features and first-order global features , which realizes the combining of the coarse-grained and fine-grained features. Attention is used by bilinear pooling model. The high-dimensional space representation of the bilinear pooling model is used to capture the subtle local differences between expressions , and the attention is used to highlight the role of important spatial positions in the feature maps. Self-attention is used in FFABP to learn the relationship between different regional features , which can obtain the global geometric features of images. Experiments show that the model can obtain competitive performance and robustness on FER2013 and CK + datasets compared with other existing models.

Keywords: feature fusion; attention; subtle facial expression recognition

(责任编辑 李秀芳)

(上接第 104 页)

Abstract ID: 1673-8020(2020) 02-0097-EA

Elegant Anti - disturbance Control for Stochastic Systems with Multiple Heterogeneous Disturbances Based on Fuzzy Logic Systems

YOU Lihong , WEI Xinjiang

(School of Mathematics and Statistics Science , Ludong University , Yantai 264039 , China)

Abstract: The problem of elegant anti - disturbance control for a class of stochastic systems with multiple heterogeneous disturbances was considered in this paper. The disturbances include white noise and non - harmonic disturbances with unknown smooth nonlinear functions , which can be approximated by fuzzy logic systems. Based on the approximation of the unknown nonlinear function , an adaptive disturbance observer was constructed. Combining disturbance observer - based control with fuzzy control , an elegant anti - disturbance control (EADC) scheme was proposed such that the composite system achieves asymptotically bounded in mean square. Finally , the effectiveness of the proposed scheme was verified by simulation.

Keywords: adaptive disturbance observer; elegant anti - disturbance control; multiple heterogeneous disturbances; fuzzy logic systems

(责任编辑 李秀芳)