

一种孤立中心损失方法及其在人脸表情识别中的应用

王军杰¹, 王泉¹, 蒋平², 刘音¹

(1. 西安电子科技大学计算机科学与技术学院, 710071, 西安; 2. 榆林学院信息工程学院, 719000, 陕西榆林)

摘要: 针对 Softmax 损失监督下各类样本之间区分度不足的问题, 提出了一种孤立中心损失监督方法。基于类间离散度尽量大、类内离散度尽量小的原则, 提出方法由 3 部分组成: 采用等角分布固定权值, 使得全部类间夹角余弦值之和最小, 确保不同类别在角度空间的距离最大化; 中心聚类思想, 最小化每个样本与其所属类别的中心之间的欧氏距离, 促使同类样本尽量聚拢; 最大化不同类之间的欧氏距离, 使得不同类样本在欧氏空间尽量分开。在表情数据集 FER2013、FERPlus 和 RAF-DB 上的测试结果显示: 提出方法的平均准确率分别达到了 73.02%、88.56% 和 86.26%, 相比于 Softmax 损失函数, 分别提升了 1.25%、0.44% 和 0.65%; 同时, 提出方法比 Softmax 损失更加稳定(相同配置下多次实验结果的变化程度更小); 提出方法的运行速度只比 Softmax 损失方法略慢一点, 仍然比一些其他方法快。

关键词: 卷积神经网络; 监督学习; 损失函数; 人脸表情识别

中图分类号: TP391.4 **文献标志码:** A

DOI: 10.7652/xjtub202204013 **文章编号:** 0253-987X(2022)04-0119-08



OSID 码

An Isolated Central Loss Method Applied in Facial Expression Recognition

WANG Junjie¹, WANG Quan¹, JIANG Ping², LIU Yin¹

(1. School of Computer Science and Technology, Xidian University, Xi'an 710071, China;

2. School of Information Engineering, Yulin University, Yulin, Shaanxi 719000, China)

Abstract: An isolated central loss supervision method is proposed in the paper due to obscure distinction among various samples under Softmax loss supervision. Based on the principle of "maximum inter-class statistical dispersion and minimum intra-class statistical dispersion", the method consists of three parts: The fixed weight of isometric distribution is adopted to minimize the sum of all inter-class angle cosine values and ensure the maximum distance between different classes in angular space; The idea of central clustering is adopted to minimize the Euclidean distance between each sample and the center of its class, so that samples in the same class would cluster as much as possible; It aims to maximize the Euclidean distance among different classes, that is, samples of different classes are separated from each other to the greatest extent in the Euclidean space. The test results on the facial expression datasets FER2013, FERPlus and RAF-DB show that the average accuracy of the isolated central loss supervision method reaches 73.02%, 88.56% and 86.26%, respectively, with an increase of 1.25%, 0.44% and 0.65% compared with the Softmax loss function. At the same time, the method is more stable than

收稿日期: 2021-09-04。 作者简介: 王军杰(1996—), 男, 硕士生; 王泉(通信作者), 男, 教授, 博士生导师。 基金项目: 国家自然科学基金资助项目(61972302); 陕西省科技厅重点研发计划资助项目(2020NY-167); 榆林市科技计划资助项目(CXY-2020-013-04)。

网络出版时间: 2021-12-20

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1069.T.20211218.0907.002.html>

Softmax loss (under the same configuration conditions, the multiple experiments results vary less). The method takes longer than the Softmax loss, but it is still faster than other measures.

Keywords: convolutional neural network; supervision and learning; loss function; facial expression recognition

人脸表情是一种重要且自然的情绪表现形式,因而人脸表情自动识别也是人机和谐交互的重要基础。近三十年来,人脸表情识别(FER)研究一直是人工智能领域的一个热点^[1-3]。近些年,随着深度学习技术发展,FER研究从实验室限制条件下的表情样本逐渐转移到了现实世界(自然)表情图像。自然表情图像复杂多变,如光照强度、头部姿态变化、有无遮挡物等,导致表情识别的难度增加^[4-6]。

FER是一个典型的多分类任务,每一个样本必归属某一个类别,且只属于一个类别。Softmax Loss (SL)是一种交叉熵损失,经过 Softmax 变换,每一个样本最终输出对应多个类别的概率值(相对概率)。SL通过优化预测概率与目标概率之间的误差,会持续拉高正确类别的概率和降低错误分类的概率。在多分类问题上,SL直观、易理解,而且反向求导简洁,是 FER 领域应用最广泛的一种监督函数。

SL尽力把不同类别样本特征在角度空间分开,在 SL 监督下学习到的深度特征的区分度不够好^[7],因此研究者们对 SL 进行了改进^[8-10]。针对 SL 的改进技术主要分两类:一类是改变 SL 本身(即 SL 的变体),例如人脸识别领域中的 ArcFace^[7]、Normface^[11]和 CosFace^[12]等;另一类是增加约束项协同 SL 一起工作,如 Center loss (CL)^[13-14]和 Island Loss (IL)^[12]等。

分类任务有个共同原则,深度特征的区分度越高越好,即不同类别样本特征之间的距离(差异性)越大越好。现有技术主要通过减小类内距离和增大不同类之间的距离两个方面来提高特征区分度。

CL采用中心聚类思想,减小同类样本之间的欧式距离,有效降低类内变化。Luo等^[16]认为同一种表情并非高度相似,同一种表情中也应该分成几个子类,从而提出了 Subcenter Loss (SCL)。Li等^[17-18]则减小每个样本与其相邻几个同类样本中心的距离,从而促使同类样本收敛在一起,这种方法兼具了 CL、SCL 的优点,但是计算复杂度高,且需要额外的内存开销。

IL在SL的基础上采用了CL,同时通过最小化不同类中心夹角的余弦值,期望不同类之间的夹角尽量大。IL需要约束每一类与其他类的夹角,难以

处理类别数量多的任务,Zhao等^[19]用于人脸识别的“排他正则化”本质上是 IL 的一种特例。Jiang等^[20]证明了类间夹角相等且等于 $\arccos(1/(1-n))$ (n 为类别数量)是类间角度距离最大化的一种理想状态,基于此提出了一种改进 Softmax 损失 (ASL)。

ASL能够有效最大化类间角度距离,CL能够有效地减小类内变化,但是 ASL、CL 单纯地协同工作,即 ASL+CL 并非更优秀,可能比 ASL 更差。针对此问题,本文提出了一种孤立中心聚类损失 (ICL)方法,以 ASL 为基础融合中心聚类思想,即改进 CL 以符合 ASL,再最大化类间欧式距离,使各类成为彻底的“孤岛”。实验结果表明,所提方法在识别精度和稳定性方面都有较好的提升。

1 相关工作

1.1 改进 Softmax 损失函数

无偏量的 SL 一般定义为

$$L_{\text{SL}} = -\frac{1}{m} \sum_{i=1}^m \log \frac{e^{w_{y_i}^T x_i}}{\sum_{k \in D} e^{w_k^T x_i}} \quad (1)$$

式中: $D=\{1, \dots, n\}$ 是样本类别标签集合; $(\cdot)^T$ 代表转置操作; $x_i \in \mathbb{R}^d$ 表示第 i 个深度特征向量; $y_i \in D$ 是对应的类别标签; $w_k \in \mathbb{R}^d$ 是对应第 k 类的权值参数向量;SL 中全体权值参数表示为 $W=[w_1, w_2, \dots, w_n] \in \mathbb{R}^{d \times n}$, n 和 m 分别表示类别数目和训练样本数量(一批)。

一个训练理想的模型,第 k 类特征的中心就在其对应权值向量 w_k 的方向上,即 w_k 可以代表第 k 类中心的方向。若用 A_{kl} 表示第 k 和 l 类(中心)的夹角,则 A_{kl} 的余弦值为

$$\cos(A_{kl}) = \frac{w_k^T w_l}{\|w_k\| \|w_l\|} = \hat{w}_k^T \hat{w}_l \quad (2)$$

式中: $\hat{w}_k$ 和 $\hat{w}_l$ 为单位化的 w_k 和 w_l ; $\|\cdot\|$ 表示向量的模。余弦值 $\cos(A_{kl})$ 越小,夹角 A_{kl} 越大。

基于类间角度距离(夹角)越大,区分度越高的原则,考虑如下损失函数

$$L_{\text{ang}} = \sum_{k \in D} \sum_{l \in D, l \neq k} \hat{w}_k^T \hat{w}_l \quad (3)$$

式中: L_{ang} 为所有夹角的余弦值之和, L_{ang} 越小表示

各类之间的夹角越大,区分度越好。

深度特征维数 $d \geq n-1$ 时, $\hat{\mathbf{w}}_k^T \hat{\mathbf{w}}_l = 1/(1-n)$ 可以使 L_{ang} 取得极小值 $-n$ (即 $L_{\text{ang}} = -n$), 即每个夹角 $A_{kl} = \arccos(1/(1-n))$ 是最小化 L_{ang} 的一种理想解。 A_{kl} 仅仅是 $L_{\text{ang}} = -n$ 的一个解, 而不是唯一解, 所以通过梯度下降方法优化 L_{ang} 损失不能确保每一个夹角 A_{kl} 都满足 $A_{kl} = \arccos(1/(1-n))$ 。文献[20]给出了一个自动生成权值参数的算法, 能够保证每一个夹角 $A_{kl} = \arccos(1/(1-n))$, 这是 ASL 的核心思想。

ASL 可定义为

$$L_{\text{SL}} = -\frac{1}{m} \sum_{i=1}^m \log \frac{\mathbf{e}^{\mathbf{w}_{y_i}^T \mathbf{x}_i}}{\sum_{k \in D} \mathbf{e}^{\mathbf{w}_k^T \mathbf{x}_i}} \quad (4)$$

$$\text{s.t. } \|\mathbf{w}_k\| = 1 \text{ and } \mathbf{w}_k^T \mathbf{w}_l = \frac{1}{1-n} \quad (k, l \in D)$$

对比式(1)、式(4), ASL 只是对 SL 的权值做了限制。ASL 在初始化时, 权值向量 $\mathbf{w}_k$ 由文献[20]的算法 1 生成^①, 在训练时固定不变。

ASL 的优势有两点: ASL 的权值向量等角度分布, 可以缓解多样本类容易挤压少样本类角度空间的问题, 在一定程度上可以解决样本不平衡的问题; ASL 采用的权值是固定的, 在训练过程中不需要更新, 所以 ASL 的训练速度比 SL 快。

1.2 中心聚类损失函数

针对 SL 只注重类间差异, 忽略类内差异性过大的问题, 文献[13]提出了一个减小类内离散度的约束函数, 即著名的中心聚类损失函数 (center loss, CL), 其定义为

$$L_{\text{CL}} = \frac{1}{2m} \sum_{i=1}^m \|\mathbf{x}_i - \mathbf{c}_{y_i}\|^2 \quad (5)$$

式中 $\mathbf{c}_{y_i}$ 表示第 y_i 类的中心。

CL 是对 SL 的一种辅助, 与 SL 一起监督学习, 学习时总损失可以表示为

$$L = L_{\text{SL}} + \lambda L_{\text{CL}} \quad (6)$$

式中 λ 是一个权值, 决定了 CL 约束效果。 $\lambda=0$ 时, CL 无效; λ 过大, 则同类样本有可能被聚集在同一个点, 这也不利于分类, 因为同类样本特征不可能完全相同, 同类样本之间也存在细微差别。

CL 的核心思想是减小同类样本之间的 (欧氏) 空间距离, 让同类样本尽量聚集在一起, 从而增大不同类样本之间的距离, 提升区分度。

2 孤立中心损失函数

增大类间差异、减小类内差异是监督学习提高

区分度的基本原则。ASL 本质上是类间角度距离最大化的结果, 本文以 ASL 为基础, 再在欧式空间减小类间离散度和增大类间距离, 提出一种孤立中心损失方法 (ICL)。

2.1 前向传播

ASL 已经假定了第 k 类的中心在其对应的权值向量 $\mathbf{w}_k$ 方向上, 因而可以设第 k 类的中心为 $\gamma_k \mathbf{w}_k$ (γ_k 是一个可学习的值), 则减小类内空间距离的损失函数可表示为

$$L_{\text{C}} = \frac{1}{2m} \sum_{i=1}^m \|\mathbf{x}_i - \gamma_{y_i} \mathbf{w}_{y_i}\|^2 \quad (7)$$

式中 $\gamma_{y_i} \mathbf{w}_{y_i}$ 代表深度特征 $\mathbf{x}_i$ 所对应的第 y_i 类的中心, $\mathbf{x}_i$ 与其类中心 $\gamma_{y_i} \mathbf{w}_{y_i}$ 的距离越小, L_{C} 的值越小。

算法 1 ICL 监督学习过程

输入: 训练数据 $\{\mathbf{I}_i, y_i\}_{i=1}^m$, 学习率 η , 超参数 λ_1 和 λ_2 , 迭代次数 T 。

输出: 神经网络参数 θ

1: 随机初始化网路参数 θ 和 γ_k ($k \in D$), 利用文献[20]的算法 1 生成权重 $\mathbf{w}_k$

2: for $t=1$ to T do

3: 根据公式(9)计算损失: L_{ICL}^t

4: 利用公式(10)计算反向梯度: $\frac{\partial L_{\text{ICL}}^t}{\partial \mathbf{x}_i^t}$

5: 利用公式(11)计算反向梯度: $\nabla_{\gamma_k^t}$

6: 更新 γ_k :

$$\gamma_k^{t+1} = \gamma_k - \eta \nabla_{\gamma_k^t}$$

7: 更新网络参数 θ :

$$\theta^{t+1} = \theta^t - \eta \sum_{i=1}^m \frac{\partial L_{\text{ICL}}^t}{\partial \mathbf{x}_i^t} \frac{\partial \mathbf{x}_i^t}{\partial \theta}$$

8: end for

不同类中心的空间距离越远越好, 可以考虑如下约束函数

$$L_{\text{D}} = \sum_{\substack{k, l \in D \\ k < l}} \frac{1}{\|\gamma_k \mathbf{w}_k - \gamma_l \mathbf{w}_l\|^2} \quad (8)$$

式中 $\|\gamma_k \mathbf{w}_k - \gamma_l \mathbf{w}_l\|$ 为第 k 类中心 $\gamma_k \mathbf{w}_k$ 到第 l 类中心 $\gamma_l \mathbf{w}_l$ 的欧式距离, 距离越大, 其倒数越小, L_{D} 越小。

ICL 以 ASL 为基础, 再用 L_{C} 损失减小类内空间距离, 以 L_{D} 损失增加类间空间距离。ICL 定义为

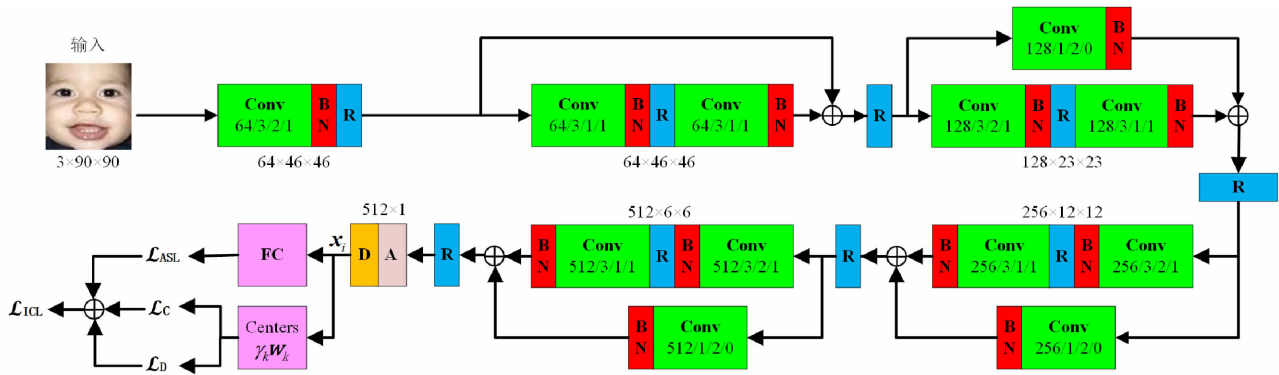
$$L_{\text{ICL}} = L_{\text{ASL}} + \lambda_1 L_{\text{C}} + \lambda_2 L_{\text{D}} \quad (9)$$

式中 λ_1, λ_2 为权值。

2.2 反向传播

在反向传播时, ASL 的权值向量 $\mathbf{w}_k$ 固定不变,

^① ASL 权值参数生成算法的源代码 (PyTorch) https://github.com/general515/Facial_Expression_recognition.



Conv—卷积层(输出通道数/卷积核尺寸/步长/边沿填充);BN—批标准化;

R—ReLU 激活函数;FC—全连接层;A—平均池化;D—随机失活。

图1 仿真实验卷积网络框

Fig. 1 The convolutional network for our experiments

只需要更新深度特征 $\mathbf{x}_i$ 和第 k 类中心的 γ_k 。 L_{ICL} 关于 $\mathbf{x}_i$ 的偏导数为

$$\begin{cases} \frac{\partial L_{ICL}}{\partial \mathbf{x}_i} = \frac{\partial L_{ASL}}{\partial \mathbf{x}_i} + \frac{\partial L_C}{\partial \mathbf{x}_i} \\ \frac{\partial L_{ASL}}{\partial \mathbf{x}_i} = \frac{1}{m} ((p_{y_i} - 1) \mathbf{w}_{y_i} + \sum_{j \in D, j \neq y_i} p_j \mathbf{w}_j) \\ \frac{\partial L_C}{\partial \mathbf{x}_i} = \mathbf{x}_i - \gamma_{y_i} \mathbf{w}_{y_i} \end{cases} \quad (10)$$

式中 $p_j = e^{\mathbf{w}_j^T \mathbf{x}_i} / \sum_{k \in D} e^{\mathbf{w}_k^T \mathbf{x}_i}$ 代表 $\mathbf{x}_i$ 属于第 j 类的概率。

只有 L_C 和 L_D 中含有 γ_k, γ_k 的反向梯度为

$$\nabla_{\gamma_k}^{L_{ICL}} = \nabla_{\gamma_k}^{L_C} + \lambda_2 \frac{\partial L_D}{\partial \gamma_k} \quad (11)$$

$$\nabla_{\gamma_k}^{L_C} = \gamma_k - \mathbf{w}_k^T \frac{\sum_{i=1}^m \delta(y_i = k) \mathbf{x}_i}{\epsilon + \sum_{i=1}^m \delta(y_i = k)}$$

$$\frac{\partial L_D}{\partial \gamma_k} = - \sum_{l \in D, l \neq k} \left(\frac{1}{\|\gamma_k \mathbf{w}_k - \gamma_l \mathbf{w}_l\|^4} \left(\gamma_k + \frac{\gamma_l}{n-1} \right) \right)$$

式中: $\delta(c)$ 当条件 c 为真时取值 1, 为假则取值 0; ϵ 是一个极小的常量, 防止除数可能为零时出错, 文中取值 10^{-6} 。

$\nabla_{\gamma_k}^{L_C}$ 不是损失 L_C 对 γ_k 的偏导数, 而是 γ_k 与当前批次中所有第 k 类样本特征的中心(均值)在 $\mathbf{w}_k$ 上的投影之间的误差。偏导数 $\frac{\partial L_D}{\partial \gamma_k}$ 的推导需要利用 $\mathbf{w}_k^T \mathbf{w}_l = 1/(1-n)$ 。

采用 ICL 监督学习过程如算法 1 所示, 在初始化时采用文献[20]的方法生成等角分布权值向量 $\mathbf{w}_k$, 以及随机初始化网络参数 θ 和 γ_k 。在第 t 次迭代时, 先计算前向传播损失 L_{ICL}^t , 然后根据反向梯度

$\frac{\partial L_{ICL}^t}{\partial \mathbf{x}_i^t}$ 、 ∇_{γ_k} 和 $\frac{\partial \mathbf{x}_i^t}{\partial \theta}$ 更新相关参数。

3 实验结果及分析

3.1 实验相关设置

实验采用 Python 开发平台和 PyTorch 库, 运行平台为一台 Dell 服务器, 其配置为 CPU: Intel Xeon Silver 4114 @ 2.20 GHz, 内存为 64 GB, 显卡为 NVIDIA GeForce RTX 3090。

3.1.1 深度卷积网络结构

实验采用经典 ResNet-18 的变体, 如图 1 所示, 网络输入大小为 90×90 的三通道图像, 输出为 512×1 特征向量。

3.1.2 实验数据集

实验采用 3 个经典的非限制性表情数据集: FER2013^[21]、FERPlus^[22] 和 RAF-DB^[17-18]。

FER2013 数据集有 7 类表情共 35 886 张图像, 用于训练、验证和测试的图像分别是 28 709、3 589 和 3 589 张。训练表情样本包括愤怒 3 995 张、厌恶 436 张、恐惧 4 097 张、悲伤 4 830 张、开心 7 215 张、惊讶 3 171 张以及中性 4 965 张。

FER2013 中含有一些标注错误的数据, 因而 FERPlus 数据集则对 FER2013 中的图像重新进行了人工标注, 在 FER2013 的 7 类基本表情上又增加了蔑视、未知和非人脸这 3 类。FERPlus 每一张图像有 10 个分类标签, 10 个标签上的得分总和为 10。本文只考虑 7 种基本表情, 采用文献[22]提供的最大投票方式筛选后的训练图像有 24 941 张、验证图像 3 186 幅、测试图像 3 137 张。

RAF-DB 有 15 339 表情图片, 其中 12 271 和

3 068分别用于训练和测试。训练图像含有的 7 类基本表情是愤怒 705 张、厌恶 717 张、恐惧 281 张、悲伤1 982张、开心4 772张、惊讶1 290张以及中性 2 524张。

FER2013 和 FERPlus 都含有验证集和测试集,本文只给出测试集的测试结果,结果如表 1~4 所示。

表 1 FER2013 数据集上的性能比较

Table 1 Performance comparison on FER2013 dataset

监督方法	Acc/%	S
SL	71.77	0.422
ASL	72.36	0.348
SL+CL	71.96	0.554
ASL+CL	72.22	0.443
IL	71.53	0.428
L2SL	71.85	0.404
LP	72.25	0.339
ICL	73.02	0.210

表 2 FERPlus 数据集上的性能比较

Table 2 Performance evaluation on FERPlus database

监督方法	Acc/%	S
SL	88.12	0.353
ASL	88.17	0.308
SL+CL	88.23	0.314
ASL+CL	88.34	0.203
IL	88.15	0.281
L2SL	88.21	0.385
LP	88.38	0.276
ICL	88.56	0.192

3.1.3 图像预处理

FER2013 和 FERPlus 的图片是 48×48 灰度图,先扩大成 100×100 图片,然后通过复制通道方式转换为 100×100×3 的图像。RAF-DB 提供 100×100 RGB 彩色图像,无需转换。

训练时,从 100×100×3 人脸图像中随机剪切出 90×90×3 图像块,再把每一个像素值除以 255 归一化。为了增强数据,在±10°范围内随机旋转,且以 50%的概率随机水平翻转。

3.1.4 训练/测试策略

采用随机梯度下降法对模型进行端对端训练,每批 128 个表情样本,动量矩参数为 0.9,采用的 Heinitialization 方法初始化卷积网络系数^[23],L2 正则系数为 3×10⁻⁴,迭代次数为 200,初始学习率为

0.005,迭代 80 次后,每 20 代衰减率 0.8。ICL 的参数 λ₁ 和 λ₂ 分别为 0.005 和 7。

表 3 RAF-DB 数据集上的性能对比

Table 3 Capability comparison on RAF-DB dataset

监督方法	Acc/%	S
SL	85.61	0.324
ASL	85.58	0.271
SL+CL	86.26	0.288
ASL+CL	85.47	0.194
IL	85.49	0.182
L2SL	85.54	0.344
LP	86.11	0.247
ICL	86.26	0.180

测试阶段采用 10-crop 方法,100×100×3 的图像被剪裁成 10 幅 90×90×3 的图像块(左上、右上、左下、右下和中心部分共 5 张,以及它们水平翻转后的 5 幅),然后对这 10 幅子图的输出结果进行平均,再选择概率最高的类别作为识别结果。

3.2 性能评价标准及实验结果

3.2.1 性能评价标准

本文采用文献[20]中的平均准确率和稳定度两个标准。对每种方法测试 N(N=10)次,以 N 次结果的平均值作为最终准确率,定义为

$$Acc = \frac{\sum_{i=1}^N Acc_i}{N}$$

(12)

式中 Acc_i 为第 i 次实验的识别率。由于网络参数随机初始化和训练样本随机分批的原因,相同设置下每一次的识别结果有一定的误差,所以采用多次实验的平均值更加公平可靠。

稳定度是 N 次实验结果的均方差,定义为

$$S = \sqrt{\frac{\sum_{i=1}^N (Acc_i - Acc)^2}{N}}$$

(13)

稳定度 S 体现了相同设置下实验结果的变化程度,S 越小,表明模型性能越稳定。

3.2.2 表情识别性能对比

本文 ICL 是在 SL 基础上的改进,为了验证 ICL 的有效性,主要与 SL 及其改进技术进行对比,对比方法有:SL、ASL^[20]、SL+CL^[13]、ASL+CL、IL^[15]、L2SL^[7]和 LP^[18]。

本文在 FER2013、FERPlus 和 RAF-DB 数据集上进行了对比测试,为了公平对比,所有方法均采用图 1 中卷积网络,且采用相同的训练步骤和网络参

数。ASL 代码由原论文作者提供,其他方法则是按照文献所述原理另行编写程序。

表 4 消融实验结果

Table 4 Ablation test

监督方法	FER2013		FERPlus		RAF-DB	
	Acc/%	S	Acc/%	S	Acc/%	S
L_{ASL}	72.36	0.348	88.17	0.308	85.58	0.271
$L_{ASL}+L_C$	72.23	0.188	88.47	0.167	85.89	0.187
$L_{ASL}+L_C+L_D$	73.02	0.210	88.56	0.192	86.26	0.180

在 FER2013 数据集上的性能对比结果如表 1 所示,ICL 的平均识别率 Acc 最高,达到了 73.02%,比 SL 和 ASL 分别提高了 1.25% 和 0.66%。在稳定度 S 方面,ICL 也取得了最佳成绩 0.210,相比于 SL 的 0.422,提升了一倍,稳定度性能提升是指 S 的数值下降。

FERPlus 数据库的测试结果如表 2 所示,在识别准确率和稳定度方面,ICL 都取得了最好成绩,分别是 88.56% 和 0.192。相比于 SL 和 ASL,ICL 的准确度提高了 0.44% 和 0.39%,其性能提升程度没有 FER2013 上显著。

RAF-DB 数据集上的性能对比如表 3 所示,在识别准确度方面,SL+CL 和 ICL 的识别率最高,达到 88.26%,但是 ICL 的稳定度是 0.180,要比 SL+CL 的稳定度 0.288 好得多。

从表 1~表 3 可知,总体而言,ASL 优于 SL,SL+CL 也很优秀,但是 ASL+CL 在 FER2013 数据库上的识别率和稳定性都比 ASL 差,在 RAF-DB 数据集上识别率也低于 ASL,说明 ASL+CL 可能比单独的 ASL 更差,这也是本文提出 ICL 的出发点。ICL 是以 ASL 为基础,融合了 CL 的思想,再加入了最大化类间空间距离的思想,ICL 在 3 个数据库上都表现优异,表明了 ICL 的合理性和有效性。

3.2.3 参数选择

ICL 有两个人为设置的权值 λ_1 、 λ_2 ,理论上应该采用网格计算最优权值。文献[13-15,17]表明,减小类内变化的约束权值在 0.001~0.01 区间比较合理,因而本文首先设置 λ_1 为 0.005,然后 λ_2 取 1 到 10 共 10 个值,在 FER2013 数据集上的识别率如图 2(b)所示,虽然不是严格的上凸曲线,整体而言, $\lambda_2=7$ 时的识别率明显优于其他值。

λ_2 设置为 7, λ_1 分别取值 0.000 5、0.001、0.005、0.01 及 0.05,在 FER2013 数据集上的识别结果如图 2(a)所示,可知 $\lambda_1=0.005$ 时的识别率最高。

综上所述,本文选取 $\lambda_1=0.005$ 和 $\lambda_2=7$ 。

3.2.4 消融实验

ICL 损失函数由 ASL 损失 L_{ASL} 、类内损失 L_C 和类间距离损失 L_D 这 3 部分构成,本次消融实验测试了 L_{ASL} (ASL)、 $L_{ASL}+L_C$ 和 $L_{ASL}+L_C+L_D$ (ICL),其结果如表 4 所示,其中 $L_{ASL}+L_C$ 与对比实验中的 ASL+CL 是不同的方法。

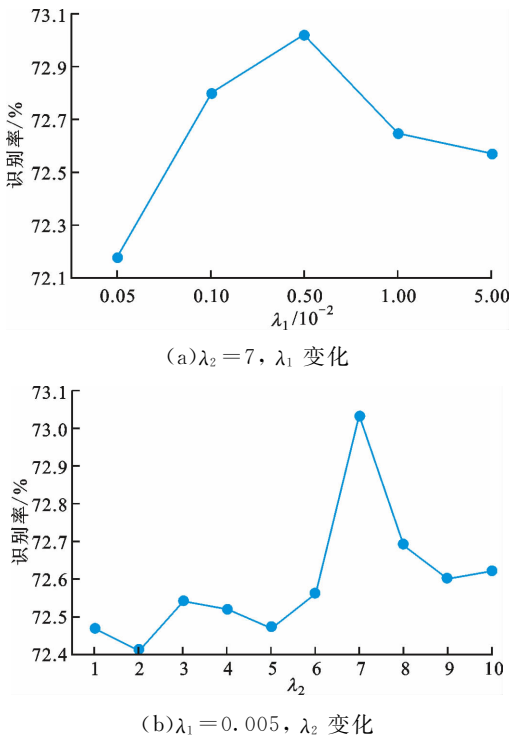


图 2 不同设置 λ_1 和 λ_2 在 FER2013 上的识别率
Fig. 2 Recognition accuracy over the FER2013 dataset with different values of λ_1 and λ_2

在 FER2013 数据库上, L_{ASL} 的识别率和稳定度分别是 72.36% 和 0.348。加入 L_C 损失($L_{ASL}+L_C$)后则分别是 72.23% 和 0.188,识别率反而略微下降,但是稳定度性能大幅提升。再加入 L_D ($L_{ASL}+L_C+L_D$)后的成绩分别是 73.02% 和 0.210, $L_{ASL}+L_C+L_D$ 在识别率方面有了明显的提高;在稳定性方面,比 $L_{ASL}+L_C$ 的稳定度稍差一点,但仍然比 L_{ASL} 优秀很多。

在 FERPlus 数据集中, L_{ASL} 、 $L_{ASL} + L_C$ 和 $L_{ASL} + L_C + L_D$ 3 种方式的识别率分别是 88.17%、88.47%和 88.56%,依次略微提升。这 3 种方式的稳定度分别是 0.308、0.167 和 0.192, $L_{ASL} + L_C$ 的稳定度好, $L_{ASL} + L_C + L_D$ 的稳定度反而比 $L_{ASL} + L_C$ 稍微差一点,但比 L_{ASL} 仍然好很多。

在 RAF-DB 数据库上, L_{ASL} 、 $L_{ASL} + L_C$ 和 $L_{ASL} + L_C + L_D$ 的识别准确率分别是 85.58%、85.89%和 86.26%,依次增高。三者的稳定度分别是 0.271、0.187 和 0.180,也是依次变好; $L_{ASL} + L_C$ 和 $L_{ASL} + L_C + L_D$ 的稳定度非常接近,几乎差不多。

表 5 FER2013 数据集上训练一个 epoch 所需平均时间
Table 5 Average training time for one epoch with respect to the FER2013 database

监督方法	SL	ASL	SL+CL	ASL+CL	IL	L2SL	ICL
时间/s	33.92	33.51	34.38	33.97	42.15	36.36	34.31

4 结 论

针对传统 SL 监督学习的深度特征区分度不足,本文提出了一种孤立中心损失方法(ICL),它主要包含了最大化类间角度距离、减小类内变化以及最大化类间(欧氏)空间距离 3 种约束,从而使各类成为彻底的“孤岛”,以提高各类之间的区分度。在 FER2013、FERPlus 和 RAF-DB 3 个数据集上的实验结果表明:ICL 在识别准确率和稳定性两个方面都有明显的提升。ICL 虽然由 3 部分组成,但是计算复杂度并未增加太多,运行速度只比传统 SL 略微慢一点,仍然比一些其他方法快。

ICL 的基础是 ASL, ASL 中生成等角分布固定权值向量过程的计算量随着类别数量增大而呈几何形式增长,难以处理类别特别多的分类任务,ICL 仍然存在这个问题,这是将来工作需要改进的一个方向。多损失函数融合模型在许多领域非常有效,未来工作将在这方面进行尝试。

参考文献:

[1] JIANG Ping, WAN Bo, WANG Quan, et al. Fast and efficient facial expression recognition using a Gabor convolutional network [J]. IEEE Signal Processing Letters, 2020, 27: 1954-1958.

[2] RUAN Delian, YAN Yan, LAI Shenqi, et al. Feature decomposition and reconstruction learning for effective facial expression recognition [C] // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2021;

由此可见,大体上 $L_{ASL} + L_C$ 可以大幅增加稳定性, $L_{ASL} + L_C + L_D$ 在保持稳定性优势的基础上可以提高一定的识别准确度。

3.2.5 时间复杂度

各种监督方法在 FER2013 数据集上训练一个 Epoch 所需的平均时间如表 5 所示, ASL 速度最快,只需 33.51 s。SL 是 33.92 s, ASL+CL 与 SL 几乎差不多,需要 33.97 s。ICL 是以 ASL 为基础融合了 CL 思想,再增加了一部分约束,所以 ICL 比 ASL+CL 计算复杂度高,需要 34.31 s,仍然比 SL+CL 略微快一点,比 IL 和 L2SL 快很多。

7656-7665.

[3] WANG Kai, PENG Xiaojiang, YANG Jianfei, et al. Suppressing uncertainties for Large-Scale facial expression recognition [C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2020: 6896-6905.

[4] YANG Huiyuan, CIFTCI U, YIN Lijun. Facial expression recognition by de-expression residue learning [C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 2168-2177.

[5] LI Danyang, WEN Guihua, HOU Zhi, et al. RTCRelief-F: an effective clustering and ordering-based ensemble pruning algorithm for facial expression recognition [J]. Knowledge and Information Systems, 2019, 59(1): 219-250.

[6] LI Shan, DENG Weihong. Deep facial expression recognition: a survey [J/OL]. IEEE Transactions on Affective Computing [2021-08-30]. https://ieeexplore.ieee.org/document/9039580.

[7] DENG Jiankang, GUO Jia, XUE Niannan, et al. ArcFace: additive angular margin loss for deep face recognition [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2019: 4685-4694.

[8] 吴慧华, 苏寒松, 刘高华, 等. 基于余弦距离损失函数的人脸表情识别算法 [J]. 激光与光电子学进展, 2019, 56(24): 188-194.

WU Huihua, SU Hansong, LIU Gaohua, et al. Facial expression recognition algorithm based on cosine distance loss function [J]. Laser & Optoelectronics Pro-

- gress, 2019, 56(24): 188-194.
- [9] 龙鑫, 苏寒松, 刘高华, 等. 一种基于角度距离损失函数和卷积神经网络的人脸识别算法 [J]. 激光与光电子学进展, 2018, 55(12): 402-413.
- LONG Xin, SU Hansong, LIU Gaohua, et al. A face recognition algorithm based on angular distance loss function and convolutional neural network [J]. Laser & Optoelectronics Progress, 2018, 55(12): 402-413.
- [10] 张文萍, 贾凯, 王宏玉, 等. 改进的 Island 损失函数在人脸表情识别上的应用 [J]. 计算机辅助设计与图形学学报, 2020, 32(12): 1910-1917.
- ZHANG Wenping, JIA Kai, WANG Hongyu, et al. Application of improved Island loss in facial expression recognition [J]. Journal of Computer-Aided Design & Computer Graphics, 2020, 32(12): 1910-1917.
- [11] WANG Feng, XIANG Xiang, CHENG Jian, et al. NormFace: L_2 hypersphere embedding for face verification [C]//Proceedings of the 25th ACM International Conference on Multimedia. New York, NY, USA: ACM, 2017: 1041-1049.
- [12] WANG Hao, WANG Yitong, ZHOU Zheng, et al. CosFace: large margin cosine loss for deep face recognition [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 5265-5274.
- [13] WEN Yandong, ZHANG Kaipeng, LI Zhifeng, et al. A discriminative feature learning approach for deep face recognition [C]//Computer Vision — ECCV 2016. Cham: Springer International Publishing, 2016: 499-515.
- [14] WEN Yandong, ZHANG Kaipeng, LI Zhifeng, et al. A comprehensive study on center loss for deep face recognition [J]. International Journal of Computer Vision, 2019, 127(6): 668-683.
- [15] CAI Jie, MENG Zibo, KHAN A S, et al. Island loss for learning discriminative features in facial expression recognition [C]//2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). Piscataway, NJ, USA: IEEE, 2018: 302-309.
- [16] LUO Zimeng, HU Jiani, DENG Weihong. Local subclass constraint for facial expression recognition in the wild [C]//2018 24th International Conference on Pattern Recognition (ICPR). Piscataway, NJ, USA: IEEE, 2018: 3132-3137.
- [17] LI Shan, DENG Weihong, DU Junping. Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2017: 2584-2593.
- [18] LI Shan, DENG Weihong. Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition [J]. IEEE Transactions on Image Processing, 2019, 28(1): 356-370.
- [19] ZHAO Kai, XU Jingyi, CHENG Mingming. RegularFace: deep face recognition via exclusive regularization [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2019: 1136-1144.
- [20] JIANG Ping, LIU Gang, WANG Quan, et al. Accurate and reliable facial expression recognition using advanced softmax loss with fixed weights [J]. IEEE Signal Processing Letters, 2020, 27: 725-729.
- [21] DHALL A, GOECKE R, JOSHI J, et al. Emotion recognition in the wild challenge 2013 [C]//Proceedings of the 15th ACM on International Conference on Multimodal Interaction. New York, NY, USA: ACM, 2013: 509-516.
- [22] BARSOUM E, ZHANG Cha, FERRER C C, et al. Training deep networks for facial expression recognition with crowd-sourced label distribution [C]//Proceedings of the 18th ACM International Conference on Multimodal Interaction. New York, NY, USA: ACM, 2016: 279-283.
- [23] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification [C]//2015 IEEE International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2015: 1026-1034.

(编辑 赵炜 亢列梅)