



# (12)发明专利申请

(10)申请公布号 CN 111160189 A

(43)申请公布日 2020.05.15

(21)申请号 201911331871.0

(22)申请日 2019.12.21

(71)申请人 华南理工大学

地址 510640 广东省广州市天河区五山路  
381号

(72)发明人 文贵华 常天元 诸俊浩

(74)专利代理机构 广州粤高专利商标代理有限公司 44102

代理人 何淑珍 江裕强

(51) Int. Cl.

G06K 9/00(2006.01)

G06K 9/62(2006.01)

G06N 3/04(2006.01)

G06N 3/08(2006.01)

权利要求书3页 说明书8页 附图2页

## (54)发明名称

一种基于动态目标训练的深度学习人脸表情识别方法

## (57)摘要

本发明公开了一种基于目标动态训练的深度学习人脸表情识别方法。所述方法包括以下步骤:输入训练样本集和测试样本集;采用多个损失函数作为训练的目标函数;将损失函数排序;根据排序后的损失函数依次在训练样本集上重复训练深度学习神经网络,获得表情识别模型;根据表情识别模型对输入的测试样本进行表情分类。本发明让模型在初期能尽快拟合训练数据,在后期通过多个损失的动态权重变化拉大不同类别的特征空间距离以达到更好的泛化性能。本发明能够有效地提升神经网络在多个人脸表情识别数据集上的准确率。



1. 一种基于动态目标训练的深度学习人脸表情识别方法,其特征在于,包括以下步骤:

S1、输入训练样本集和测试样本集;

S2、采用多个损失函数作为训练的目标函数;

S3、将损失函数排序;

S4、根据排序后的损失函数依次在训练样本集上重复对神经网络进行动态目标训练,获得表情识别模型;

S5、根据表情识别模型对输入的测试样本进行表情分类。

2. 根据权利要求1所述的一种基于动态目标训练的深度学习人脸表情识别方法,其特征在于,步骤S1中,对输入的任何人脸图像,属于三通道的彩色图像,通过双线性插值的方式进行放缩,每张图像进行随机的水平翻转、随机平移、随机图像旋转、随机图像缩放。

3. 根据权利要求1所述的一种基于动态目标训练的深度学习人脸表情识别方法,其特征在于,步骤S2中损失函数包括最大间隔高斯混合损失函数L-GM、似然正则化损失函数、基于表情分类领域知识的损失函数和L<sub>2</sub>正则项损失函数。

4. 根据权利要求3所述的一种基于动态目标训练的深度学习人脸表情识别方法,其特征在于,L-GM损失函数在训练集上提取的深度特征x遵循高斯混合概率分布,其中p(k)代表深度特征x的第k类的先验概率, $\mu_k$ 和 $\Sigma_k$ 分别表示类别k的均值和协方差,公式如下:

$$p(x) = \sum_{k=1}^K N(x; \mu_k, \Sigma_k) p(k);$$

其中N表示高斯分布函数,K表示分类的总类别个数;

基于上述约束,训练集上提取的深度特征 $x_i$ 在给定的类别标签 $z_i \in [1, K]$ 下的条件概率的计算如下所示:

$$p(x_i | z_i) = N(x_i; \mu_{z_i}, \Sigma_{z_i});$$

相应的后验概率分布的计算如下所示:

$$p(z_i | x_i) = \frac{N(x_i; \mu_{z_i}, \Sigma_{z_i}) p(z_i)}{\sum_{k=1}^K N(x_i; \mu_k, \Sigma_k) p(k)};$$

混合高斯分布的类别损失函数 $L_{cls}$ 通过后验概率分布与类别标签的交叉熵来计算,公式如下:

$$\begin{aligned} L_{cls} &= -\frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K 1(z_i = k) \log p(k | x_i) \\ &= -\frac{1}{N} \sum_{i=1}^N \log \frac{N(x_i; \mu_{z_i}, \Sigma_{z_i}) p(z_i)}{\sum_{k=1}^K N(x_i; \mu_k, \Sigma_k) p(k)}; \end{aligned}$$

转换用距离来度量后的混合高斯分布的类别损失函数 $L_{cls}$ ,其中 $d_k$ 的计算采用马氏距离来衡量相关特征间的距离,具体公式如下:

$$L_{cls,i} = -\log \frac{p(z_i) |\Sigma_{z_i}|^{-\frac{1}{2}} e^{-d_{zi}}}{\sum_k p(k) |\Sigma_k|^{-\frac{1}{2}} e^{-d_k}};$$

$$d_k = (x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k) / 2;$$

其中,  $T$ 表示矩阵转置,  $e$ 为自然常数,  $L_{cls, i}$ 表示第 $i$ 个样本的类别损失函数的大小。

5. 根据权利要求3所述的一种基于动态目标训练的深度学习人脸表情识别方法, 其特征在于, 似然正则化损失函数用于测量训练样本在何种程度上符合假设的混合高斯分布, 通过对最大似然估计取负来进行计算, 公式如下:

$$L_{lkd} = - \sum_{i=1}^N \log N(x_i; \mu_{zi}, \Sigma_{zi})。$$

6. 根据权利要求3所述的一种基于动态目标训练的深度学习人脸表情识别方法, 其特征在于, 基于表情识别领域知识的损失函数的目标是在人脸表情样本在距离自身类中心距离尽可能小的同时引入先验领域信息知识拉大其与易混分表情类中心的距离; 领域损失函数公式如下:

$$L_{domain\_information} = \sum_i^N [ \|f(x_i) - f(x_i^s)\|_2^2 - \|f(x_i) - f(x_i^m)\|_2^2 + \alpha ]_+;$$

其中参数 $\alpha$ 的作用是在网络模型的学习中加入强约束用来定义训练样本特征与同类中心特征间距离和对应易混分类中心特征间距离的一个最小间隔;  $f(x_i)$ 代表训练样本 $x_i$ 的特征表示,  $f(x_i^s)$ 代表训练样本 $x_i$ 的同类表情类别中心的特征表示,  $f(x_i^m)$ 表示训练样本 $x_i$ 对应易混分表情标签类别中心的特征表示;  $+$ 代表中括号内的值大于0时取该值作为损失值, 当中括号内的值小于0时损失值为0。

7. 根据权利要求3所述的一种基于动态目标训练的深度学习人脸表情识别方法, 其特征在于, 正则项就是稀疏性, 以保证抽取的特征尽可能简单;  $L_2$ 正则项损失函数的公式如下:

$$L_2 = \sum_{i=1}^M w_i^2$$

其中 $M$ 代表神经网络的参数个数,  $w_i$ 是模型用于更新的第 $i$ 个参数。

8. 根据权利要求1所述的一种基于动态目标训练的深度学习人脸表情识别方法, 其特征在于, 步骤S3中, 依据损失函数对模型的作用将四个损失函数排序分为三个等级, 第一等级: 分类损失函数; 第二等级: 似然正则化损失函数, 基于表情识别领域知识的损失函数; 第三等级:  $L_2$ 正则项损失函数。

9. 根据权利要求1所述的一种基于动态目标训练的深度学习人脸表情识别方法, 其特征在于, 步骤S4中, 搭建深度学习模型, 依次设置为卷积层, 池化层, 3层残差模块, 协方差层和全连接层; 其中卷积层采用卷积核大小为 $3 \times 3$ , 步长为1的卷积, 池化层卷积核大小为 $2 \times 2$ , 步长为2, 残差模块包括两层卷积核大小为 $3 \times 3$ , 步长为1的卷积层和一个卷积核大小为 $2 \times 2$ , 步长为2的池化层; 其中全连接层由两层线性连接层组成, 第一层线性层输入维度为2000, 输出维度为512, 第二层线性层输入维度为512, 输出维度为7; 每一层卷积层后面都使用Leaky Relu作为激活函数, 并接上BN层。

10. 根据权利要求1所述的一种基于动态目标训练的深度学习人脸表情识别方法,

其特征在于,步骤S4中,根据损失函数的等级类别分阶段训练搭建的神经网络;深度神经网络的损失函数为:

$$L=L_{cls}+\lambda_{1kd}L_{1kd}+\lambda_{di}L_{di}+\lambda_{L2}L_2;$$

其中, $\lambda_{1kd}$ 、 $\lambda_{di}$ 、 $\lambda_{L2}$ 分别为似然正则化损失函数、基于表情识别领域知识的损失函数、L2正则项损失函数的权重; $L_{di}$ 为基于表情识别领域知识的损失函数Ldomain\_information;

然后通过权重来实现三个训练阶段,而每个训练阶段损失函数与步骤S3中损失函数的排序一致;

整个网络模型训练过程划分为三个不同的阶段,每个阶段损失项的加权系数都会随着训练过程动态上升,通过动态调整加权系数的值在不同阶段达到峰值,使得每一个训练阶段都有网络模型要重点关注的损失项;

在第一阶段中,给 $\lambda_{1kd}$ 、 $\lambda_{di}$ 以及 $\lambda_{L2}$ 很小的初始值, $\lambda_{1kd}$ 取[0.001,0.005], $\lambda_{di}$ 取[0.01,0.05], $\lambda_{L2}$ 取[0.00001,0.00005],使得网络模型首先着重训练Lcls确保网络模型尽快的拟合数据的真实分布提高网络模型分类准确率;

在第二阶段中,通过动态调整 $\lambda_{1kd}$ 和 $\lambda_{di}$ 加权系数的值达到预设最大值,使网络模型在这一阶段着重训练似然正则损失项和领域信息损失项,目标是使得类内的样本更加紧凑的同时拉大易混分类别特征空间的距离;

在第三阶段中,通过动态调整 $\lambda_{L2}$ 加权系数值到预设最大值,使网络模型在这一阶段着重训练L2损失项,让模型更简单泛化能力更强;

三个损失项的动态加权系数 $\lambda_{dynamic}$ 的通用公式化形式如下:

$$\lambda_{dynamic}=\min(\lambda_{initial}10^{3\cdot epoch/\max\_epoch},\max\_value)$$

其中 $\lambda_{initial}$ 为不同损失项的初始权重系数,epoch是当前迭代轮数,max\_epoch是总迭代轮数,max\_value为损失项加权系数的预设最大值;训练完成后,得到表情识别模型。

## 一种基于动态目标训练的深度神经网络人脸表情识别方法

### 技术领域

[0001] 本发明涉及深度学习机器视觉领域,具体涉及一种基于动态目标训练的深度神经网络人脸表情识别方法。

### 背景技术

[0002] 人脸表情是人们用来表达情绪的最主要的一种方式,研究表明:人与人之间的交流仅有7%的信息是靠言语来传达的,38%的信息是靠声音来传递的,剩余的55%均是靠人脸的面部表情来进行传达。由此结论可以看出人脸表情在人与人交流过程中的重要性,因此人脸面部表情识别有着重要的研究价值和意义。

[0003] 人脸表情识别目前在社会生活中有着广泛的应用场景,比如在智能安防、测谎、智慧医疗、互联网教育等方面。在智能安防领域中,人脸表情识别技术可以实时监测路人的情绪状态,识别可疑人物等潜在风险并提前采取必要的防护措施,在银行、商城、小区等公共场所有着很大的用途。在智慧医疗领域中,人脸表情识别技术可以24小时不间断的获取病人的面部表情图像进而判断病人当前的状态,当病人出现紧急危险情况时可以第一时间反馈给医生做好及时治疗。在远程互联网教育中,人脸表情识别对于老师在线及时了解学习的听课状态并根据学生当前的情绪状态来动态调整授课的形式和内容,进而提升教学质量。

[0004] 由于表情具有复杂性和多变性的特点,加上不同种族,年龄,周围环境等因素的干扰(Li,S.;Deng,W.Deep Facial Expression Recognition:A Survey.In Proceedings of the International Conference on Computer Vision and Pattern Recognition,Salt Lake City,UT,USA,18-22June 2018.),很难针对数据集去明确的划分每一个表情的特征空间,表情类别中的大部分样本的面部特征在特征空间中距离同类样本的可能比较近,而某些表情类别中的样本特征空间距离同类样本比较远,仅仅使用图像分类中具有普适性的分类损失函数引导网络学习使得网络泛化性能不足。虽然有一些先进的方法提出来解决这个问题,比如Jie Cai et al提出的概率属性树方法(Cai J,Meng Z,Khan AS,et al.Probabilistic Attribute Tree in Convolutional Neural Networks for Facial Expression Recognition[J].Computer Vision and Pattern Recognition,2018,1(c):10-24.)和Yong Li et al提出的注意力方法(Li Y,Zeng J,Shan S,et al.Occlusion aware facial expression recognition using CNN with attention mechanism[J].IEEE Transactions on Image Processing,IEEE,2018,28(5):2439-2450.),但它们更为复杂且精度较低。

### 发明内容

[0005] 本发明为了解决现有技术的不足,提出了一种基于动态目标训练的深度神经网络人脸表情识别方法,有效提高对人脸表情识别分类的能力,同时具备较强的鲁棒性。

[0006] 本发明的目的至少通过如下技术方案之一实现。

[0007] 一种基于动态目标训练的深度学习人脸表情识别方法,包括以下步骤:

[0008] S1、输入训练样本集和测试样本集;

[0009] S2、采用多个损失函数作为训练的目标函数;

[0010] S3、将损失函数排序;

[0011] S4、根据排序后的损失函数依次在训练样本集上重复对神经网络进行动态目标训练,获得表情识别模型;

[0012] S5、根据表情识别模型对输入的测试样本进行表情分类。

[0013] 进一步地,步骤S1中,对输入的任何人脸图像,属于三通道的彩色图像,通过双线性插值的方式进行放缩,每张图像进行随机的水平翻转、随机平移、随机图像旋转、随机图像缩放。

[0014] 进一步地,步骤S2中损失函数包括最大间隔高斯混合损失函数L-GM、似然正则化损失函数、基于表情分类领域知识的损失函数和L<sub>2</sub>正则项损失函数。

[0015] 进一步地,L-GM损失函数在训练集上提取的深度特征x遵循高斯混合概率分布,其中p(k)代表深度特征x的第k类的先验概率, $\mu_k$ 和 $\Sigma_k$ 分别表示类别k的均值和协方差,公式如下:

$$[0016] \quad p(x) = \sum_{k=1}^K N(x; \mu_k, \Sigma_k) p(k);$$

[0017] 其中N表示高斯分布函数,K表示分类的总类别个数;

[0018] 基于上述约束,训练集上提取的深度特征 $x_i$ 在给定的类别标签 $z_i \in [1, K]$ 下的条件概率的计算如下所示:

$$[0019] \quad p(x_i | z_i) = N(x_i; \mu_{z_i}, \Sigma_{z_i});$$

[0020] 相应的后验概率分布的计算如下所示:

$$[0021] \quad p(z_i | x_i) = \frac{N(x_i; \mu_{z_i}, \Sigma_{z_i}) p(z_i)}{\sum_{k=1}^K N(x_i; \mu_k, \Sigma_k) p(k)};$$

[0022] 混合高斯分布的类别损失函数 $L_{cls}$ 通过后验概率分布与类别标签的交叉熵来计算,公式如下:

$$[0023] \quad \begin{aligned} L_{cls} &= -\frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K 1(z_i = k) \log p(k | x_i) \\ &= -\frac{1}{N} \sum_{i=1}^N \log \frac{N(x_i; \mu_{z_i}, \Sigma_{z_i}) p(z_i)}{\sum_{k=1}^K N(x_i; \mu_k, \Sigma_k) p(k)}; \end{aligned}$$

[0024] 转换用距离来度量后的混合高斯分布的类别损失函数 $L_{cls}$ ,其中 $d_k$ 的计算采用马氏距离来衡量相关特征间的距离,具体公式如下:

$$[0025] \quad L_{cls,i} = -\log \frac{p(z_i) |\Sigma_{z_i}|^{-\frac{1}{2}} e^{-d_{z_i}}}{\sum_k p(k) |\Sigma_k|^{-\frac{1}{2}} e^{-d_k}};$$

$$[0026] \quad d_k = (x_i - \mu_k)^T \sum_k^{-1} (x_i - \mu_k) / 2;$$

[0027] 其中,T表示矩阵转置,e为自然常数, $L_{cls,i}$ 表示第i个样本的类别损失函数的大小。

[0028] 进一步地,似然正则化损失函数用于测量训练样本在何种程度上符合假设的混合

高斯分布,通过对最大似然估计取负来进行计算,公式如下:

$$[0029] \quad L_{lkd} = - \sum_{i=1}^N \log N(x_i; \mu_{zi}, \Sigma_{zi})。$$

[0030] 进一步地,基于表情识别领域知识的损失函数的目标是在人脸表情样本在距离自身类中心距离尽可能小的同时引入先验领域信息知识拉大其与易混分表情类中心的距离;领域损失函数公式如下:

$$[0031] \quad L_{domain\_information} = \sum_i^N [ \|f(x_i) - f(x_i^s)\|_2^2 - \|f(x_i) - f(x_i^m)\|_2^2 + \alpha ]_+;$$

[0032] 其中参数 $\alpha$ 的作用是在网络模型的学习中加入强约束用来定义训练样本特征与同类中心特征间距离和对应易混分类中心特征间距离的一个最小间隔; $f(x_i)$ 代表训练样本 $x_i$ 的特征表示, $f(x_i^s)$ 代表训练样本 $x_i$ 的同类表情类别中心的特征表示, $f(x_i^m)$ 表示训练样本 $x_i$ 对应易混分表情标签类别中心的特征表示;+代表中括号内的值大于0时取该值作为损失值,当中括号内的值小于0时损失值为0。

[0033] 进一步地,正则项就是稀疏性,以保证抽取的特征尽可能简单; $L_2$ 正则项损失函数的公式如下:

$$[0034] \quad L_2 = \sum_{i=1}^M w_i^2$$

[0035] 其中M代表神经网络的参数个数, $w_i$ 是模型用于更新的第i个参数。

[0036] 进一步地,步骤S3中,依据损失函数对模型的作用将四个损失函数排序分为三个等级,第一等级:分类损失函数;第二等级:似然正则化损失函数,基于表情识别领域知识的损失函数;第三等级: $L_2$ 正则项损失函数。

[0037] 进一步地,步骤S4中,搭建深度神经网络模型,依次设置为卷积层,池化层,3层残差模块,协方差层和全连接层;其中卷积层采用卷积核大小为 $3*3$ ,步长为1的卷积,池化层卷积核大小为 $2*2$ ,步长为2,残差模块包括两层卷积核大小为 $3*3$ ,步长为1的卷积层和一个卷积核大小为 $2*2$ ,步长为2的池化层;其中全连接层由两层线性连接层组成,第一层线性层输入维度为2000,输出维度为512,第二层线性层输入维度为512,输出维度为7;每一层卷积层后面都使用Leaky Relu作为激活函数,并接上BN层。

[0038] 进一步地,步骤S4中,根据损失函数的等级类别分阶段训练搭建的深度神经网络;深度神经网络的损失函数为:

$$[0039] \quad L = L_{cls} + \lambda_{lkd} L_{lkd} + \lambda_{di} L_{di} + \lambda_{L2} L_2;$$

[0040] 其中, $\lambda_{lkd}$ 、 $\lambda_{di}$ 、 $\lambda_{L2}$ 分别为似然正则化损失函数、基于表情识别领域知识的损失函数、 $L_2$ 正则项损失函数的权重; $L_{di}$ 为基于表情识别领域知识的损失函数 $L_{domain\_information}$ ;

[0041] 然后通过权重来实现三个训练阶段,而每个训练阶段损失函数与步骤S3中损失函数的排序一致;

[0042] 整个网络模型训练过程划分为三个不同的阶段,每个阶段损失项的加权系数都会随着训练过程动态上升,通过动态调整加权系数的值在不同阶段达到峰值,使得每一个训

练阶段都有网络模型要重点关注的损失项；

[0043] 在第一阶段中，给 $\lambda_{1kd}$ ， $\lambda_{di}$ 以及 $\lambda_{L2}$ 很小的初始值， $\lambda_{1kd}$ 取 $[0.001, 0.005]$ ， $\lambda_{di}$ 取 $[0.01, 0.05]$ ， $\lambda_{L2}$ 取 $[0.00001, 0.00005]$ ，使得网络模型首先着重训练 $L_{cls}$ 确保网络模型最快的拟合数据的真实分布提高网络模型分类准确率；

[0044] 在第二阶段中，通过动态调整 $\lambda_{1kd}$ 和 $\lambda_{di}$ 加权系数的值达到预设最大值，使网络模型在这一阶段着重训练似然正则损失项和领域信息损失项，目标是使得类内的样本更加紧凑的同时拉大易混分类别特征空间的距离；

[0045] 三个损失项的动态加权系数 $\lambda_{dynamic}$ 的通用公式化形式如下：

[0046]  $\lambda_{dynamic} = \min(\lambda_{initial} 10^{3 \cdot \text{epoch}/\text{max\_epoch}}, \text{max\_value})$

[0047] 其中 $\lambda_{initial}$ 为不同损失项的初始权重系数，epoch是当前迭代轮数，max\_epoch是总迭代轮数，max\_value为损失项加权系数的预设最大值；训练完成后，得到表情识别模型。

[0048] 本发明相对于现有技术具有如下的优点及效果：

[0049] 1. 本发明更有机会避免模型的参数陷入局部最优（请说明此处是避免什么陷入局部最优），提高了模型的表情识别泛化性能。

[0050] 2. 本发明提高了表情识别的准确率，有更广泛的应用。

## 附图说明

[0051] 图1是本发明实施例中一种基于动态目标训练的深度学习人脸表情识别方法的流程图；

[0052] 图2是本发明实施例中一种基于动态目标训练的深度学习人脸表情识别方法的神经网络结构图。

## 具体实施方式

[0053] 为使本发明实施例的目的、技术方案和优点更加清楚，下面将结合本发明实施例以及附图，对本发明的具体实施进行清楚、完整地描述，显然，所描述的实施例是本发明一部分实施例，而不是全部的实施例。

[0054] 实施例：

[0055] 一种基于动态目标训练的深度学习人脸表情识别方法，如图1所示，包括以下步骤：

[0056] S1、输入训练样本集和测试样本集。

[0057] 本实施例中，对输入的任何人脸图像，属于三通道的彩色图像，通过双线性插值的方式放缩到 $100 \times 100$ 尺寸大小的图片，每张图像随机的水平翻转、随机平移正负0.1的宽度或正负0.1的高度、随机图像旋转正负30度、随机图像缩放0.2的比例。

[0058] S2、采用多个损失函数作为训练的目标函数，包括最大间隔高斯混合损失函数L-GM、似然正则化损失函数、基于表情分类领域知识的损失函数和 $L_2$ 正则项损失函数。

[0059] L-GM损失函数在训练集上提取的深度特征 $x$ 遵循高斯混合概率分布，其中 $p(k)$ 代表深度特征 $x$ 的第 $k$ 类的先验概率， $\mu_k$ 和 $\Sigma_k$ 分别表示类别 $k$ 的均值和协方差，公式如下：

[0060] 
$$p(x) = \sum_{k=1}^K N(x; \mu_k, \Sigma_k) p(k) ;$$

[0061] 其中 $N$ 表示高斯分布函数， $K$ 表示分类的总类别个数。



[0062] 基于上述约束,训练集上提取的深度特征 $x_i$ 在给定的类别标签 $z_i \in [1, K]$ 下的条件概率的计算如下所示:

[0063]  $p(x_i | z_i) = N(x_i; \mu_{z_i}, \Sigma_{z_i})$ ;

[0064] 相应的后验概率分布的计算如下所示:

$$[0065] \quad p(z_i | x_i) = \frac{N(x_i; \mu_{z_i}, \Sigma_{z_i}) p(z_i)}{\sum_{k=1}^K N(x_i; \mu_k, \Sigma_k) p(k)};$$

[0066] 混合高斯分布的类别损失函数 $L_{cls}$ 通过后验概率分布与类别标签的交叉熵来计算,公式如下:

$$[0067] \quad L_{cls} = -\frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K 1(z_i = k) \log p(k | x_i) \\ = -\frac{1}{N} \sum_{i=1}^N \log \frac{N(x_i; \mu_{z_i}, \Sigma_{z_i}) p(z_i)}{\sum_{k=1}^K N(x_i; \mu_k, \Sigma_k) p(k)};$$

[0068] 转换用距离来度量后的混合高斯分布的类别损失函数 $L_{cls}$ ,其中 $d_k$ 的计算采用马氏距离来衡量相关特征间的距离,具体公式如下:

$$[0069] \quad L_{cls,i} = -\log \frac{p(z_i | \Sigma_{z_i}) |\Sigma_{z_i}|^{-\frac{1}{2}} e^{-d_{z_i}}}{\sum_k p(k) |\Sigma_k|^{-\frac{1}{2}} e^{-d_k}};$$

$$[0070] \quad d_k = (x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k) / 2;$$

[0071] 其中, $T$ 表示矩阵转置, $e$ 为自然常数; $L_{cls,i}$ 表示第 $i$ 个样本的类别损失函数的大小。

[0072] 似然正则化损失函数用于测量训练样本在何种程度上符合假设的混合高斯分布,通过对最大似然估计取负来进行计算,公式如下:

$$[0073] \quad L_{lkd} = -\sum_{i=1}^N \log N(x_i; \mu_{z_i}, \Sigma_{z_i})。$$

[0074] 基于表情识别领域知识的损失函数的目标是在人脸表情样本在距离自身类中心距离尽可能小的同时引入先验领域信息知识拉大其与易混分表情类中心的距离;领域损失函数公式如下:

$$[0075] \quad L_{domain\_information} = \sum_i^N [||f(x_i) - f(x_i^s)||_2^2 - ||f(x_i) - f(x_i^m)||_2^2 + \alpha]_+;$$

[0076] 其中参数 $\alpha$ 的作用是在网络模型的学习中加入强约束用来定义训练样本特征与同类中心特征间距离和对应易混分类中心特征间距离的一个最小间隔; $f(x_i)$ 代表训练样本 $x_i$ 的特征表示, $f(x_i^s)$ 代表训练样本 $x_i$ 的同类表情类别中心的特征表示, $f(x_i^m)$ 表示训练样本 $x_i$ 对应易混分表情标签类别中心的特征表示; $+$ 代表中括号内的值大于0时取该值作为损失值,当中括号内的值小于0时损失值为0。

[0077] 正则项就是稀疏性,以保证抽取的特征尽可能简单; $L_2$ 正则项损失函数的公式如下:

[0078] 
$$L_2 = \sum_{i=1}^M w_i^2$$

[0079] 其中M代表神经网络的参数个数,wi是模型用于更新的第i个参数。

[0080] S3、将损失函数排序。本实施例中,将损失函数分成三个等级,从第一个等级开始训练。

[0081] 第一等级:分类损失函数;

[0082] 第二等级:似然正则化损失函数,基于表情识别领域知识的损失函数;

[0083] 第三等级:L<sub>2</sub>正则项损失函数。

[0084] S4、根据排序后的损失函数依次在训练样本集上重复对神经网络进行动态目标训练,获得表情识别模型。

[0085] 首先搭建深度神经网络模型,如图2所示,依次设置为卷积层,池化层,3层残差模块,协方差层和全连接层;其中卷积层采用卷积核大小为3\*3,步长为1的卷积,池化层卷积核大小为2\*2,步长为2,残差模块包括两层卷积核大小为3\*3,步长为1的卷积层和一个卷积核大小为2\*2,步长为2的池化层。其中全连接层由两层线性连接层组成,第一层线性层输入维度为2000,输出维度为512,第二层线性层输入维度为512,输出维度为7;每一层卷积层后面都使用Leaky Relu作为激活函数,并接上BN层。

[0086] 每一层的具体参数配置如表1所示。

[0087] 表1.深度神经网络结构参数表

[0088]

网络层	网络参数				Active fun	BN
	Kernel Size	stride	padding	Kernel num		
卷积层 1	3*3	1	SAME	64	Leaky ReLU	Yes
池化层 2	2*2	2	VALID	64	Leaky ReLU	Yes
卷积层 3	3*3	1	SAME	96	Leaky ReLU	Yes
卷积层 4	3*3	1	SAME	96	Leaky ReLU	Yes
池化层 5	2*2	2	VALID	96	-	-
卷积层 6	3*3	1	SAME	128	Leaky ReLU	Yes
卷积层 7	3*3	1	SAME	128	Leaky ReLU	Yes
卷积层 8	3*3	1	SAME	128	Leaky ReLU	Yes
池化层 9	2*2	2	SAME	128	-	-
卷积层 10	3*3	1	SAME	256	Leaky ReLU	Yes
卷积层 11	3*3	1	SAME	256	Leaky ReLU	Yes
卷积层 12	3*3	1	SAME	256	Leaky ReLU	Yes
池化层 13	2*2	2	SAME	256	-	-
协方差层	-	-	-	-	-	-

全连接层	输入 2000 维, 输出 512 维
------	---------------------

[0089]

全连接层	输入 512 维, 输出 7 维
------	------------------

[0090] 然后根据损失函数的等级类别分阶段训练搭建的深度神经网络;深度神经网络的

损失函数为：

[0091]  $L=L_{c1s}+\lambda_{1kd}L_{1kd}+\lambda_{di}L_{di}+\lambda_{L2}L_2$ ;

[0092] 其中,  $\lambda_{1kd}$ 、 $\lambda_{di}$ 、 $\lambda_{L2}$ 分别为似然正则化损失函数、基于表情识别领域知识的损失函数、 $L_2$ 正则项损失函数的权重;  $L_{di}$ 为基于表情识别领域知识的损失函数 $L_{domain\_information}$ ;

[0093] 然后通过权重来实现三个训练阶段,而每个训练阶段损失函数与步骤S3中损失函数的排序一致;

[0094] 整个网络模型训练过程划分为三个不同的阶段,每个阶段损失项的加权系数都会随着训练过程动态上升,通过动态调整加权系数的值在不同阶段达到峰值,使得每一个训练阶段都有网络模型要重点关注的损失项;

[0095] 在第一阶段中,给 $\lambda_{1kd}$ 、 $\lambda_{di}$ 以及 $\lambda_{L2}$ 很小的初始值,本实施例中, $\lambda_{1kd}$ 取0.003, $\lambda_{di}$ 取0.03, $\lambda_{L2}$ 取0.00003,使得网络模型首先着重训练 $L_{c1s}$ 确保网络模型尽快的拟合数据的真实分布提高网络模型的分类准确率;

[0096] 在第二阶段中,通过动态调整 $\lambda_{1kd}$ 和 $\lambda_{di}$ 加权系数的值达到预设最大值,使网络模型在这一阶段着重训练似然正则损失项和领域信息损失项,目标是使得类内的样本更加紧凑的同时拉大易混分类别特征空间的距离;

[0097] 在第三阶段中,通过动态调整 $\lambda_{L2}$ 加权系数值到预设最大值,使网络模型在这一阶段着重训练 $L_2$ 损失项,让模型更简单泛化能力更强;

[0098] 三个损失项的动态加权系数 $\lambda_{dynamic}$ 的通用公式化形式如下:

[0099]  $\lambda_{dynamic}=\min(\lambda_{initial}10^{3\cdot epoch/max\_epoch},max\_value)$

[0100] 其中 $\lambda_{initial}$ 为不同损失项的初始权重系数,epoch是当前迭代轮数,max\_epoch是总迭代轮数,max\_value为损失项加权系数的预设最大值;

[0101] 本实施例中,训练过程中总迭代轮数为600轮,第一阶段(0-200轮)网络训练的注意力集中在分类损失函数 $L_{c1s}$ 上,其余损失函数的加权系数初始化值均很小,将损失项加权系数的初始值设置为 $\lambda_{initial\_1kd}=0.003$ , $\lambda_{initial\_di}=0.03$ , $\lambda_{initial\_L2}=3e-5$ 。第二阶段(200-400轮)动态调整加权系数使 $\lambda_{1kd}$ 和 $\lambda_{di}$ 加权系数大约在300轮时达到预设最大值,其中似然正则项加权系数 $\lambda_{1kd}$ 的最大值为0.1,领域信息损失项加权系数 $\lambda_{di}$ 的最大值为1.0,这一阶段网络模型特征学习的注意力集中在使类内的样本更加紧凑同时拉大易混分类别间的距离。第三个阶段(400-600轮)使 $\lambda_{L2}$ 加权系数大约在500轮时达到最大值, $L_2$ 正则项加权系数的预设最大值为 $1e-2$ ;

[0102] 训练完成后,得到表情识别模型。

[0103] S5、根据表情识别模型对输入的测试样本进行表情分类。

[0104] 依据本发明的实现在RAF-DB(Li S,Deng W,Du J.Reliable Crowdsourcing and Deep Locality-Preserving Learning for Expression Recognition in the Wild[J].2017IEEE Conference on Computer Vision and Pattern Recognition(CVPR),2017:2852-2861.)测试集上与其他方法比较,如表2所示,本发明优于本领域其他方法,包括概率属性树方法(Cai J,Meng Z,Khan AS,et al.Probabilistic Attribute Tree in Convolutional Neural Networks for Facial Expression Recognition[J].Computer Vision and Pattern Recognition,2018,1(c):10-24.)以及注意力机制方法(Li Y,Zeng J,Shan S,et al.Occlusion aware facial expression recognition using CNN with

attention mechanism[J].IEEE Transactions on Image Processing,IEEE,2018,28(5):2439-2450.)。

[0105] 表2.RAF-DB测试集上不同先进方法平均识别准确率比较

	Systems	Recognition rate
[0106]	注意力机制方法	85.01%
	概率属性树方法	84.19%
[0107]	本发明	<b>86.47%</b>

[0108] 上述实施例为本发明较佳的实施方式,但本发明的实施方式并不受上述实施例的限制,其他的任何未背离本发明的精神实质与原理下所作的改变、修饰、替代、组合、简化,均应为等效的置换方式,都包含在本发明的保护范围之内。



图1

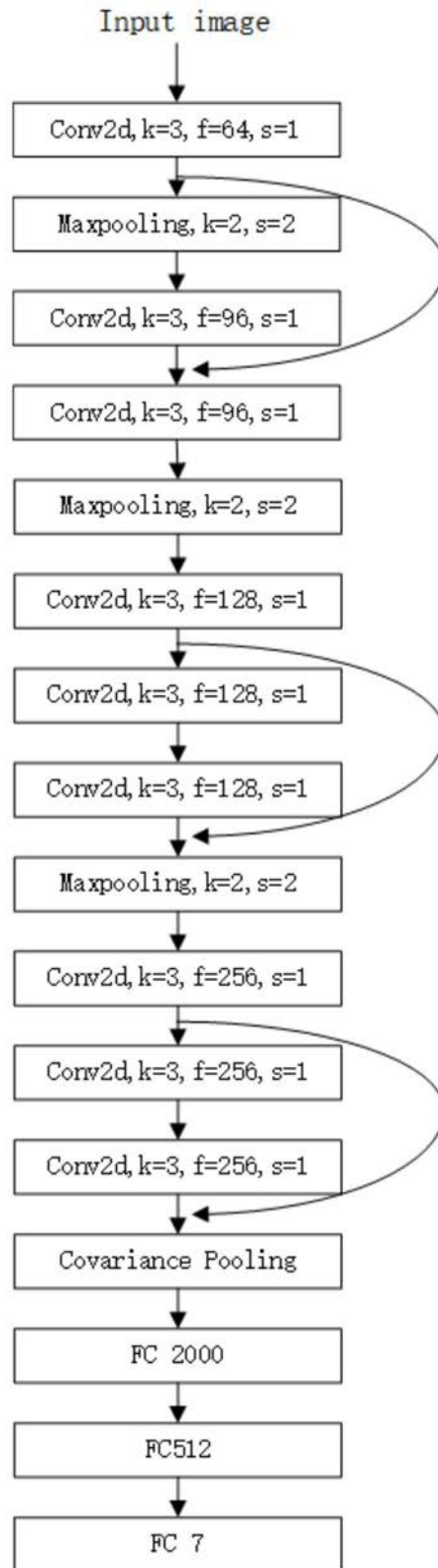


图2