

基于残差整流增强卷积神经网络的表情识别

陈斌, 朱晋宁, 东一舟

引用本文:

陈斌, 朱晋宁, 东一舟. 基于残差整流增强卷积神经网络的表情识别[J]. 液晶与显示, 2020, 35(12): 1299–1308.

CHEN Bin, ZHU Jin-ning, DONG Yi-zhou. Expression recognition based on residual rectifier enhanced convolution neural network[J]. *Chinese Journal of Liquid Crystals and Displays*, 2020, 35(12): 1299–1308.

在线阅读 View online: <https://doi.org/10.37188/YJYXS20203512.1299>

您可能感兴趣的其他文章

Articles you may be interested in

多尺度特征融合与极限学习机结合的交通标志识别

Traffic sign recognition based on multi-scale feature fusion and extreme learning machine

液晶与显示. 2020, 35(6): 572–582 <https://doi.org/10.3788/YJYXS20203506.0572>

基于条件生成对抗网络的手写数字识别

Handwritten digit recognition based on conditional generative adversarial network

液晶与显示. 2020, 35(12): 1284–1290 <https://doi.org/10.37188/YJYXS20203512.1284>

基于深度卷积生成对抗网络的图像识别算法

Image recognition algorithms based on deep convolution generative adversarial network

液晶与显示. 2020, 35(4): 383–388 <https://doi.org/10.3788/YJYXS20203504.0383>

基于LeNet-5的卷积神经图像识别算法

Convolutional neural image recognition algorithm based on LeNet-5

液晶与显示. 2020, 35(5): 486–490 <https://doi.org/10.3788/YJYXS20203505.0486>

基于局部路径特征信息神经网络的图像去噪

Image denoising based on local path feature in formation neural network

液晶与显示. 2020, 35(1): 70–79 <https://doi.org/10.3788/YJYXS20203501.0070>

文章编号:1007-2780(2020)12-1299-10

基于残差整流增强卷积神经网络的表情识别

陈 斌*, 朱晋宁, 东一舟

(南京师范大学 信息化建设管理处, 江苏 南京 210046)

摘要:针对受控场景下人脸表情识别率低的问题,提出了一种受控场景下基于残差整流增强卷积神经网络的表情识别算法。该方法以卷积神经网络为原型,在训练模型的过程中,引入残差网络的思想,修正测试集效果与训练集效果之间的差。通过内嵌于卷积层中的激励函数对残差块的线性整流操作,协助表达复杂特征。同时,通过数据增强方法抑制深度神经网络模型在训练过程中过快的拟合,提升其在给定识别任务上的泛化性能,进而提高模型学习效果的鲁棒性。实验中通过将该方法运用于模拟在线授课环境中,最终达到受控场景下有效人脸表情识别的效果。实验结果表明,该方法可以有效对受控情况下的人脸图像输入进行表情分类,最高准确率达到了 91.7%。该研究对人脸表情识别领域及人机交互领域的发展具有积极意义。

关键词:受控场景;卷积神经网络;残差整流;数据增强;激励函数;表情识别

中图分类号:TP39;TH691.9 **文献标识码:**A **doi:**10.37188/YJYXS20203512.1299

Expression recognition based on residual rectifier enhanced convolution neural network

CHEN Bin*, ZHU Jin-ning, DONG Yi-zhou

(Informatization Office of Nanjing Normal University, Nanjing 210046, China)

Abstract: In order to solve the problem of low face recognition rate in controlled scene, an expression recognition algorithm based on residual rectification enhanced convolutional neural network is proposed. This method takes convolutional neural network as the prototype. In the process of training model, the idea of residual network is introduced to correct the difference between the effect of test set and the effect of training set. The linear rectification operation of the residual block by the excitation function embedded in the convolution layer helps to express complex features. At the same time, the data enhancement method is used to suppress the fast fitting of the deep neural network model during the training process, and also to improve its generalization performance on a given recognition task, and then to improve the robustness of the model learning effect. In the experiment, the method

收稿日期:2020-08-02;修订日期:2020-08-31.

基金项目:国家自然科学基金(No.61603193);江苏省高校自然科学研究项目(No.18KJB520027);赛尔网络下一代互联网技术创新项目(No.NGII20170524)

Supported by National Natural Science Foundation of China(No.61603193); Natural Science Research Project of Colleges and Universities in Jiangsu Province(No.18KJB520027); CERNET Innovation Project of Next Generation of Internet Technology(No.NGII20170524)

* 通信联系人, E-mail: njnuchenbin@njnu.edu.cn

is applied to simulate the online teaching environment, and the effect of effective facial expression recognition in controlled scene is achieved. According to the experimental data, this method can effectively classify the facial image input under controlled conditions, and the highest accuracy is up to 91.7%. This research is helpful to the development of facial expression recognition and human-computer interaction.

Key words: controlled scene; convolutional neural network; residual rectification; data enhancement; excitation function; expression recognition

1 引言

表情是人类情感对外表现的重要途径之一。由于表情之中蕴含着丰富的信息,这些信息是人类内心世界的折射,所以使得计算机可以识别表情,从而探知人类的心理状态,使其达到更好的人机交互效果和服务目标^[1]。美国著名心理学家保罗·艾克曼研究了人的不同精神状态与表情的关联关系,发现了基本情感表达要素的 7 种表情分类:惊讶、悲伤、愤怒、恐惧、快乐、厌恶及轻蔑^[2]。随着计算机图像处理技术的蓬勃发展,以及其在人脸识别领域的推进,人脸表情识别研究逐渐成为一个重要方向。

目前计算机表情识别的主要方法涵盖 3 个方向:整体识别法与局部识别法、形变提取法与运动提取法、几何特征法与容貌特征法^[3]。计算机识别表情的过程主要分 3 个步骤:图像预处理,表情特征提取,表情识别(分类)^[4]。计算机识别的表情特征主要以 3 种形式表示:灰度特征、运动特征和频率特征^[5]。表情识别主要分模型设计与训练过程及模型应用与决策过程^[6]。在模型设计与训练阶段,主要有以下模型设计方法:线性分类器、神经网络分类器、支持向量机、隐马尔可夫模型、人脸物理模型建模、模型图像编码等分类模型^[7-9]。针对设计的模型在人脸表情数据集上通常进行有监督的机器学习训练,从而得到最优参数归一化模型的目标结果^[10]。目前较为公认的人脸表情数据集有:Extended Cohn-Kanada (CK+), Kaggle-Facial Expression Recognition 2013 (FER2013), Japanese Female Facial Expression (JAFPE) 等^[11-13]。训练完毕的模型结果将用于应用与决策过程进行表情识别。近年来,在表情识别领域以卷积神经网络(Convolutional Neural

Network, CNN)为基础的分类模型研究得到重视^[14]。Su 等人通过多角度卷积神经网络(Multi-view Convolutional Neural Network, MvCNN)^[15]进行多种姿态下的物体识别,获得较好效果,但训练所需条件苛刻,实践性差;Kaya 等人将图像深层特征与 HOG 及局部 Gabor 二元模式的人工提取特征结合使用,以提高性能^[16],但识别准确率受到影响;Gu 等人对分块图进行 Gabor 变换,并通过径向网络对其进行联合编码,进而通过多分类器级联实现全局分类^[17]。这类方法由于融合高维特征,识别实时性受到约束;Soyel 等人等在卷积过程中引入了加权投票法^[18],Liu 等人基于 Dempster-Shafer 证据理论引入了最优支持度算法实现了多区域决策级融合^[19],但二者的特征提取方法通常均需人工干预,而人工特征提取质量造成了表情特征损失的不确定性,直接影响了模型分类质量的鲁棒性。Pathak 提出基于语义的深度卷积神经网络,通过学习图像特征,生成图像缺失部分的预测图,最终实现缺失图像的修复^[20]。

针对上述问题,本文提出了基于残差整流增强卷积神经网络的表情识别方法。在深度学习结构中, CNN 在图像分类问题中得到了广泛的应用^[21]。近几年来, CNN 在目标分割、识别与检测等领域表现尤为出色^[22]。以 CNN 为原型,在训练模型的过程中,引入残差网络(Residual Net, ResNets)的思想^[23],修正 CNN 测试集效果与训练集效果之间,也即实际观察值与拟合值之间的差。通过内嵌于卷积层中的激励函数(Activation Function)对残差块(Residual Block)的线性整流(Rectified Linear)操作,协助表达复杂特征,并降低由于神经网络层数堆叠后在训练集上学习效果受到的影响。同时通过数据增强的方法抑制深度神经网络模型在训练过程中过快的拟合,提升其在给定识别任务上的泛化性能,进而提高模型学

习效果的鲁棒性。在表情识别领域,算法的结果精准与性能稳定是重要指标。本文致力于结合 CNN 模型的优点,内嵌残差整流及数据增强,在网络层数恒等且复杂性不变的前提下提升模型分类的准确性与鲁棒性。

2 理论基础与模型构架

图 1 是基于残差整流增强卷积神经网络的表情识别模型构架。卷积神经网络是一类包含卷积计算且具有深度结构的前馈神经网络,是深度学习的代表算法之一^[24-25]。卷积神经网络主要包括输入层、隐含层和输出层,隐含层主要由卷积层和池化层构成。通过嵌入各层网络的短连接 (Shortcut Connection) 组成的残差块在卷积过程中的不断带入以抑制训练的退化,以及线性整流函数防止梯度弥散并保持稀疏激活性作用,反复在训练过程中校准直到模型生成,过程中融入了数据增强手段进行训练条件的优化,较好地增加了模型分类效果的准确性,同时增强了模型的鲁棒性。

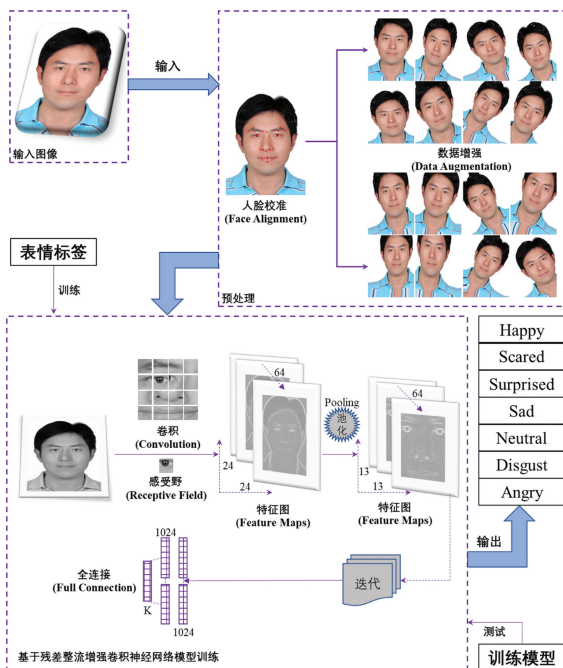


图 1 基于残差整流增强卷积神经网络的表情识别模型构架

Fig.1 Expression recognition model based on residual rectifier enhanced convolution neural network

2.1 残差堆叠伴线性整流卷积神经网络的图像特征处理

CNN 将数据特征的提取、模型的训练,以及分类等操作作为一个端到端的整体识别过程来看待^[26],对于大规模数据集上的应用取得了良好的效果,但对类似于人脸表情识别的小规模数据集,其分类过程中容易出现过拟合现象,从而导致识别效果和稳定性都不理想^[27]。图 2 给出了卷积神经网络对表情图像处理的基本过程。网络深度的增加理论上可以提升分类的精度,但随着网络深度(层数)的不断增加,训练过程越发低效,参数调优越发困难,模型训练衰退情况也越发严重。研究发现残差中蕴含着模型基本关系之间的重要信息^[28],通过回溯观测值的误差,对模型参数进行修正,可以达到理想的抑制训练退化的目的。激活函数通过线性整流的方式将嵌入层的残差修正,并将激活信息传入下一层网络。该处理方法本质上就是将激活函数的非线性因素融入 CNN,补齐模型的线性表达能力,将激活整流后的残差特征保留传递至下一层映射。

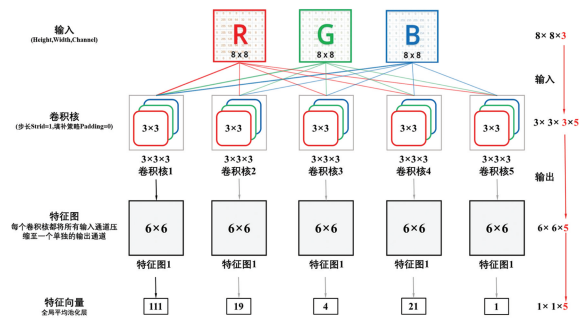


图 2 卷积神经网络图像处理基本过程

Fig.2 Basic process of image processing in convolutional neural network

模型训练过程分为前向计算阶段及方向传播阶段。在两个过程交叠往复中,持续对权值参数保持更新,并将残差激活后迭代进各层各训练阶段,进而使得损失函数结果达到最小。

残差块与激活整流作用过程如图 3 所示, x 表示输入, $F(x)$ 表示残差块在前一层激活函数整流之前的输入,公式(1)

$$F(x) = W_2 \sigma(W_1 x), \quad (1)$$

式中, W_1 和 W_2 分别表示上一层和下一层的权重, σ 表示 ReLU 激活函数,残差整流后的输出结果为 $\sigma(F(x) + x)$ 。卷积层计算方式为:

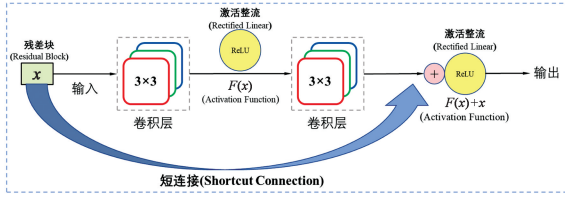


图 3 残差块与激活整流作用过程

Fig.3 Residual block and activation rectification process

$$\text{Conv}^i = F(x * k^i + b^i), \quad (2)$$

式中, Conv 结果为第 i 次卷积与输入图像卷积计算后得到的特征图, F 为激活整流函数 ReLU, k^i 为第 i 个卷积核, b^i 为其对应偏置量。池化层的计算方式为:

$$P^i = F(\alpha^i \text{pooling}(\text{Conv}^i) + b^i), \quad (3)$$

式中, P^i 为池化层中第 i 个图像的输出, α^i 和 b^i 为第 i 个图像的偏置值, $\text{pooling}()$ 为池化函数。合并层的计算方式为:

$$F = \sum_{i=1}^3 m(i), \quad (4)$$

式中, $m(i)$ 为第 i 个池化层的特征图, F 是池化层特征图的和值。全连接层的计算方式为:

$$Fl = F(w \times u + b), \quad (5)$$

式中, u 为特征值, w 为权值, b 为偏置量, F 为激活函数。

2.2 数据增强处理

人脸表情数据集数据样本数量不足已经成为共性问题^[29], 已存在的样本也有着相似度高的问题, 各类表情分类样本分布不均衡容易导致不同表情的分类效果差别较大^[30]。数据增强是目前较好地解决表情样本量不足及分布不均衡问题的方式。图像生成方式在表情识别领域较常用的为生成对抗网络(Generative Adversarial Network, GAN)及其改进方法。但究其根本, 生成方法的基本思路均为二人零和博弈, 即二人的利益之和为零, 一方的所得必然为另一方的所失^[31]。它是一种无监督概率学习方法, 能够对数据的真实分布进行学习并产生与之相关的新数据集。其模型网络由生成模型 G 和判别模型 D 构成, 生成模型用来探寻训练数据的分布情况, 判别模型本质上是一个二分类, 用来判断输入的数据是来源于真实的数据还是生成的数据。生成模型与判别模型之间的关系可以按式(6)表示为:

$$\begin{aligned} \min_G \max_D V(D, G) = & F_{x \sim P_d(x)} [\log D(x)] + \\ & F_{z \sim P_z(z)} [\log(1 - D(z))], \end{aligned} \quad (6)$$

式中, x 为真实数据, 服从分布 $p_d(x)$, z 为随机噪声, 服从高斯分布 $p_z(z)$ 。但传统 GAN 方式存在训练过程不稳定, 生成样本多样性差等问题, 导致生成的图像不可控, 通常的处理方式是给生成网络添加生成约束条件。添加了条件约束的 GAN 本质上是对 GAN 的一种有条件扩展, 其中最具代表性的是带循环约束的生成对抗网络, 也是本文选择使用的表情数据集增强方式。

带循环约束的生成对抗网络由两个生成模型 G_u 和 G_v , 以及两个判别模型 D_u 和 D_v 构成。 $\{u_i\}_{i=1}^{S_1} \in U$ 与 $\{v_j\}_{j=1}^{S_2} \in V$ 为不同的样本数据集。其中 U, V 为源域及目标域样本集容量, u_i 为源域的第 i 张图片, v_j 为目标域中第 j 张图片。 G_u 输出由源域 U 至目标域 V 的对应关系, G_v 输出从目标域 V 到源域 U 的映射关系。判别模型 D_u 用于判定 u 与 G_v 的真假; 判别器 D_v 用于判定 v 与 G_u 的真假关系。带循环约束的生成对抗网络中的损失最主要由两部分内容所组成: 即对抗损失以及循环一致性损失。对于对抗损失, 可以按照式(7)进行映射关系定义:

$$\begin{aligned} \text{Loss}_{\text{GAN}}(G_u, D_v, U, V) = & F_{v \sim P_d(v)} [\log D_v(v)] + \\ & F_{u \sim P_d(u)} [\log(1 - D_v(G_u(u)))], \end{aligned} \quad (7)$$

式(7)用于表示源域样本集容量 U 与目标域样本集容量 V 之间的映射关系所对应的由判别模型 D_v 进行优化所带来的损失代价计算数值。

$$\begin{aligned} \text{Loss}_{\text{GAN}}(G_v, D_u, V, U) = & F_{u \sim P_d(u)} [\log D_u(u)] + \\ & F_{v \sim P_d(v)} [\log(1 - D_u(G_v(v)))], \end{aligned} \quad (8)$$

式(8)表示目标域样本集容量 V 与源域样本集容量 U 的映射关系所需由判别模型 D_u 进行优化的损失。理想情况下对抗损失优化的结果将近似接近目标域的分布, 使得反转后的分布接近源域的分布, 然而实际情况下生成模型通常只能学习到目标域的部分分布, 导致优化结果不能达到预期的目标。然而循环一致性损失的添加可以较好地解决该问题。循环一致性损失函数的定义如式(9)所示:

$$\begin{aligned} \text{Loss}_{\text{Cycle}} = & F_{u \sim P_d(u)} [\|G_v(G_u(u)) - u\|] + \\ & F_{v \sim P_d(v)} [\|G_u(G_v(v)) - v\|], \end{aligned} \quad (9)$$

式中, $\|\dots\|$ 是 L_1 的正则化范式, 用于计算两个分布之间的间距差值。总的损失函数的计算方式

可定义为:

$$\text{Loss}(G_u, G_v, D_u, D_v) = \text{Loss}_{\text{GAN}}(G_u, D_v, U, V) + \text{Loss}_{\text{GAN}}(G_v, D_u, V, U) + \sigma \text{Loss}_{\text{Cycle}}, \quad (10)$$

式中, σ 为超参数,用于协调对抗损失与循环一致性损失函数的权重。

2.3 基于残差整流增强卷积神经网络的表情识别模型

本文基于残差整流增强卷积神经网络的表情识别模型主要分为输入人脸图像预处理(包含数据增强)、表情识别模型构建(包括模型训练和测试),以及表情识别模型的决策,模型整体构架如图 1 所示。首先,对输入的人脸图像进行人脸配准和归一化处理,同时通过数据增强手段对训练数据集进行补充,并对训练集和测试集进行划分;进而对表情信息含量丰富的眼眉区域、唇口区域的关键特征进行提取,过滤低表情信息熵的鼻、脸颊、头发等区域,通过不断卷积迭代及特征图生成和池化后,最终产出训练模型结果;在最终的模型决策阶段,将待决策图片信息提取送入模型进行判断,寻找最优解,获得待测对象表情分类的判定结果,完成表情识别过程。

本文的网络结构以 VGG16 为基础,卷积网络共 12 个卷积层,卷积核大小在整个过程中保持 3×3 不变,最终的输出特征图尺寸为 7×7 ,由于每一步都携带残差信息并伴有线性整流操作,测算出每一步感受比未残差整流处理的情况提升了 3 倍,这对整个卷积过程最终结果有很大的向好影响。本文分别在基础表情数据集 JAFFE 和 FER2013 基础上进行图像生成增强,训练数据 W 包含了 n 个训练样本 w ,划分为 7 类表情数据:惊讶(Surprised)、悲伤(Sad)、愤怒(Angry)、恐惧(Scared)、快乐(Happy)、厌恶(Disgust)和平静(Neutral),经过 SoftMax 分类器进行归一化处理,处理过程表示如下:

$$k(w_i) = \text{SoftMax}(w_i) = e^i / \sum_{x=1}^n e^x, \quad (11)$$

式中, w 代表训练样本, n 代表样本数量, e 代表样本概率。样本模型预测分布 h 与真实样本分布 k 之间的差异关系可以按照式(12)表示:

$$\text{KL}(k \| h) = \sum_{i=1}^{\text{tensor}} k(w_i) \log(k(w_i)) - \sum_{i=1}^{\text{tensor}} k(w_i) \log(h(w_i)), \quad (12)$$

式中, tensor 为输出特征图张量,此处定义其值为

7。在卷积的整个过程中,由于真实分布的 k 值表情信息熵始终保持稳定,所以只需要考虑如何对共享表情信息熵的数值进行计算,方式如下:

$$\begin{aligned} \text{KL}(k \| h) &= -X(k(w)) + \\ &\sum_{i=1}^{\text{tensor}} k(w_i) \log(h(w_i)), \quad (13) \\ X(k, h) &= -\sum_{i=1}^{\text{tensor}} k(w_i) \log(h(w_i)). \end{aligned} \quad (14)$$

同时,对于损失的计算以及各样本分类均衡处理需要添加不同的权值,以降低样本数量的差异化所带来的分类效果的差异化,权值赋予方式定义如下:

$$\text{AW} = v_i \left[-\sum_{i=1}^{\text{tensor}} k(w_i) \log(h(w_i)) \right], \quad (15)$$

式中, v 为权值数据。训练样本类别不均衡是造成不同类别预测结果不对称的最常见原因之一。本文中在训练时为避免正负样本比例失调,一方面通过数据增强补差,另一方面通过增加和调节权值偏置量(超参数)来改善训练稳固性,增强抗干扰性,提升鲁棒性。权值偏置量作用过程可以表示为:

$$\text{AWB} = \sigma_i \left[-(1 - k(w_i)) \right]^\lambda \times \text{AW}, \quad (16)$$

式中, λ 表示表情交叉信息熵, σ 表示调节权值偏置量。

本文通过基于残差整流增强卷积神经网络实现人脸表情检测与识别,通过迭代残差块带入卷积抑制了退化作用,并通过线性整流函数激活残差效果,通过生成高质量候选区域,优化面部表情分类过程。为避免数据集密度不足导致的训练过程过拟合情况,辅助数据增强手段对训练过程进行微调,以获得较优的效果增益。

3 实验结果及分析

本实验所使用的硬件环境为: Intel(R) Core (TM) i5-6200U CPU @ 2.30GHz 处理器, 16GB 内存 (RAM), NVIDIA GeForce GTX 1070Ti GPU 加速。支撑软件及开发软件环境为: Anaconda 4.7.5(内嵌 Python 3.7), PyCharm 2019.3.3(Professional Edition), CNN 所采用的基础框架为 TensorFlow 2.0.0。

3.1 实验数据集

为了证明基于残差整流增强卷积神经网络的



图 4 JAFFE 数据集 7 类表情图像示例

Fig.4 Examples of seven expressions images in JAFFE datasets



图 5 JAFFE 数据集 7 类表情生成图示例

Fig.5 Examples of seven expressions images generation graphs for JAFFE datasets

表情识别方法的有效性,本文分别引入业界具有代表性的 JAFFE 数据集和 FER2013 数据集进行实验评估,文中列出了 JAFFE 数据集的示例图及生成图示例(如图 4 及图 5 所示),FER2013 数据集情况类似。JAFFE 数据集使用日本 ATR (Advanced Telecommunication Research Institute International)专门用于表情识别研究的基本表情数据库。该数据库中包含了 213 幅(每幅图像的分辨率:256×256)日本女性的脸相,每幅图像都有原始的表情定义。表情库中共有 10 个人。JAFFE 数据库均为正面脸相,且把原始图像进行重新调整和修剪,使得眼睛在数据库图像中的位置大致相同,脸部尺寸基本一致,光照均为正面光源,但光照强度有差异。由于此表情数据库完全开放,且表情标定非常标准,所以现在大多数研究表情识别的文章中都使用它来训练与测试。FER2013 的表情数据来源于 ICML(International Conference on Machine Learning)2013 年的挑战赛。FER2013 是自然环境下的人脸表情

数据库,样本总数为 35 887 幅,所有图像来自于互联网。由于样本差异性较大,部分人脸有遮挡和旋转等情况,因此在该数据库上进行表情识别具有一定挑战性。FER2013 数据集图像大小为 48×48 ,由于生成对抗网络输出图像大小为 128×128 ,在使用生成样本时,本文统一将图像样本归一化处理成 48×48 。两个数据集都包括愤怒(Angry)、中性(Neutral)、厌恶(Disgust)、恐惧(Scared)、快乐(Happy)、悲伤(Sad)和惊讶(Surprised)7 种分类表情数据。

3.2 实验结果及分析

本实验基于受控场景下近距离实时摄像头或上传照片的表情分析,相对非受控场景下具备较好排除干扰因素条件。该环境特点是连续(待测对象大致位置确定)、光线阴影变化来源单一(正常情况下人员位置固定,正对摄像头或有小幅偏转,并且一般光照情况均较好,光照影响情况突变性较小)、遮挡引起表情变化情况简单(主要为低头、侧脸、托腮等小范围动作),混合干扰情况少(除了小幅动作影响,几乎没有其他遮挡和光线阴影变化因素)。通过使用已增强数据后的数据集对模型进行了 102 轮的训练,并经过近半年时间的算法改进,在基于 JAFFE 数据集的训练分类效果上,最终最高正确率已收敛至 91.7%。所有训练完成耗时达 3 017 min,在训练过程中,以参数零为均值,满足高斯分布的权重值标准差参数 0.1 为条件,按照偏置量参数零作为初始化值,每次选择 8 张图像代入训练过程。为了减少测试集与训练集数据划分偶然性带来的影响,实验中进行了 3 次随机抽样,对 3 次实验结果进行了平均值计算处理。

从实验结果来看,7 类表情中分类准确率最高的是中性(Neutral,最高达到 91.7%,平均值接近 80%),分类准确率最低的是厌恶(Disgust,最高达到 57.6%,平均值接近 22%)。该差异主要由两方面原因造成:(1)中性表情的训练集样本最为丰富,中性表情的特征性较为明确,与其他表情的交叉特征信息少;(2)厌恶表情的训练集样本最少,厌恶表情的特征并不十分明确,与其他表情例如愤怒(Angry)、恐惧(Fear)、悲伤(Sad)的特征重叠性较强。面部表情信息主要通过眉眼区域或唇口区域的动作形态变化及纹理特征变化进行传递,如果面部器官动作特征或纹理特征不明确,则

易引起识别结果不明确,从而造成分类错误的可能性增大,误识率增高,最终导致识别准确率降低。图 6 是本文对应的表情分析实验系统示例图,展示了本实验系统对 7 种表情图像(动态实时摄像或静态图片输入)的分析结果,对单张静态照片或动态实时摄像单帧的平均计算耗时均在 0.04 s 左右。本实验系统同时展现了同源表情识别的归一化概率分布,从中可以得出不同表情特征的独立排他性和重叠相关性。



图 6 表情分析实验系统结果示例

Fig.6 Examples of the results of the experimental expression analysis system

3.3 实验结果比较

针对增强后的 FER2013 数据集,本实验模型经过了 38 961 个样本的训练,其中用于测试集的为 4 012 张,训练中采用 300 epoch 为一组的单网格模式,针对 JAFFE 数据集,由于增强后还是属于样本量偏少类型,所以为了避免过拟合,训练中采用 8 epoch 为一组的单网格模式,并且使用了交叉验证的方式。对于这两个增强数据集的不同训练模型结果的应用情况,从实验分析数据来看,在 JAFFE 增强数据集上训练的模型结果的表情分类准确率高于在 FER2013 增强数据集上训练的模型结果。在 JAFFE 增强数据集上的识别分类准确率最高达到了 91.7%,而在 FER2013 增强数据集上的识别分类准确率最高只能达到 68.2%。造成两种情况下模型识别率差别的最主要因素,在于 FER2013 数据集的图片来源于互联网随机获取的多元数据,相较于 JAFFE 实验室特定情况下(10 名日本女性)约束型数据集而言,虽然样本数量占优,但样本质量缺乏规约性。然而,从识别率的鲁棒性角度而言,FER2013 较 JAFFE 样

本集的测试结果更优,这与样本量存在着直接的关系。图 7 和图 8 分别给出了基于 JAFFE 增强数据集以及 FER2013 增强数据集上训练的模型结果应用在实验系统中的表情分类识别率的混淆矩阵。

Happy	60.41	8.53	3.83	6.21	16.58	0.26	4.18
Scared	0.80	45.46	1.07	18.21	24.04	1.31	9.10
Surprised	8.71	10.41	44.13	6.46	26.82	0.01	3.46
Sad	18.53	21.82	4.95	23.69	15.21	7.35	8.45
Neutral	2.58	4.06	0.99	9.44	79.70	0.15	3.09
Disgust	16.83	20.59	2.36	14.97	9.90	21.11	14.23
Angry	8.02	14.95	2.21	12.74	28.57	0.42	33.09
	Happy	Scared	Surprised	Sad	Neutral	Disgust	Angry

图 7 基于 JAFFE 增强表情数据集训练模型平均识别率

Fig.7 Average recognition rate of training model based on JAFFE enhanced expression dataset

Happy	68.15	3.24	22.78	2.19	0.17	2.64	0.83
Scared	7.17	56.10	11.30	4.05	1.82	19.01	0.55
Surprised	17.03	5.58	48.05	3.61	0.70	7.76	17.27
Sad	0.28	31.05	6.03	51.02	2.16	6.03	3.43
Neutral	3.96	7.04	8.75	15.13	58.11	5.03	1.98
Disgust	11.72	4.08	8.15	19.02	1.17	37.72	18.14
Angry	4.01	13.60	17.33	7.29	0.22	15.84	41.71
	Happy	Scared	Surprised	Sad	Neutral	Disgust	Angry

图 8 基于 FER2013 增强表情数据集训练模型平均识别率

Fig.8 Average recognition rate of training model based on FER2013 enhanced expression dataset

4 结 语

本文使用了基于残差整流增强卷积神经网络的表情识别算法对人脸表情进行分类识别,该工作基于对现有卷积神经网络的残差补齐和激励整流同时辅以数据增强手段。实验结果表明,该方

法可以有效对受控情况下的人脸图像输入进行表情分类,最高准确率达到了 91.7%,在同训练集基础上,较其他识别算法效果更优。本文提及了该算法以及实验均在受控情况下进行,仅适用于特

定受控环境下的表情识别,而针对开放性非受控场景的密集型群体或人流表情的识别效果不佳,并未在本文中给出具体实验数据,这也是下一步需要改进和研究的方向。

参 考 文 献:

- [1] GUROVICH Y, HANANI Y, BAR O, *et al.* Identifying facial phenotypes of genetic disorders using deep learning [J]. *Nature Medicine*, 2019, 25(1): 60-69.
- [2] EKMAN P, FRIESEN W V. *Pictures of Facial Affect* [M]. Palo Alto, USA: Consulting Psychologists' Press, 1976: 1-7.
- [3] 张飞飞,张天柱,毛启容,等.基于生成对抗网络的多姿态人脸表情识别[J].计算机学报,2019,42(120):1-16.
ZHANG F F, ZHANG T Z, MAO Q R, *et al.* Multi-pose facial expression recognition via generative adversarial network [J]. *Chinese Journal of Computers*, 2019, 42(120): 1-16. (in Chinese)
- [4] 谭小慧,李昭伟,樊亚春.基于多尺度细节增强的面部表情识别方法[J].电子与信息学报,2019,41(11):2752-2759.
TAN X H, LI Z W, FAN Y C. Facial expression recognition method based on multi-scale detail enhancement [J]. *Journal of Electronics & Information Technology*, 2019, 41(11): 2752-2759. (in Chinese)
- [5] 李宏菲,李庆,周莉.基于多视觉描述子及音频特征的动态序列人脸表情识别[J].电子学报,2019,47(8):1643-1653.
LI H F, LI Q, ZHOU L. Dynamic facial expression recognition based on multi-visual and audio descriptors [J]. *Acta Electronica Sinica*, 2019, 47(8): 1643-1653. (in Chinese)
- [6] LEKDIOUI K, MESSOUSSE R, RUICHEK Y, *et al.* Facial decomposition for expression recognition using texture/shape descriptors and SVM classifier [J]. *Signal Processing: Image Communication*, 2017, 58: 300-312.
- [7] 夏添,张毅锋,刘袁.基于特征点与多网络联合训练的表情识别[J].计算机辅助设计与图形学学报,2019,31(4):552-559.
XIA T, ZHANG Y F, LIU Y. Landmark-based facial expression recognition by joint training of multiple networks [J]. *Journal of Computer-Aided Design & Computer Graphics*, 2019, 31(4): 552-559. (in Chinese)
- [8] 孙晓,潘汀.基于兴趣区域深度神经网络的静态面部表情识别[J].电子学报,2017,45(5):1189-1197.
SUN X, PAN T. Static facial expression recognition system using ROI deep neural networks [J]. *Acta Electronica Sinica*, 2017, 45(5): 1189-1197. (in Chinese)
- [9] LIU Y P, LI Y B, MA X, *et al.* Facial expression recognition with fusion features extracted from salient facial areas [J]. *Sensors*, 2017, 17(4): 712.
- [10] 贾明兴,杜俊强,宋鹏飞,等.基于不同分块多特征优化融合的人脸识别研究[J].东北大学学报(自然科学版),2017,38(3):310-314.
JIA M X, DU J Q, SONG P F, *et al.* Face recognition based on multi-feature optimization fusion of LBP, LPQ and Gabor with multi-scale blocks [J]. *Journal of Northeastern University (Natural Science)*, 2017, 38(3): 310-314. (in Chinese)
- [11] 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 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*. San Francisco, USA: IEEE, 2010: 94-101.
- [12] GOODFELLOW I J, ERHAN D, CARRIER P L, *et al.* Challenges in representation learning: A report on three machine learning contests [C]//*Proceedings of the 20th International Conference on Neural Information Processing*. Daegu, Korea: Springer, 2013: 117-124.
- [13] ZHENG W M, ZHOU X Y, ZOU C R, *et al.* Facial expression recognition using kernel canonical correlation analysis (KCCA) [J]. *IEEE Transactions on Neural Networks*, 2006, 17(1): 233-238.
- [14] ZHOU H L, LAM K M, HE X J. Shape-appearance-correlated active appearance model [J]. *Pattern Recognition*,

- 2016, 56: 88-99.
- [15] SU H, MAJI S, KALOGERAKIS E, *et al.* Multi-view convolutional neural networks for 3D shape recognition [C]//*Proceedings of the IEEE International Conference on Computer Vision*. Santiago, Chile: IEEE, 2015: 945-953.
- [16] KAYA H, GRPINAR F, SALAH A A. Video-based emotion recognition in the wild using deep transfer learning and score fusion [J]. *Image and Vision Computing*, 2017, 65: 66-75.
- [17] 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.
- [18] SOYEL H, DEMIREL H. Localized discriminative scale invariant feature transform based facial expression recognition [J]. *Computers & Electrical Engineering*, 2012, 38(5): 1299-1309.
- [19] 刘娟,胡敏,黄忠.基于最优支持度的证据融合表情识别方法[J].电子测量与仪器学报,2016,30(5):714-721.
LIU J, HU M, HUANG Z. Expression recognition method based on evidence fusion of optimal supportability [J]. *Journal of Electronic Measurement and Instrumentation*, 2016, 30(5): 714-721. (in Chinese)
- [20] 付傲威,赵敏,罗令,等.自由立体显示中基于深度卷积神经网络的虚拟视点生成方法[J].液晶与显示,2019,34(11):1031-1036.
FU A W, ZHAO M, LUO L, *et al.* Virtual viewpoint image generation method based on deep convolutional neural network in autostereoscopic display [J]. *Chinese Journal of Liquid Crystals and Displays*, 2019, 34(11): 1031-1036. (in Chinese)
- [21] 张万征,胡志坤,李小龙.基于 LeNet-5 的卷积神经图像识别算法[J].液晶与显示,2020,35(5):486-490.
ZHANG W Z, HU Z K, LI X L. Convolutional neural image recognition algorithm based on LeNet-5 [J]. *Chinese Journal of Liquid Crystals and Displays*, 2020, 35(5): 486-490. (in Chinese)
- [22] 王春哲,安军社,姜秀杰,等.基于卷积神经网络的候选区域优化算法[J].中国光学,2019,12(6):1348-1361.
WANG C Z, AN J S, JIANG X J, *et al.* Region proposal optimization algorithm based on convolutional neural networks [J]. *Chinese Optics*, 2019, 12(6): 1348-1361. (in Chinese)
- [23] 黄倩露,王强.基于跨连特征融合网络的面部表情识别[J].计算机工程与设计,2019,40(10):2969-2973.
HUANG Q L, WANG Q. Facial expression recognition based on cross-connect feature fusion network [J]. *Computer Engineering and Design*, 2019, 40(10): 2969-2973. (in Chinese)
- [24] GOODFELLOW I, BENGIO Y, COURVILLE A. *Deep Learning (Vol. 1)* [M]. Cambridge: MIT Press, 2016: 326-366.
- [25] GU J X, WANG Z H, KUEN J, *et al.* Recent advances in convolutional neural networks[J]. *arXiv preprint arXiv*: 1512.07108.
- [26] 王秀友,范建中,刘华明,等.基于概率协作表示的多表情序列融合识别[J].激光与光电子学进展,2019,56(13):131004.
WANG X Y, FAN J Z, LIU H M, *et al.* Multi-expression sequence fusion recognition based on probabilistic co-operative representation [J]. *Laser & Optoelectronics Progress*, 2019, 56(13): 131004. (in Chinese)
- [27] 姚丽莎,徐国明,赵凤.基于 CNN 局部特征融合的人脸表情识别[J].激光与光电子学进展,2020,57(3):031501.
YAO L S, XU G M, ZHAO F. Facial expression recognition based on CNN local feature fusion [J]. *Laser & Optoelectronics Progress*, 2020, 57(3): 031501. (in Chinese)
- [28] BARNOUTI N H, AL-DABBAGH S S M, MATTI W E, *et al.* Face detection and recognition using viola-jones with PCA-LDA and square Euclidean distance [J]. *International Journal of Advanced Computer Science and Applications*, 2016, 7(5): 371-377.
- [29] 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 Y, LIN X Z, JIANG M Y. Facial expression recognition with cross-connect LeNet-5 network [J]. *Acta Automatica Sinica*, 2018, 44(1): 176-182. (in Chinese)
- [30] GOYAL S J, UPADHYAY A K, JADON R S, *et al.* RETRACTED CHAPTER: Real-life facial expression rec-

- ognition systems; A review [M]//SATAPATHY S S, BHATEJA V, DAS S. *Smart Computing and Informatics*. Singapore: Springer, 2018: 311-331.
- [31] MAO Q R, ZHANG F F, WANG L J, *et al.* Cascaded multi-level transformed dirichlet process for multi-pose facial expression recognition [J]. *The Computer Journal*, 2018, 61(11): 1605-1619.

作者简介:



陈 斌(1978—),男,江苏南京人,博士,高级工程师,2015 年于河海大学获得博士学位,主要从事云计算、大数据分析、模式识别方面的研究。E-mail: njnuchenbin@njnu.edu.cn