DOI: 10.11992/tis.202212030
网络出版地址: https://link.cnki.net/urlid/23.1538.TP.20230731.1904.012

融合全局与局部特征的跨数据集表情识别方法

梁艳, 温兴, 潘家辉
(华南师范大学 软件学院, 广东 佛山 528225)

摘 要: 人脸表情数据集在收集过程中存在主观的标注差异和客观的条件差异, 导致表情识别模型在不同数据集间呈现明显的性能差异。为了提高跨数据集表情识别精度、减少表情识别在实际应用中进行样本打标重训练的过程, 本文提出了一种基于表情融合特征的域对抗网络模型, 用于跨数据集人脸表情识别。采用残差神经网络提取人脸表情的全局特征与局部特征。利用 Encoder 模块对全局特征与局部特征进行融合, 学习更深层次的表情信息。使用细粒度的域鉴别器进行源数据集与目标数据集对抗, 对齐数据集的边缘分布和条件分布, 使模型能迁移到无标签的目标数据集中。以 RAF-DB 为源数据集, 以 CK+、JAFFE、SFEW2.0、FER2013、Expw 分别作为目标数据集进行跨数据集人脸表情识别实验。与其他跨数据集人脸表情识别算法相比, 所提方法获得了最高的平均识别率。实验结果表明, 所提方法能有效提高跨数据集人脸表情识别的性能。

关键词: 跨数据集; 人脸表情识别; 领域自适应; 特征融合; 自注意力机制; 迁移学习; 细粒度域鉴别器; 残差网络
中图分类号: TP391 **文献标志码**: A **文章编号**: 1673−4785(2023)06−1205−08

中文引用格式: 梁艳, 温兴, 潘家辉. 融合全局与局部特征的跨数据集表情识别方法 [J]. 智能系统学报, 2023, 18(6): 1205–1212.
英文引用格式: LIANG Yan, WEN Xing, PAN Jiahui. Cross-dataset facial expression recognition method fusing global and local features[J]. CAAI transactions on intelligent systems, 2023, 18(6): 1205–1212.

Cross-dataset facial expression recognition method fusing global and local features

LIANG Yan, WEN Xing, PAN Jiahui
(School of Software, South China Normal University, Foshan 528225, China)

Abstract: The expression recognition model shows significant performance differences between datasets due to subjective annotation and objective condition differences in the collection of facial expression datasets. A domain adversarial network model based on expression fusion features is proposed for cross-dataset facial expression recognition. This model aims to improve the accuracy of cross-dataset expression recognition and reduce the sample marking and retraining processes for expression recognition in practical applications. Residual neural networks are used to extract the global and local features of facial expressions. An encoder module is then employed to fuse global and local features to learn deep expression information. A fine-grained domain discriminator is adopted to antagonize the source dataset against the target dataset, aligning the edge and conditional distributions of the dataset and facilitating the migration of the model to the unlabeled target dataset. RAF-DB is used as the source dataset, and CK+, JAFFE, SFEW2.0, FER2013, and Expw are used as the target datasets for cross-dataset facial expression recognition experiments. Compared with other cross-dataset facial expression recognition algorithms, the proposed method achieves the highest average recognition rate. Experimental results show that the proposed method can effectively improve the performance of cross-dataset facial expression recognition.

Keywords: cross-dataset; facial expression recognition; domain adaptation; feature fusion; self-attention mechanism; transfer learning; fine-grained domain discriminator; residual network

人脸表情是人类最自然、最直接的情绪表达方式之一。研究发现, 在人们日常交流沟通的过程中, 有 55% 的情感信息靠人脸表情进行传递[1]。

收稿日期: 2022−12−29. **网络出版日期**: 2023−08−01.
基金项目: 国家科技创新 2030 重点项目(2022ZD0208900); 国家自然科学基金项目(62076103).
通信作者: 梁艳. E-mail: liangyan@m.scnu.edu.cn.

研究人脸表情识别有效促进人机交互系统的发展。目前，该技术已广泛应用在医学、安全监控、教育等领域[2]。

为了推动人脸表情识别的理论研究与实际应用，在过去的十几年里，研究者们已公开了多个表情数据集，并提出了多种方法来提高表情识别的性能。但是，大部分的表情识别算法都基于一个前提，即：训练集和测试集来自同一个数据集，训练数据和测试数据特征分布相同。然而这一假设并不总是成立，在实际应用中，测试集与训练集通常来自不同的数据分布，因此模型需要进行跨数据集表情识别验证。

近年来，领域自适应方法成为迁移学习中最为热门的研究之一，其核心问题是解决数据分布不一致对模型性能的影响。Xu 等[3] 证明，把源域和目标域的特征范数调整到一个较大范围的值可以获得显著的迁移收益。Lee 等[4] 利用特定任务的决策边界和 Wasserstein 度量在领域之间进行特征分布对齐。考虑到领域自适应方法在解决跨域问题的有效性，有学者尝试把基于统计差异的领域自适应方法用于跨数据集表情识别任务。莫宏伟等[5] 利用一个特征变换矩阵，把源域和目标域数据映射到公共子空间，减小域间分布差异。Long 等[6] 基于统计的思想提出了一种新的深度自适应网络 (deep adaptation network, DAN) 架构，把领域自适应方法与深度学习技术结合起来。Li 等[7] 将 DAN 网络应用到人脸表情识别，引入最大均值误差 (maximum mean discrepancy, MMD) 来测量源域与目标域的特征散度，减小源域与目标域的分布距离。Xu 等[8-9] 基于 MMD 损失寻找远离表情特征中心的异常样本，并在训练过程中通过抑制异常样本来提高跨数据集表情识别准确率。

受对抗学习技术的启发，有部分学者采用基于对抗学习的领域自适应方法，即域对抗自适应方法，实现跨数据集表情识别。该类方法的核心思想是加入一个域鉴别器，使之与表情分类器进行对抗，在对抗过程中学习到同时适用于两个数据集的表情特征。Chen 等[10] 将经典的域对抗自适应方法：领域对抗神经网络 (domain-adversarial neural network, DANN)[11]、条件域对抗自适应网络 (conditional domain adversarial network, CDAN)[12] 应用到跨数据集表情识别任务，学习领域不变性特征。Wang 等[13] 在域对抗中通过缩小目标数据集样本与源数据集对应类别的特征中心的距离，扩大与源数据集不同类别的特征中心的距离，实现类级别的对齐。

领域自适应方法仅在特征分布层面上对齐不同域特征分布，目标数据集无需提供标签信息，因此可应用于无监督的跨数据集表情识别[14]。但是，目前大部分基于领域自适应的跨数据集表情识别方法仅对齐表情特征的边缘分布，未关注不同数据集间的表情类内差异导致特征的条件分布差异。而使用通用的域对抗自适应算法强行对齐两个数据集间的整体分布，将不可避免地把来自源数据集和目标数据集的不同表情类别样本混合在一起，导致不同表情数据集间类别不匹配问题。

因此，为了提高跨数据集表情识别的特征可迁移性，解决跨数据集表情类别不匹配问题，本文提出一种利用表情融合特征对齐不同数据集联合分布的领域自适应方法，利用编码器（Encoder）模块融合表情的全局特征和局部特征，并通过表情分类器与细粒度域鉴别器联合对抗训练，提高分类器在无标签的目标数据集的识别效果。

1　本文方法

在跨数据集表情识别任务中，给定一个源数据集 $D_s=\{(x_i^s,y_i^s)\}_{i=1}^{n_s}$ 和目标数据集 $D_t=\{(x_j^t)\}_{j=1}^{n_t}$，其中 x 表示样本，n 表示样本数量。这两个数据集在两种不同环境下采样，具有不同的分布 $p_s(X,Y)$ 和 $p_t(X,Y)$，其中目标数据集样本不提供标签。为了提升跨数据集表情识别性能，本文从两方面进行改进，提高跨数据集表情识别性能。1）通过关注人脸表情的关键区域，学习更多表情相关信息，提高表情特征的可迁移性，抑制数据集自带的偏差。2）使用细粒度的对抗领域自适应策略，对齐表情类级别的信息。本文提出的域对抗网络模型框架如图 1 所示。该模型主要由特征提取器、表情分类器和域鉴别器 3 部分组成。特征提取器利用多残差网络 (multi-ResNet) 提取人脸表情的全局和局部特征，然后利用 Encoder 层进行表情特征融合。表情分类器由两层全连接网络构成，根据融合特征对表情进行分类。域鉴别器用于与表情分类器进行联合对抗，本文通过把传统域鉴别器的 2 个域判别通道（即源域和目标域）扩展为 $2K$ 通道（K 为表情类别数），进行不同数据集间的整体对抗和不同数据集相同表情类别间的细粒度对抗，达到同时对齐数据边缘分布和条件分布的效果。

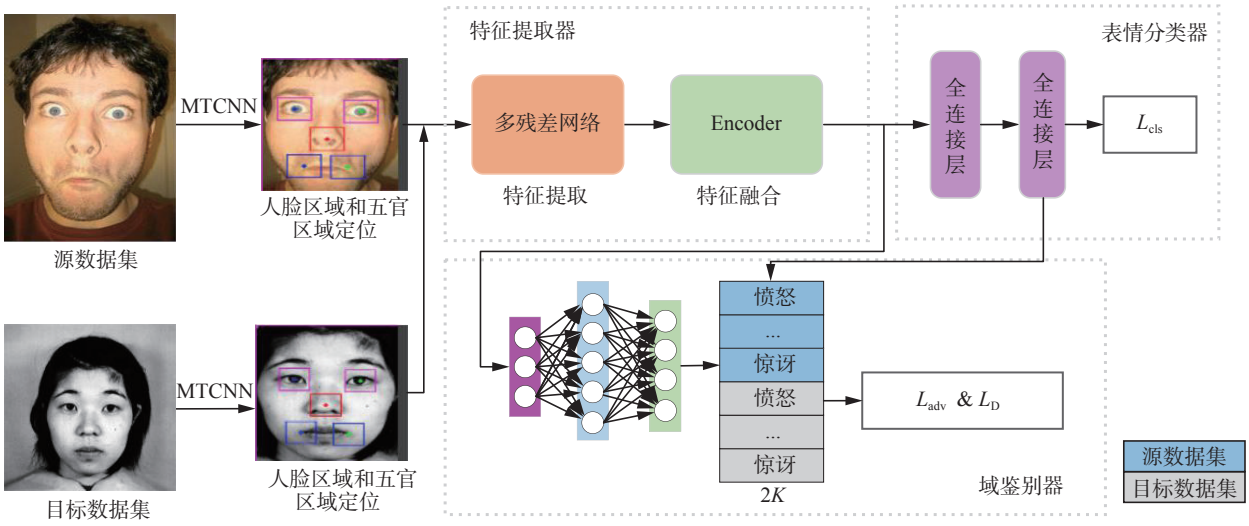


图 1　基于表情融合特征的域对抗网络模型框架

Fig. 1　Framework for domain adversarial network based on facial expression fusion feature

1.1　表情融合特征的提取

根据人脸动作单元(action unit, AU)[15]的划分可知，表情的决定性信息聚集在人脸的五官位置。为了提高表情特征的可迁移性，本文提取人脸区域的全局特征和五官区域的局部特征，并利用 Encoder 模型进行特征融合。特征提取器的具体结构如图 2 所示。

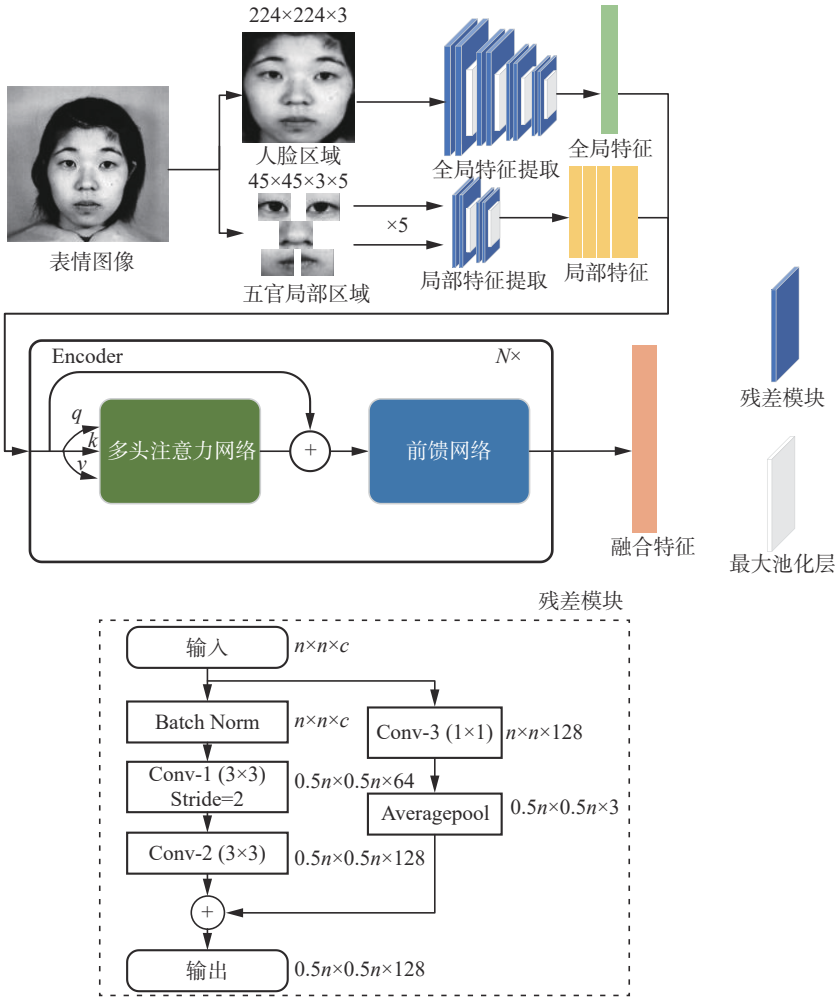


图 2　特征提取器的结构

Fig. 2　Structure of feature extractor

首先使用多任务卷积神经网络(multi-task convolutional neural network, MTCNN)[16]对表情数据集的人脸图像进行人脸定位以及 5 个关键点(左眼、右眼、鼻子、左嘴角、右嘴角)定位。然后,将人脸区域输入到四层残差网络[17],提取表情的全局特征。此外,以关键点为中心,截取 5 个大小为$0.2W\times0.2H$ (W、H分别为人脸区域的宽和高) 的子图作为判断表情类别的关键区域,输入两层残差网络,提取表情的局部特征。

为了使模型学习到领域不变性的表情特征,本文基于 Transformer[18] 的 Encoder 模块,设计了一个具有N层的表情 Encoder 层,将上述提取的全局和局部表情特征输入 Encoder 层进行表情特征的融合。Encoder 层包括一个多头注意力网络和一个前馈网络。首先根据全局和局部表情特征获得 3 个自注意力向量$\boldsymbol{q}$、$\boldsymbol{k}$和$\boldsymbol{v}$[19],然后,输入多头注意力网络,根据下式计算特征间的权重,获得加权后的特征$\boldsymbol{c}_i$:

$$\boldsymbol{c}_i=\mathrm{softmax}(\frac{\boldsymbol{q}_i\boldsymbol{k}_i^{\mathrm{T}}}{\sqrt{d}})\boldsymbol{v}_i \tag{1}$$

其中:d为特征维度,这里为 128。把加权特征$\boldsymbol{c}_i$输入前馈网络进行学习,最终获得表情融合特征$\boldsymbol{x}_i$。

1.2　细粒度域鉴别器

无监督的跨数据集表情识别任务中,其目标是学习一个表情识别模型G,令G可以在不带标签的目标数据集上实现较高表情识别准确率。具体来说,表情识别模型G由特征提取器F和表情分类器C构成。域对抗自适应方法在解决跨域表情识别问题时,在表情识别模型G的基础上引入了域鉴别器D。通过域鉴别器D对表情识别模型G提取的表情特征进行域来源判断,在反向传播时加入梯度反转层,使模型混淆来自不同数据集的表情特征,从而使表情分类器C能应用到目标数据集。最后,通过表情分类器C和域鉴别器D联合对抗训练,实现在无标签的目标数据集上进行表情分类。

大部分域对抗自适应方法中,域鉴别器D采用二分类方式区分表情特征来自源数据集还是目标数据集,再由梯度反转进行特征混淆,对齐数据集间边缘分布。但是,由于人脸表情存在类内差异大、类间差异小的特性,仅仅混淆源、目标数据集内的所有特征,会引起表情数据集间跨域类别不匹配问题。因此,本文对算法进行改进,令表情分类器C与域鉴别器D不仅在数据集间进行宏观的对抗,还增加了表情相同类间的细粒度对抗,使数据集相同类间能实现协调自适应。

传统域对抗自适应损失为

$$L=\alpha L_{\mathrm{cls}}+\beta L_{\mathrm{d}} \tag{2}$$

式中:L_{cls}为表情的分类损失,L_{d}为域判别损失,α和β分别是分类损失和域判别损失的权重。L_{cls}的目的是帮助G学习到表情分类信息,它采用交叉熵损失在源数据集上最小化预测分类与真实表情分类间的区别,计算公式为

$$L_{\mathrm{cls}}=-\sum_{i=1}^{S}\sum_{k=1}^{K}y_{ik}\log(p_{ik}) \tag{3}$$

式中:S表示源域样本数量,K表示表情类别,y_{ik}为源域样本i第k类的类别信息,p_{ik}为表情识别模型G预测源域样本i为第k类表情的类别信息。

式 (2) 中的域判别损失L_{d}目的是帮助域鉴别器D区分来自不同数据集的表情特征,使提取的特征能对齐源数据集和目标数据集,损失计算公式为

$$\begin{aligned}L_{\mathrm{d}}=&-\sum_{i=1}^{S}\Big[(1-d)\log P(d=0|x_i)\Big]-\\&\sum_{i=1}^{T}\Big[d\log P(d=1|x_i)\Big]\end{aligned} \tag{4}$$

式中:d为 0 代表特征来自源数据集,为 1 则代表特征来自目标数据集;S为源数据集样本数量;T为目标数据集样本数量;$P(d=0|x)$为域鉴别器预测特征为源数据集的概率。

传统的域鉴别器只能判别$d=0$或者$d=1$,即特征标签为[1,0]或[0,1]。为了将表情类别信息纳入对抗性学习框架,达到同时对齐表情特征的边缘分布和条件分布的效果,本文修改了传统的域鉴别器D,将2个域判别通道扩展为$2K$通道(K为表情类别数),进行不同数据集间的整体对抗以及不同数据集相同表情类别间的细粒度对抗。通过更细粒度的对抗性学习,不仅仅对齐数据集间表情特征的边缘分布,而且对齐特征的类内条件分布。

本文使用表情特征提取器和分类器对目标域进行软标签的标注,然后将源数据集表情图像与目标数据集表情图像的标签扩展为$2K$维标签,其中源域标签在1至K维使用原来的标签信息,在$K+1$至$2K$维数据置为 0;目标域标签在 1 至K维数据置为 0,在$K+1$至$2K$维使用软标签标注。通过对i和$K+i$类进行对抗自适应即可实现不同数据集间表情分布对齐。

为了实现基于类别的对抗,本文将提取的融合特征输入细粒度域鉴别器中计算细粒度类判别损失。与传统域判别损失L_{d}不同的是,本文在L_{D}加入了类别信息,具体计算公式如下:

$$L_{\mathrm{D}}=-\sum_{i=1}^{S}\sum_{k=1}^{K}\left[a_{ik}\log P(c=k,d=0)|x_i\right]-\sum_{j=1}^{T}\sum_{k=1}^{K}\left[a_{jk}\log P(c=k,d=1)|x_j\right] \tag{5}$$

式中：a_{ik}和a_{jk}分别为源域样本i和目标域样本j为第k类的信息，即上文所述构建$2K$维的标签信息。

此外，为了引导特征提取器F学习到两个数据集共用的表情特征，我们还增加了一个整体判别损失L_{adv}，其目的是帮助域鉴别器获取目标数据集的类别信息，从而经过梯度翻转后可以混淆两个数据集的类别信息,进而引导特征提取器F学习共用表情特征，L_{adv}的计算公式如下：

$$L_{\mathrm{adv}}=-\sum_{j=1}^{T}\sum_{k=1}^{K}\left[a_{jk}\log P(c=k,d=0)|x_j\right] \tag{6}$$

综上所述，本文采用的总损失 L 为

$$L=\omega_1 L_{\mathrm{cls}}+\omega_2 L_{\mathrm{D}}+\omega_3 L_{\mathrm{adv}} \tag{7}$$

其中：ω_1、ω_2和ω_3分别是表情分类损失、细粒度类判别损失和整体判别损失的权重。

在训练过程中，将源数据集表情图像的特征输入表情分类器中计算表情分类损失L_{cls}，将源、目标数据集表情图像的特征输入域鉴别器计算域判别损失L_{D}和L_{adv}，最终，在域鉴别器D和表情分类器C的对抗学习下对齐不同表情数据集间的联合分布。

2　实验结果与分析

2.1　表情数据集

本文采用 6 个表情数据集进行算法测试，具体包括实验室环境下的 CK+[20] 和 JAFFE[21] 数据集和自然场景下的 SFEW2.0[22]、FER2013[23]、ExpW[24]、RAF-DB[25] 数据集。这些数据集都包含愤怒、厌恶、恐惧、高兴、悲伤、惊讶、中性等 7 种表情。

CK+数据集包含来自 123 个实验对象的 593 个图像序列，每个图像序列都是从中性表情到峰值表情。本文参照文献 [7] 的方法，从每个序列中抽取 1 帧中性表情图像和 3 帧表情图像，去除无效数据后共获得 1236 张图像进行实验。

JAFFE 数据集包括来自 10 位日本女性共 213 张图像。本文使用了所有图像进行实验。

SFEW2.0 数据集由不同电影的表情图像构成，具有不同的头部姿势、年龄范围、遮挡和照明。该数据集分为训练集、验证集和测试集，分别有 958、436 和 372 个样本。

FER2013 是一个自然场景下获得的表情数据集，包含 35 887 张大小为 48 像素×48 像素的图像。数据集进一步分为 28 709 张图像的训练集、3 589 张图像的验证集和 3 589 张图像的测试集。

ExpW 数据集由谷歌图像搜索中下载的表情图像构成，包含 91 793 张人脸图像。

RAF-DB 数据集也是由互联网上收集的图像构成，共 29 672 张表情图像，其中 15 339 张图像有 7 种基本表情，分为 12 271 个训练样本和 3 068 个测试样本。

2.2　评估标准

遵循跨数据集表情识别的通用标准[14]，本文选取平均准确率作为评价指标。首先计算出某表情类别的准确率，然后再计算所有类别的准确率均值，即为跨数据集表情识别算法的平均准确率。

2.3　实现细节

本文方法的训练目标为最小化式 (7) 的总损失 L，以目标数据集获得最高平均准确率作为标准，训练表情识别模型G和域鉴别器D。本文分两个阶段进行训练。第一阶段，在源数据集采用随机梯度下降（stochastic gradient descent, SGD）算法训练特征提取器F和表情分类器C，初始学习率设为 0.01，SGD 的动量设为 0.9，训练 100 轮后获得初始的表情识别模型G；第二阶段，加入域鉴别器D，使用总损失 L 进行对抗训练，使初始表情识别模型G迁移到不带标签的域鉴别器中，在这步骤中同样使用 SGD 算法训练模型，除了特征提取器F和表情分类器C的学习率降到 0.001 外，其余超参数均与第一阶段相同，本阶段训练采用学习率递减策略，每 20 轮学习率乘以 0.5。式 (7) 中 3 个损失权重ω_1、ω_2和ω_3的比值设为 50∶50∶1。

2.4　消融实验

为探究融合特征对表情识别性能的影响，本文采用相同的网络提取全局特征、局部特征和融合特征，在 6 个数据集进行表情识别实验，结果如表 1 所示（文中表格加粗数据为最佳结果）。

从实验结果可知，本文提出的融合特征方法在 6 个数据集的表情识别性能均优于仅采用全局特征或局部特征的方法，它的平均表情识别准确率比仅采用全局特征的方法提高了 4.95%，比仅采用局部特征的方法则提高了 24.56%。由此可见，表情全局特征与局部特征存在互补性，对两种特征进行融合，可以大幅提高表情识别的准确率。

此外，为了验证细粒度域对抗自适应方法在跨数据集表情识别任务中的有效性，我们参照文献 [14] 的做法，采用 RAF-DB 作为源域，其余 5 个

数据集作为目标域，使用融合特征进行对抗，与无域对抗方法、两种通用域对抗自适应方法（DANN[11] 和 CDAN[12]）进行模型迁移效果对比，实验结果如表 2 所示。

表 1　分别采用全局特征、局部特征、融合特征进行表情识别的结果对比

Table 1　Comparison of expression recognition results using global features, local features, and fusion features, respectively

%

特征类型	CK+	RAF-DB	JAFFE	SFEW2.0	FER2013	ExpW	平均准确率
全局特征	91.47	79.03	93.75	34.64	65.63	68.63	72.19
局部特征	70.93	53.94	65.66	29.59	48.81	46.53	52.58
融合特征	**96.90**	**79.20**	**98.12**	**51.52**	**66.84**	**70.23**	**77.14**

表 2　无域对抗、通用域对抗、细粒度域对抗的跨数据集识别结果对比

Table 2　Comparison of cross-dataset recognition results for non-domain adversarial, general domain adversarial, and fine-grained domain adversarial

%

方法	CK+	JAFFE	SFEW2.0	FER2013	ExpW	平均准确率
无域对抗	53.57	49.25	29.27	44.18	31.30	41.51
通用域对抗 (DANN)[11]	80.62	54.46	45.18	51.36	63.80	59.08
通用域对抗 (CDAN)[12]	**80.95**	53.52	**52.72**	54.18	64.63	61.20
细粒度域对抗	80.92	**61.54**	51.13	**55.95**	**68.94**	**63.70**

从表 2 可知，采用细粒度域对抗自适应方法的结果均优于无域对抗方法和 DANN 方法，其平均准确率相较于无域对抗方法提高了 22.19%，相较于 DANN 和 CDAN 方法，分别提高了 4.62% 和 2.50%。实验结果证明，细粒度域对抗自适应方法能有效地提高跨数据集的表情识别性能。

2.5　实验效果对比

为验证本文方法的性能，我们把本文方法与近五年的几个跨数据集算法进行对比。所有方法均使用相同的源数据集 RAF-DB 和主干网络 ResNet-18，分别以 CK+、JAFFE、SFEW2.0、FER2013、ExpW 作为目标域进行测试，结果如表 3 所示。其中，POCAN[13] 和 ESSRN[9] 方法的数据来源于原文献，其他几种方法的数据则来自文献 [10] 对这些算法的复现结果。

表 3　本文方法与其他方法的比较

Table 3　Comparison of the proposed method with other methods

%

方法	CK+	JAFFE	SFEW2.0	FER2013	ExpW	平均准确率
SAFN[3]	68.99	49.30	50.46	53.31	68.32	58.08
SWD[4]	72.09	53.52	49.31	53.70	65.85	58.89
DETN[26]	64.19	52.11	42.25	42.01	43.92	48.90
ECAN[7]	66.51	52.11	48.21	50.76	48.73	53.26
AGRA[10]	77.52	61.03	**52.75**	54.94	**69.70**	63.19
POCAN[13]	76.74	52.11	—	—	—	—
ESSRN[9]	80.83	**63.85**	—	50.98	—	—
本文方法	**80.92**	61.54	51.13	**55.95**	68.94	**63.70**

从表 3 可以看出，本文方法在 CK+和 FER2013 进行跨数据集表情识别时，获得最优识别结果。在 JAFFE、SFEW2.0 和 ExpW 数据集也获得了次优的准确率。本文方法的平均准确率达到 63.70%，高于其他方法。

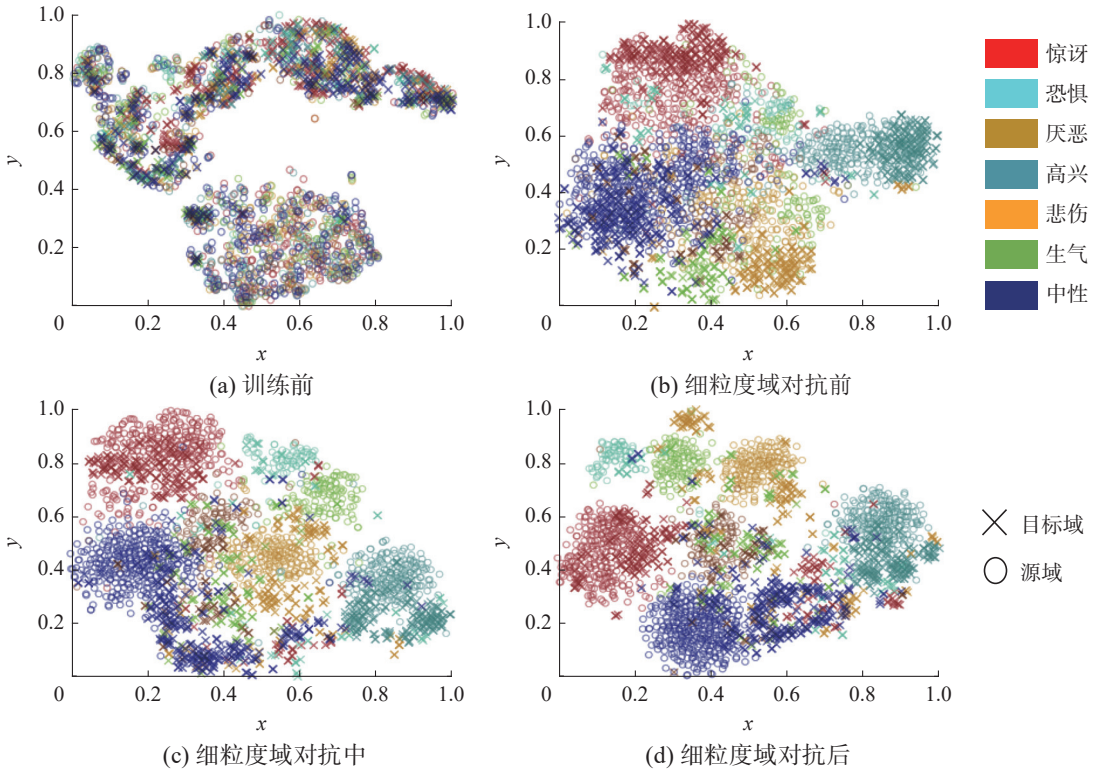
值得注意的是，本文方法在 SFEW2.0 和 ExpW 数据集的准确率稍低于 AGRA 方法。这可能是因为两个数据集均为自然场景下获取的数据集，部分人脸存在较大的头部姿态变化以及面部遮挡等问题，导致局部表情特征获取失败，影响了本文提出的表情识别模型的性能。

2.6　特征分布可视化

为了进一步证明细粒度域鉴别器能有效地对齐不同数据集表情类别间的分布，我们把迁移过程中不同阶段的表情特征进行可视化展示和对比。具体来说，我们以 RAF-DB 为源数据集，CK+为目标数据集，将迁移过程的 4 个阶段：训练前，细粒度域对抗前（仅在源数据集训练），细粒度域对抗中（加入目标数据集后，经过 30 轮的训练），细粒度域对抗后。这四种情况的表情特征使用 t-SNE 算法[27] 降维，进行可视化展示，如图 3 所示。

从图 3 可以看到，在模型训练前，两个数据集表情类别的特征分布非常混杂，无法进行表情分类。在细粒度域对抗前，由于已经在源数据集进行了第一阶段的表情分类训练，两个数据集的相同表情类别的特征聚类开始显现。在细粒度域对抗训练过程中，两个数据集的相同表情类别聚类更明显，类间差距也逐渐扩大。细粒度域对抗训练完成后，两个数据集的特征已呈现聚类，表情

的类间分布差异明显。这表明，通过细粒度域对抗训练，可以学习到不同数据集的相同表情类别信息，并聚合在一起，同时加大不同表情类间距离，从而降低两个数据集间的特征分布差异。



(a) 训练前　(b) 细粒度域对抗前

(c) 细粒度域对抗中　(d) 细粒度域对抗后

图 3　RAF-DB 迁移到 CK+的 4 个阶段的特征分布

Fig. 3　Feature distribution of four stages of RAF-DB transfer to CK+

3　结束语

为了解决跨数据集表情识别的问题，本文提出了一种基于表情融合特征的域对抗网络模型。该模型利用 Encoder 模块融合表情的全局和局部特征，在提高表情特征的鲁棒性的同时，减少了表情特征的跨域差异，有利于后续表情模型的迁移。此外，为了解决不同表情数据集的类别不匹配导致跨数据集识别精度下降的问题，本文基于表情类别进行细粒度的对抗学习。在实验部分，本文通过消融实验及可视化实验证明特征融合以及细粒度域对抗自适应方法的有效性。通过与近年几个表现优异的算法比较，证明了本文方法的有效性。目前，本文算法仅在公开表情数据集进行跨数据集实验达到较为理想的效果，在未来研究中，我们将尝试构建个人数据集验证算法的鲁棒性和实用性，并把算法推广到动态表情数据上，提高动态表情的跨数据集效果。

参考文献：

[1] MEHRABIAN A. Communication without words[M]// Communication theory. [S. l.]: Routledge, 2017: 193−200.

[2] LI Shan, DENG Weihong. Deep facial expression recognition: a survey[J]. IEEE transactions on affective computing, 2022, 13(3): 1195–1215.

[3] XU Ruijia, LI Guanbin, YANG Jihan, et al. Larger norm more transferable: an adaptive feature norm approach for unsupervised domain adaptation[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2020: 1426−1435.

[4] LEE Chenyu, BATRA T, BAIG M H, et al. Sliced Wasserstein discrepancy for unsupervised domain adaptation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2020: 10277−10287.

[5] 莫宏伟, 傅智杰. 基于迁移学习的无监督跨域人脸表情识别 [J]. 智能系统学报, 2021, 16(3): 397–406.
MO Hongwei, FU Zhijie. Unsupervised cross-domain expression recognition based on transfer learning[J]. CAAI transactions on intelligent systems, 2021, 16(3): 397–406.

[6] LONG Mingsheng, CAO Yue, WANG Jianmin, et al. Learning transferable features with deep adaptation networks[C]//32nd International Conference on Machine Learning. Lille: ICML, 2015, 1: 97−105.

[7] LI Shan, DENG Weihong. A deeper look at facial expression dataset bias[J]. IEEE transactions on affective computing, 2022, 13(2): 881–893.

[8] XU Xiaolin, ZHENG Wenming, ZONG Yuan, et al.

Sample self-revised network for cross-dataset facial expression recognition[C]//2022 International Joint Conference on Neural Networks. Padua: IEEE, 2022: 1−8.

[9] XU Xiaolin, ZONG Yuan, LU Cheng, et al. Enhanced sample self-revised network for cross-dataset facial expression recognition[J]. Entropy, 2022, 24(10): 1475.

[10] CHEN Tianshui, PU Tao, WU Hefeng, et al. Cross-domain facial expression recognition: a unified evaluation benchmark and adversarial graph learning[J]. IEEE transactions on pattern analysis and machine intelligence, 2022, 44(12): 9887–9903.

[11] GANIN Y, LEMPITSKY V. Unsupervised domain adaptation by backpropagation[C]//Proceedings of the 32nd International Conference on International Conference on Machine Learning. New York: ACM, 2015: 1180−1189.

[12] LONG Mingsheng, CAO Zhangjie, WANG Jianmin, et al. Conditional adversarial domain adaptation[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal: ACM, 2018: 1647−1657.

[13] WANG Chao, DING Jundi, YAN Hui, et al. A prototype-oriented contrastive adaption network for cross-domain facial expression recognition[C]//Asian Conference on Computer Vision. Cham: Springer, 2023: 324−340.

[14] XIE Yuan, CHEN Tianshui, PU Tao, et al. Adversarial graph representation adaptation for cross-domain facial expression recognition[C]//Proceedings of the 28th ACM International Conference on Multimedia. Seattle: ACM, 2020: 1255−1264.

[15] TIAN Yingli, KANADE T, COHN J F. Recognizing action units for facial expression analysis[J]. IEEE transactions on pattern analysis and machine intelligence, 2001, 23(2): 97–115.

[16] ZHANG Kaipeng, ZHANG Zhanpeng, LI Zhifeng, et al. Joint face detection and alignment using multitask cascaded convolutional networks[J]. IEEE signal processing letters, 2016, 23(10): 1499–1503.

[17] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770−778.

[18] ARNAB A, DEHGHANI M, HEIGOLD G, et al. ViViT: a video vision transformer[C]//2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2022: 6816−6826.

[19] KHAN S, NASEER M, HAYAT M, et al. Transformers in vision: a survey[J]. ACM computing surveys, 2022, 54(10s): 1–41.

[20] 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]//2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco: IEEE, 2010: 94−101.

[21] LYONS M, AKAMATSU S, KAMACHI M, et al. Coding facial expressions with Gabor wavelets[C]//Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition. Nara: IEEE, 2002: 200−205.

[22] DHALL A, GOECKE R, LUCEY S, et al. Static facial expression analysis in tough conditions: data, evaluation protocol and benchmark[C]//2011 IEEE International Conference on Computer Vision Workshops. Barcelona: IEEE, 2012: 2106−2112.

[23] GOODFELLOW I J, ERHAN D, LUC CARRIER P, et al. Challenges in representation learning: a report on three machine learning contests[J]. Neural networks, 2015, 64: 59–63.

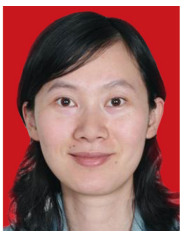
[24] ZHANG Zhanpeng, LUO Ping, LOY C C, et al. Learning social relation traits from face images[C]//2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2016: 3631−3639.

[25] 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. Honolulu: IEEE, 2017: 2584−2593.

[26] LI Shan, DENG Weihong. Deep emotion transfer network for cross-database facial expression recognition[C]//2018 24th International Conference on Pattern Recognition. Beijing: IEEE, 2018: 3092−3099.

[27] LAURENS V D M, HINTON G. Visualizing data using t-SNE[J]. Journal of machine learning research, 2008, 9(2605): 2579–2605.

作者简介：



梁艳，讲师，博士，主要研究方向为计算机视觉、模式识别与智能系统等。发表学术论文 20 余篇。

温兴，硕士研究生，主要研究方向为深度学习、计算机视觉、迁移学习。

潘家辉，教授，博士，中国人工智能学会脑机融合与生物机器智能专业委员会委员，主要研究方向为模式识别与智能系统、脑机交互。主持 3 项国家自然科学基金项目，2 项广东省自然科学基金项目，发表学术论文 60 余篇。