



基于域适应卷积神经网络的人脸表情识别

亢 洁¹, 李佳伟¹, 杨思力²

(1. 陕西科技大学 电气与信息工程学院, 西安 710021; 2. 294188 部队 航空管制室, 西安 710077)

摘 要: 在利用卷积神经网络进行人脸表情识别时, 可借助其他数据集进行辅助训练以应对缺少标记数据的情况, 但源域数据库和目标域数据库之间的数据分布差异会影响分类正确率。为此, 以 AlexNet 网络为原型构建基于域适应的卷积神经网络结构。通过引入包含注意力机制的 SE 模块进行特征重标定, 同时利用域适应方法减小领域差异性。在人脸识别公开数据集上的实验结果表明, 与 AlexNet 和 GoingDeep 等网络相比, 该网络能够以较少的参数量获得较高的识别正确率。

关键词: 卷积神经网络; 人脸表情识别; 数据分布; 域适应; 迁移学习

开放科学(资源服务)标志码(OSID):



中文引用格式: 亢洁, 李佳伟, 杨思力. 基于域适应卷积神经网络的人脸表情识别[J]. 计算机工程, 2019, 45(12): 201-206.

英文引用格式: KANG Jie, LI Jiawei, YANG Sili. Facial expression recognition based on convolutional neural network with domain adaption[J]. Computer Engineering, 2019, 45(12): 201-206.

Facial Expression Recognition Based on Convolutional Neural Network with Domain Adaption

KANG Jie¹, LI Jiawei¹, YANG Sili²

(1. School of Electrical and Information Engineering, Shaanxi University of Science and Technology, Xi'an 710021, China;

2. Air Control Room, Unit 294188, Xi'an 710077, China)

[Abstract] In facial expression recognition using Convolutional Neural Network (CNN), other data sets can assist in training to deal with the lack of tag data. However, the classification accuracy will be affected by the data distribution differences between the source domain database and the target domain database. To address the problem, this paper constructs a domain adaptation-based convolutional neural network structure modeled after AlexNet. The network introduces a SE module including the attention mechanism for feature re-location, and uses the domain adaption method to reduce the differences between domains. Experimental results on the public data sets of facial expression recognition show that the proposed network can achieve a higher recognition accuracy rate than AlexNet and GoingDeep with fewer parameters.

[Key words] Convolutional Neural Network (CNN); facial expression recognition; data distribution; domain adaptation; transfer learning

DOI: 10.19678/j.issn.1000-3428.0054950

0 概述

人脸表情是人类表达情感中最自然的信息, 其在人们表达情感的过程中传达了 55% 的有用信息, 而语言所传达的有用信息仅占 7%^[1]。因此, 近年来人脸表情识别被广泛应用于人机交互、医疗、教育和游戏等领域, 已经成为计算机视觉和机器学习方面的研究热点^[2]。

早期的人脸表情识别方法主要采用手动提取特征(如 LBP^[3]、HOG^[4]和 Gabor^[5]等)的方法, 在此基

础上再利用 SVM^[6]进行分类。文献[7]提出一种改进的多特征表情识别方法, 首先利用链码编码刻画表情形状特征并构建形变特征描述面部几何变化, 然后构造 Gabor 特征融合图以表征表情局部纹理细节, 最后采用支持向量机对表情进行分类。文献[8]提出一种基于高斯过程隐变量模型的表情识别方法, 首先运用高斯过程隐变量对提取的表情特征进行降维, 然后利用变分稀疏高斯过程分类方法进行表情识别, 但此方法受人干扰影响较大。

基金项目: 国家自然科学基金(61603233); 陕西省自然科学基金基础研究计划(2017JQ6076)。

作者简介: 亢洁(1973—), 女, 副教授、博士, 主研方向为计算机视觉、模式识别; 李佳伟, 硕士研究生; 杨思力, 工程师、硕士。

收稿日期: 2019-05-20 修回日期: 2019-06-25 E-mail: 316938770@qq.com

卷积神经网络通过构建多个卷积层来提取人脸表情更深层次的特征,可避免人为提取特征带来的误差且具有很强的鲁棒性,逐渐成为目前的主流方法。文献[9]提出一种多分辨率特征融合的卷积神经网络,首先利用2个相互独立且深度不同的通道对图片进行特征提取,使卷积神经网络自主学习同一图像下不同分辨率的特征,然后将不同分辨率的特征送入全连接层并进行特征融合,最后基于 SoftMax 分类器进行表情分类。文献[10]设计一个隔离损失函数,相对于一般网络所采用的多分类损失函数,该函数能更准确地区分不同类之间的中心距离,提升网络分类效果。

卷积神经网络具有较高的识别率和鲁棒性,但其需要大量的训练数据进行网络训练,在实际中往往很难付出高昂的代价利用人工标记大量表情数据集,因此,很多表情识别任务因缺少标记数据而无法利用卷积神经网络。针对此问题,可利用其他有丰富标记的数据集进行辅助训练,但不同的人脸表情数据集通常具有不同的姿态、光照和清晰度,会使目标域数据集和源域数据集的数据分布存在很大差异,导致正确率下降。针对此问题,可采用迁移学习进行辅助训练,同时减少源域数据集和目标域数据集之间的分布差异。目前迁移学习主要有基于实例^[11]、基于特征^[12]和基于模型^[13]3种方法,领域自适应是基于特征的方法。文献[14]提出一种域再生成的无监督学习方法,通过减少源域数据和目标域数据之间分布的差距来提高识别的准确率。文献[15]提出一个新的域适应卷积神经网络,将 coral 损失函数和多分类损失函数共同作为训练优化目标对源域知识进行迁移。文献[16]提出一种基于 LBP 特征的核均值匹配方法,通过计算源域和目标域的核均值进行数据分布的自动匹配,但该方法需要标记一定量准确的目标域数据。

为高效识别不同分布且无标注的人脸表情数据,本文引入 Squeeze-and-Excitation (SE) 模块^[17]设计基于域适应的卷积神经网络结构。通过对全连接层进行适配,把源域数据和目标域数据映射到再生核希尔伯特空间得到域混淆损失函数,并将其加入到分类损失函数中,在减小源域及目标域分布差异的同时得到准确的目标域表情分类结果。

1 基于域适应的卷积神经网络

1.1 网络结构

本文以 AlexNet 为原型,提出一种基于域适应的卷积神经网络用于表情识别,该网络结构如图1所示。

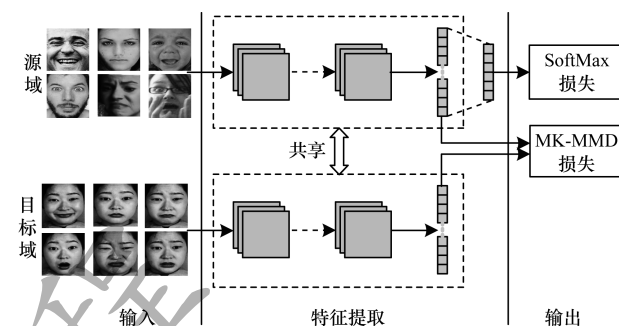


图1 基于域适应的卷积神经网络结构

基于域适应的卷积神经网络由共享参数的源域卷积神经网络和目标域卷积神经网络组成。本文在对 AlexNet 进行改进的同时,引入具有注意力机制的 SE 模块来提高网络的特征提取能力,最终通过输入源域 X_s 、源域表情标签 L_s 和无标签的目标域 X_t 来训练网络,使 X_t 的表情得到正确分类。训练过程中的域适应通过第一个全连接层来实现,作为域适应适配层,本文分别提取 X_s 和 X_t 在该层的特征,利用多核最大均值差异 (Multi-Kernel Maximum Mean Discrepancies, MK-MMD)^[18] 计算提取到的特征之间的分布距离,最终将该距离作为2个域之间的分布差异和多分类损失一起形成优化目标。此外,网络中所有的输入都参与计算 MK-MMD 损失,但只有被标记的源域数据用于计算多分类损失。

● 网络结构及参数信息如表1所示。本文设计的卷积神经网络主要由6个卷积层和2个全连接层依次相连组成,卷积层的所有卷积核大小为 3×3 (步长为1),通道数分别为 64、64、256、256、128、128,第一个全连接层是2 048 维的域适应适配层,最后一个全连接层是7 维的 SoftMax 输出层,用于对7类表情的预测。网络中的每一个卷积层后都依次采用批量归一化 (Batch Normalization, BN)^[19]、修正线性单元 (Rectified Linear Unit, ReLU)^[20] 和 SE 模块并且在第2个、第3个和第6个 SE 模块后嵌入窗口大小为3、步长为2的最大池化层。

表1 网络结构及参数信息

| 层数 | 层类型 | 卷积核大小及步长 | 输出(宽×高×通道数) | 参数量 |
|----|-------------|---------------------|---------------------------|---------|
| 1 | Convolution | 3×3 , 步长为1 | $56 \times 56 \times 64$ | 2 048 |
| 2 | SE | — | $56 \times 56 \times 64$ | 580 |
| 3 | Convolution | 3×3 , 步长为1 | $56 \times 56 \times 64$ | 37 184 |
| 4 | SE | — | $56 \times 56 \times 64$ | 580 |
| 5 | Max pooling | 3×3 , 步长为2 | $28 \times 28 \times 64$ | 0 |
| 6 | Convolution | 3×3 , 步长为1 | $28 \times 28 \times 256$ | 148 736 |
| 7 | SE | — | $28 \times 28 \times 256$ | 8 464 |
| 8 | Max pooling | 3×3 , 步长为2 | $14 \times 14 \times 256$ | 0 |
| 9 | Convolution | 3×3 , 步长为1 | $14 \times 14 \times 256$ | 591 104 |
| 10 | SE | — | $14 \times 14 \times 256$ | 8 464 |

表 1 网络结构及参数信息(续)

| 层数 | 层类型 | 卷积核大小及步长 | 输出(宽×高×通道数) | 参数量 |
|----|-----------------|------------|-------------|------------|
| 11 | Convolution | 3×3, 步长为 1 | 14×14×128 | 295 552 |
| 12 | SE | — | 14×14×128 | 2 184 |
| 13 | Convolution | 3×3, 步长为 1 | 14×14×128 | 148 096 |
| 14 | SE | — | 14×14×128 | 2 184 |
| 15 | Max pooling | 3×3, 步长为 2 | 7×7×128 | 0 |
| 16 | Full Connection | — | 1×1×2 048 | 12 847 104 |
| 17 | Full Connection | — | 1×1×7 | 14 343 |

1.2 人脸表情特征提取

本文主要以卷积层和全连接层为基础来提取人脸表情特征。为使网络具有更好的特征提取能力,在原网络中加入一些新的模块。首先在每个卷积层后都连接具有学习能力的 BN 层,使中间层数据分布在训练过程中不发生改变,在加快网络收敛速度的同时提高初始学习率;然后在 BN 层后引入具有注意力机制的 SE 模块,该模块由压缩、激发和重分配 3 个部分构成,当上一层网络结构输出一个宽为 W 、高为 H 、通道数为 C 的特征 Z 时,SE 模块可以学习到该特征不同通道的重要性并进行加权,最终输出一个表征能力更强的特征 Z' ,SE 模块结构如图 2 所示。

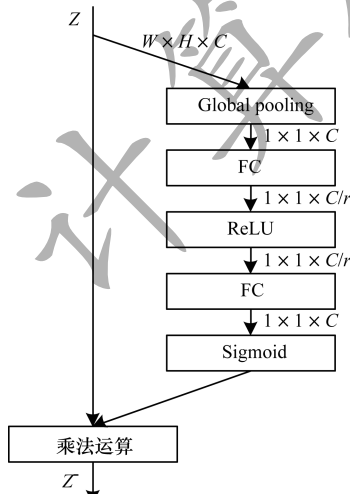


图 2 SE 模块结构

SE 模块首先进行压缩操作,通过全局平均池化方法将 Z 的 $W \times H \times C$ 维特征压缩成 $1 \times 1 \times C$ 维特征;然后进行激发操作,采用 2 个全连接层去表示通道间的相关性,第 1 个全连接层将全局池化后的 $1 \times 1 \times C$ 维特征的通道维度降低到输入的 $1/r$,第 2 个全连接层将 C/r 维通道特征升回到原来的 C 维,2 个全连接层之间采用一个 ReLU 函数激活;最后进行重分配操作,对激发操作的输出通过一个 Sigmoid 激活函数获得 $0 \sim 1$ 之间归一化的权重,再通过乘法运算将归一化后的权重加权到每个通道的特征上,以此使网络可依照权重提升有用的特征并抑制对当前任务作用不大的特征。在上述操作基础上,将 ReLU 激活函数应用到全部的卷积层和全连接层。

ReLU 定义为 $f(x) = \max(0, x)$,当输入 $x < 0$ 时,输出为 0;当 $x > 0$ 时,输出为 x 。该激活函数能使网络更快速地收敛并且可避免使用其他激活函数产生的梯度消失问题。

1.3 基于域适应的人脸表情分类

本文设计的网络是一个分类模型,其通过域适应策略将源域表情特征知识迁移到目标域上,从而实现目标域表情的正确分类。该网络首先分别提取源域和目标域在适配层的表情特征来计算 2 个域之间的 MK-MMD 距离。MK-MMD 用于表示不同域之间的分布距离,它是在最大均值差异(Maximum Mean Discrepancies, MMD)^[21]距离基础上的改进。MMD 距离定义如下:

$$MMD[F, p, q] := \sup_{f \in F} (E_p[f(X_s)] - E_q[f(X_t)]) \quad (1)$$

其中, \sup 表示函数的最小上界, E_p 和 E_q 是分布为 p 的源域数据 X_s 和分布为 q 的目标域数据 X_t 的数学期望, F 是所有特征空间映射函数 f 的集合。由于当 F 为再生核希尔伯特空间中的单位球时分布度量效果最佳,因此本文采用的 MK-MMD 是通过多个高斯核函数将 $f(\cdot)$ 表示为再生核希尔伯特空间 H 来度量源域和目标域的分布差异,高斯核函数为:

$$k(X_s, X_t) = \exp\left(-\frac{\|X_s - X_t\|^2}{2\sigma^2}\right) \quad (2)$$

其中, σ 是由中值算法得到的高斯核函数的带宽^[20]。再生核希尔伯特空间的 MK-MMD 定义如下:

$$MMD_{MK}[H, p, q] := \|\Phi(X_s) - \Phi(X_t)\|_H \quad (3)$$

其中, $\Phi(\cdot)$ 为 $f(\cdot)$ 的高斯核函数映射。

为实现域迁移能力的最大化,本文采用核选择策略^[19],计算式如下:

$$K := \left\{ k = \sum_{u=1}^d \beta_u k_u, \sum_{u=1}^d \beta_u = 1, \beta_u \geq 0, \forall u \right\} \quad (4)$$

其中, d 为高斯核的数量, β 是由学习所得到的不同高斯核的权重^[18]。

文献[22]提出一种无偏差的 MK-MMD 计算公式,定义如下:

$$MMD_{MK}^2[H, p, q] = \frac{1}{m(m-1)} \sum_{i \neq j}^m k(X_{si}, X_{sj}) + \frac{1}{n(n-1)} \sum_{i \neq j}^n k(X_{ti}, X_{tj}) + \frac{2}{mn} \sum_{i=1}^m \sum_{j=1}^n k(X_{si}, X_{tj}) \quad (5)$$

其中, m 为 X_s 的数量, n 为 X_t 的数量, 当且仅当 p 和 q 分布相同时 MK-MMD 距离为 0。当适配层的计算出 MK-MMD 距离后, 将其与多分类损失函数一起作为域适应卷积神经网络的优化目标来最小化源域和目标域的分布。多分类损失函数定义为:

$$L_c(y_p, y) = \sum_{i=0}^6 -y_{pi} \ln y_i \quad (6)$$

其中, $L_c(y_p, y)$ 为 X_s 的多分类损失函数, y_p 为网络预测的类别, y 为真实的类别, i 表示表情类别。最终的域适应任务中总损失函数定义为:

$$L = L_c(y_p, y) + \lambda MMD_{MK}^2[H, p, q] \quad (7)$$

其中, L 为域适应总损失, $L_c(y_p, y)$ 为多分类损失函数, $MMD_{MK}^2[H, p, q]$ 是源域和目标域之间分布的距离, 通过在适配层上计算得到, 参数 λ 控制着分布距离在多大程度上参与网络的训练。在训练过程中, 将总损失函数 L 作为优化目标来更新网络参数, 使得卷积神经网络可以不断缩小源域和目标域数据之间分布的差异, 以此实现域自适应。

1.4 训练策略

本文设计的网络在训练时参数更新策略采用带动量优化的随机梯度下降, 动量设置为 0.9, 训练每批的数据量为 128, 初始学习率设置为 0.008, 每轮的权重衰减系数为 0.000 1。依据文献[18], 本文共选取 5 个高斯核, 不同核的带宽分别设置为 $\sigma/4$ 、 $\sigma/2$ 、 σ 、 2σ 、 4σ 。

本文所有实验均在 python3.6.5 上实现, 硬件平台为 Intel(R) Core(TM) i7-8700 CPU, 内存为 16 GB, GPU 为 11 GB 的 NVIDIA GeForce GTX 1080Ti。

2 实验结果与分析

2.1 数据集及数据预处理

本文所采用的数据集为 JAFFE 数据集^[23]、RAF-DB 数据集^[24] 和 CK+ 数据集^[25]。JAFFE 数据集是在实验室环境下采集的数据集, 该表情数据库由 10 名日本女性做出的 7 种基本表情(生气、厌恶、害怕、高兴、悲伤、兴奋和正常)构成, 每个人每种表情有 2 张~4 张, 共 213 张图片。本文在使用该数据集之前对数据进行了裁剪, 将原始表情图像的人脸区域裁剪出来以减少表情无关区域的影响, 裁剪效果如图 3 所示。

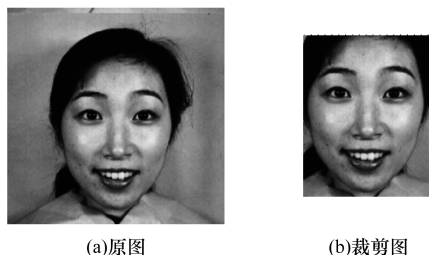


图 3 JAFFE 裁剪效果示例

RAF-DB 数据集是自然环境下的数据集, 它包含 29 672 张图片, 由 7 种基本表情子集和 11 种混合表情子集组成。本文利用该数据集的 7 种基本表情子集(生气、高兴、厌恶、正常、惊讶、悲伤、害怕)进行训练, 包含 15 339 张图片。由于该数据集不同类的样本数量相差较大, 因此对该数据集做均衡化处理: 对高兴和正常的表情数据采用欠采样的方法; 对生气、厌恶和害怕的表情数据采用随机旋转和水平翻转的数据增强方法。平衡后的数据集包含 14 535 张图片, 平衡效果如图 4 所示。

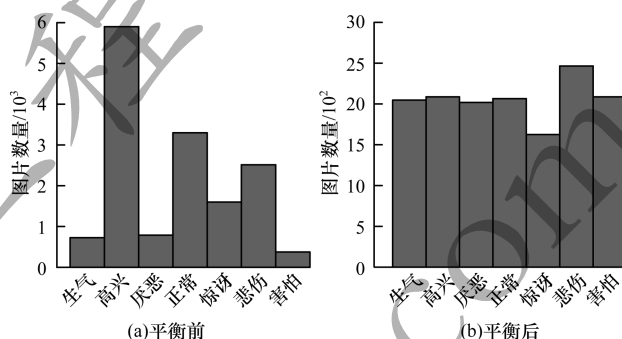


图 4 RAF-DB 数据平衡效果

CK+ 数据集是在实验室环境下对 123 个人采集的 593 段表情序列。在采集的表情序列中, 只有 309 段表情序列被标记为 6 种基本表情(生气、厌恶、害怕、高兴、悲伤和惊讶)之一。本文按照常用的选择方法, 从 309 段表情序列中抽取第 1 帧(正常表情)和最后 3 帧组成具有 7 种基本表情的表情库, 共计 1 236 张表情图片。下文中提及的 CK+ 数据集指所组成的表情库。

在本文的所有实验中, 将有标签的 RAF-DB 数据集作为源域数据集, 将无标签的 JAFFE 数据集和无标签的 CK+ 数据集作为目标域数据集。源域数据集全部作为训练集, 目标域数据集的训练集、验证集和测试集划分比例为 8:1:1。数据集中的图片统一缩放到 56 像素 × 56 像素大小再作为网络的输入。

2.2 实验对比与分析

本文采用表情识别准确率作为评价准则, 其定义:

$$Accuracy = \sum_{i=1}^7 \frac{N_i}{N} \quad (8)$$

其中, N 为 X_t 的样本数量, N_i 为识别第 i 类表情正确的样本数量。本节首先讨论压缩率对比实验来确定基础的卷积神经网络分类模型; 然后讨论惩罚系数对比实验使得分类模型具有最好的迁移效果; 最后通过对比不同方法来确定本文方法的有效性。

2.2.1 压缩率对比实验

在 SE 模块中设置不同的压缩率 r 可以使模块具有不同的通道加权能力。本文在未使用域适应策略的条件下将 r 分别设置为 2、4、8、16、32 进行对比实验, 实验结果如图 5 所示。从中可以看出, 当 SE 模块的压缩率逐渐变大时, 表情识别正确率先升高后下降, 这是因为压缩率太小会使模块学习不到特征的主要知识, 压缩率太大会使得模块容量变小而无法有效学习, 因此, 选择合适的压缩率可使网络具有最好的特征提取能力。由图 5 可知, 当 r 为 16 时网络具有最高的表情识别正确率, 此时 SE 模块最好地学习到了通道之间不同贡献率的知识, 因此, 在本文的所有实验中将 r 设置为 16。

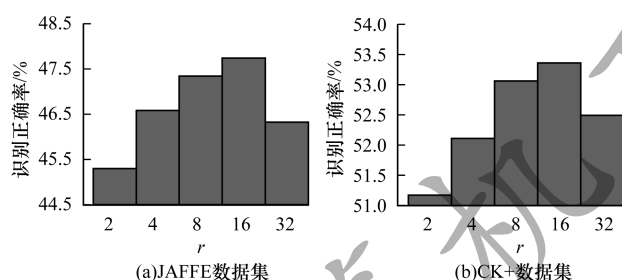


图 5 不同压缩率下的表情识别正确率

2.2.2 惩罚系数对比实验

图 6 是不同惩罚系数 λ 的对比实验结果。域适应任务总损失中 λ 决定了迁移的程度, 本文将 λ 分别取值为 0.2、0.4、0.6、0.8、1.0 来测试域适应程度不同对分类准确率产生的影响。由图 6 可以看出, 当 λ 取 0.2 时迁移效果最差, 网络在此时的迁移能力最弱, 但比不加迁移时的识别正确率要高, 证明了迁移方法的有效性。当 λ 逐渐变大, 表情识别正确率先升高后下降, 当 λ 为 0.8 时正确率达到最高值, 此时网络能够最好地学习到源域数据集的知识。由此可以证明, 最大的迁移程度未必能够得到最好的迁移效果, 但惩罚系数的选择必须恰当。

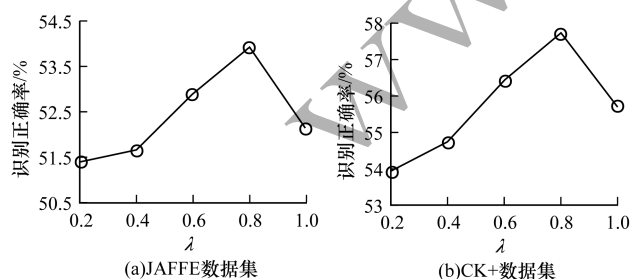


图 6 不同惩罚系数下的表情识别正确率

2.2.3 不同方法对比实验

本文通过表情识别正确率和网络结构参数量 2 个指标来与经典的 AlexNet 网络^[20]、GoingDeep 网

络^[26] 和 Deep CORAL 网络^[15] 进行对比, 以此证明提出方法的有效性, 对比结果如表 2 所示。

表 2 不同网络的识别正确率对比

| 网络 | 识别正确率/% | | 参数量 |
|------------|-----------|----------|------------|
| | JAFFE 数据集 | CK + 数据集 | |
| GoingDeep | 37.85 | 41.66 | 27 490 375 |
| AlexNet | 44.13 | 39.24 | 58 310 023 |
| Deep CORAL | 48.18 | 52.53 | 58 310 023 |
| CNN | 47.75 | 53.33 | 14 106 623 |
| CNN + MMD | 53.93 | 57.72 | 14 106 623 |

由表 2 可以看出, 本文设计的网络结构在不进行域适应时, 对 JAFFE 和 CK + 数据库图片中的表情识别正确率比 AlexNet 和 GoingDeep 高, 证明该结构可以更有效地提取表情特征信息。同时, 其参数量最少, 在实现卷积神经网络轻量化的同时取得了较高的表情识别准确率。当加上提出的域适应方法后, JAFFE 库的正确率在 47.75% 的基础上提升了 6.18%, CK + 库的正确率在 53.33% 的基础上提升了 4.39%。另外, 本文方法在 JAFFE 和 CK + 数据库上的正确率比 Deep CORAL 网络分别高 5.75% 和 5.19%, 足以证明本文提出的域适应卷积神经网络可以有效地减小源域和目标域数据之间的分布误差, 更适用于人脸表情识别问题。目标域人脸表情部分识别结果如图 7 和图 8 所示。

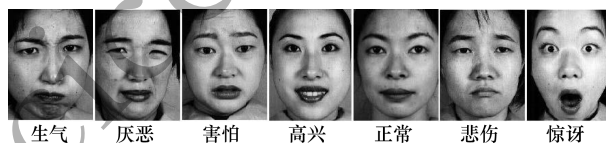


图 7 JAFFE 人脸表情部分识别结果



图 8 CK + 人脸表情部分识别结果

3 结束语

针对人脸表情识别因缺少标注数据难以使用卷积神经网络的情况, 本文设计一个基于域适应的卷积神经网络。对源域表情图片和目标域表情图片, 分别通过 2 个权重共享且带有 SE 模块的卷积通道进行特征提取。在此基础上, 在全连接层计算不同域特征之间的最大均值距离, 把减小最大均值距离作为网络的优化目标, 以此将源域的知识迁移到目标域上, 实现对目标域中无标注表情数据的正确分类。下一步将针对卷积神经网络设计性能更优的域适应策略, 使网络具有更强的表情迁移能力。

参考文献

- [1] MEHRABIAN A, RUSSELL J A. An approach to environmental psychology[M]. Cambridge, USA: MIT Press, 1974.
- [2] ZHANG Yanliang, LU Bing. Micro-expression recognition method based on information gain feature selection[J]. Computer Engineering, 2019, 45(5): 261-266. (in Chinese)
张延良, 卢冰. 基于信息增量特征选择的微表情识别方法[J]. 计算机工程, 2019, 45(5): 261-266.
- [3] ZHAO Guoying, PIETIKÄINEN M. Dynamic texture recognition using local binary patterns with an application to facial expressions[J]. Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6): 915-928.
- [4] DAHMANE M, MEUNIER J. Emotion recognition using dynamic grid-based HoG features[C]//Proceedings of International Conference on Automatic Face and Gesture Recognition. Washington D. C., USA: IEEE Press, 2011: 884-888.
- [5] BARTLETT M S, LITTLEWORT G, FRANK M, et al. Recognizing facial expression: machine learning and application to spontaneous behavior[C]//Proceedings of Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2005: 568-573.
- [6] CORTES C, VAPNIK V N. Support vector networks[J]. Machine Learning, 1995, 20(3): 273-297.
- [7] HUANG Zhong, HU Min, LIU Juan. Facial expression recognition method based on multi-feature decision-level fusion[J]. Computer Engineering, 2015, 41(10): 171-176. (in Chinese)
黄忠, 胡敏, 刘娟. 基于多特征决策级融合的表情识别方法[J]. 计算机工程, 2015, 41(10): 171-176.
- [8] PAN Wusheng, HUANG Yushui. A facial expression recognition method based on gaussian process latent variable method[J]. Computer Simulation, 2018, 35(3): 341-344. (in Chinese)
潘武生, 黄玉水. 一种基于高斯过程隐变量模型的表情识别方法[J]. 计算机仿真, 2018, 35(3): 341-344.
- [9] HE Zhichao, ZHAO Longzhang, CHEN Chuang. Convolution neural network with multi-resolution feature fusion for facial expression recognition[J]. Laser & Optoelectronics Progress, 2018, 55(7): 370-375. (in Chinese)
何志超, 赵龙章, 陈闯. 用于人脸表情识别的多分辨率特征融合卷积神经网络[J]. 激光与光电子学进展, 2018, 55(7): 370-375.
- [10] CAI Jie, MENG Zibo, KHAN A S, et al. Island loss for learning discriminative features in facial expression recognition[C]//Proceedings of the 13th IEEE International Conference on Automatic Face and Gesture Recognition. Washington D. C., USA: IEEE Press, 2018: 302-309.
- [11] DAI Wenyuan, YANG Qiang, XUE Guirong, et al. Boosting for transfer learning[C]//Proceedings of the 24th International Conference on Machine Learning. New York, USA: ACM Press, 2007: 193-200.
- [12] SHEN Junge, ZHENG Enrang, CHENG Zhiyong, et al. Assisting attraction classification by harvesting Web data[J]. IEEE Access, 2017, 10(5): 1600-1608.
- [13] ZHAO Zhongtang, CHEN Yiqiang, LIU Junfa, et al. Cross-mobile ELM based activity recognition[J]. International Journal of Engineering and Industries, 2010, 1(1): 30-38.
- [14] YAN Keyu, ZHENG Wenming, CUI Zhen, et al. Cross-database facial expression recognition via unsupervised domain adaptive dictionary learning[C]//Proceedings of International Conference on Neural Information Processing. Berlin, Germany: Springer, 2016: 427-434.
- [15] SUN Baochen, SAENKO K. Deep coral: correlation alignment for deep domain adaptation[C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2016: 443-450.
- [16] MIAO Yunqian, ARAUJO R, KAMEL M S. Cross-domain facial expression recognition using supervised kernel mean matching[C]//Proceedings of the 11th International Conference on Machine Learning and Applications. Washington D. C., USA: IEEE Press, 2012: 326-332.
- [17] HU Jie, SHEN Li, ALBANIE S, et al. Squeeze-and-excitation networks[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 7132-7141.
- [18] GRETTON A, SEJDINOVIC D, STRATHMANN H, et al. Optimal kernel choice for large-scale two-sample tests[M]//PEREIRA F, BURGESS C J C, BOTTOU L, et al. Advances in neural information processing systems. Cambridge, USA: MIT Press, 2012: 1205-1213.
- [19] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]//Proceedings of International Conference on Machine Learning. New York, USA: ACM Press, 2015: 448-456.
- [20] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with deep convolutional neural networks[J]. Advances in Neural Information Processing Systems, 2012, 25(2): 1097-1105.
- [21] SEJDINOVIC D, SRIPERUMBUDUR B, GRETTON A, et al. Equivalence of distance-based and RKHS-based statistics in hypothesis testing[J]. The Annals of Statistics, 2013, 41(5): 2263-2291.
- [22] GRETTON A. A kernel two-sample test[J]. Journal of Machine Learning Research, 2012, 13(1): 723-773.
- [23] LYONS M J, AKAMATSU S, KAMACHI M, et al. Coding facial expressions with Gabor wavelets[C]//Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition. Washington D. C., USA: IEEE Press, 1998: 200-205.
- [24] LI Shan, DENG Weihong, DU Junping. Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 2584-2593.
- [25] 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 Computer Vision and Pattern Recognition Workshops. Washington D. C., USA: IEEE Press, 2010: 94-101.
- [26] MOLLAHOSSEINI A, CHAN D, MAHOOR M H. Going deeper in facial expression recognition using deep neural networks[C]//Proceedings of 2016 IEEE Winter Conference on Applications of Computer Vision. Washington D. C., USA: IEEE Press, 2016: 1-10.