

基于因果干预的无偏面部动作单元识别

邵志文^{1,2,3}, 陈必宽^{1,2}, 祝汉城^{1,2*}, 周 勇^{1,2}, 姚 睿^{1,2}, 马利庄^{3,4}

(1. 中国矿业大学计算机科学与技术学院, 江苏徐州 221116; 2. 矿山数字化教育部工程研究中心, 江苏徐州 221116;
3. 上海交通大学计算机科学与工程系, 上海 200240; 4. 华东师范大学计算机科学与技术学院, 上海 200062)

摘 要: 面部动作单元 (Action Unit, AU) 识别是计算机视觉与情感计算领域的热点课题。AU 识别属于多标签二分类任务, 目前面临着标签不均衡等挑战。现有的主流算法利用 AU 之间的关联, 通过调整采样率和 AU 的权重来进行标签重均衡化。然而, 这些方法仅仅使模型预测时从偏向出现频率高的标签转为偏向出现频率低的标签, 并未解决偏置问题。根据出现频率的高低可将 AU 划分为头类和尾类, 公平对待每一类是实现 AU 无偏识别的关键。本文引入因果推理理论, 提出基于因果干预的无偏化方法 (Causal Intervention for Unbiased facial action unit recognition, CIU), 以解决多 AU 间不均衡的问题。通过调整不平衡域和平衡但不可见域上的经验风险实现模型的无偏性。大量实验结果表明, 本方法在基准数据集 BP4D、DISFA 上超越已有的方法, 其中在 DISFA 上超越当前最先进方法 1.1%, 且可以学习到无偏的特征表示。

关键词: 因果推理; 无偏性; 面部动作单元识别; 多标签二分类; 标签不均衡; 经验风险

基金项目: 国家自然科学基金 (No.62106268, No.62101555, No.62272461, No.62172417, No.72192821); 江苏省自然科学基金 (No.BK20210488, No.BK20201346); 上海市“科技创新行动计划” (No.21511101200); 中国博士后科学基金 (No.2023M732223); “香江学者计划” (No.XJ2023037)

中图分类号: TP391.4

文献标识码: A

文章编号: 0372-2112(2024)10-3312-10

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20240279

Causal Intervention for Unbiased Facial Action Unit Recognition

SHAO Zhi-wen^{1,2,3}, CHEN Bi-kuan^{1,2}, ZHU Han-cheng^{1,2*}, ZHOU Yong^{1,2}, YAO Rui^{1,2}, MA Li-zhuang^{3,4}

(1. School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, Jiangsu 221116, China;
2. Mine Digitization Engineering Research Center of the Ministry of Education, Xuzhou, Jiangsu 221116, China;
3. Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China;
4. School of Computer Science and Technology, East China Normal University, Shanghai 200062, China)

Abstract: Facial action unit (AU) recognition is a hot topic in the fields of computer vision and affective computing. AU recognition is a multi-label binary classification task, and currently faces challenges such as label imbalance. Most existing methods re-balance labels by adjusting the sampling rate and weights of AUs based on the correlations among AUs. However, these methods only shift the model's prediction bias from high-frequency labels to low-frequency ones, and the bias is still unresolved. Fair treatment of each AU class, including the head and tail classes, is the key to achieve unbiased AU recognition. By introducing causal inference theory, we propose an unbiased AU recognition method CIU (Causal Intervention for Unbiased facial action unit recognition), which adjusts the empirical risks in both the imbalanced and balanced but invisible domains to achieve model unbiasedness. Extensive experiments demonstrate that our method outperforms state-of-the-art methods on BP4D and DISFA benchmarks, in which 1.1% margin over previous best method is achieved on DISFA, and can learn unbiased feature representation.

Key words: causal inference; unbiasedness; facial action unit recognition; multi-label binary classification; label imbalance; empirical risk

Foundation Item(s): National Natural Science Foundation of China (No.62106268, No.62101555, No.62272461, No.62172417, No.72192821); Natural Science Foundation of Jiangsu Province (No. BK20210488, No. BK20201346); Shanghai Action Plan for Science, Technology and Innovation (No.21511101200); China Postdoctoral Science Foundation (No.2023M732223); Hong Kong Scholars Program (No.XJ2023037)

1 引言

面部表情是人类表达情感的主要方式之一,对其的分析和识别在智慧医疗^[1]、公共安全^[2]等领域具有广泛的应用前景.面部动作单元(Action Unit, AU)为面部局部区域的基本面部动作,由面部动作编码系统(Facial Action Coding System, FACS)^[3]定义,可以描述面部表情的细粒度变化.近年来, AU 识别受到越来越多研究者的关注,由于 AU 具有细微性、动态性和多样性,其仍是一个挑战性任务.

AU 识别指判定人脸图像中每个 AU 的状态是否激活,属于多标签二分类问题.在人们常见的表情中,某些 AU 出现的频率比其他 AU 更高,比如 AU7、AU10、AU12 等是出现频率高的 AU,而 AU2、AU4 等是出现频率低的 AU.同时,每一 AU 不出现的频率多数情况下高于其出现的频率,因此 AU 标签存在不平衡性^[4]. AU 的人工标注成本较高,导致当前大多数 AU 数据集的规模较小且样本多样性较低,使得 AU 标签的不平衡性更严重. AU 标签的不平衡性会导致 AU 识别模型在同时预测多个 AU 时容易偏向出现频率较高的 AU,而抑制其他 AU 的预测准确性,并且也更倾向于将其他 AU 预测为不出现.

已有的一些 AU 识别方法未处理标签不平衡问题,为不平衡模型,其偏向于出现频率高的 AU,且倾向于预测为出现,预测结果存在大量假阳性(False Positive, FP),降低了预测的精度.为解决标签不平衡问题,许多方法考虑利用 AU 间的关联^[5],实现对标签的重采样^[6,7]和重加权^[8-12],如 Li 等人^[6]通过增加训练集中低频 AU 的随机采样率,确保每个小批量(mini-batch)中 AU 出现频率均衡. Wang 等人^[8]提出了 MSR 损失函数,通过在交叉熵损失函数中引入平衡因子,放大正样本(含有特定 AU 的样本)对损失的影响.由于 F_1 分数指标适用于标签不平衡的情况, Shao 等人^[11]使用基于 F_1 分数指标的频率加权 Dice 系数损失函数. Song 等人^[12]提出自适应加权损失函数等.这些方法虽然在一定程度上缓解了标签不平衡问题,但会让网络的特征学习能力下降.通过增加尾部类别样本的权重,使模型更关注这些类别,可能会导致模型过度偏向尾部类别.

由于在训练过程中存在着混淆因子,平衡模型和不平衡模型均存在偏置.近年来,因果推理方法被广泛应用于其他计算机视觉任务^[13-17],以消除混淆因子,如 Wang 等人^[16]利用后门调整(backdoor adjustment)阻断视觉特征提取中的后门路径,实现了不同目标视觉特征的解耦.但这种方法仅考虑了视觉域的混淆因子,而忽视了语言域的混淆因子,导致虚假关联的后门路径没有被完全阻断,制约了性能的提升. Yang 等人^[17]进一步提出了因果注意力机制,并应用于 Transformer 结

构,虽然这一方法实现简单以及适用性较强,但其忽视了视觉域混淆因子对视觉目标特征的干扰.鉴于此,本文融合经验误差最小化理论^[18]和因果推理理论^[19],提出了基于因果干预的无偏面部动作单元识别(Causal Intervention for Unbiased facial action unit recognition, CIU)方法,以解决混淆因子引起的偏置问题.

具体而言,大多数先前的方法实质上采用了训练集数据分布是不均衡的,而测试集数据分布是均衡的假设,在训练时通过平衡数据来提升模型性能.然而,测试集上的数据分布时常并不均衡,因此平衡后的模型在测试时会更偏向出现频率低的 AU.这实质是不平衡模型和平衡模型的域选择不同而导致对不同频率 AU 的偏好不同,其中,不平衡域经验风险鼓励模型从不平衡的数据分布中学习,这有利于头部类别(出现频率高的 AU);平衡域经验风险鼓励模型从平衡的数据分布中学习,这有利于尾部类别(出现频率低的 AU).本文提出对这两种风险进行自适应加权,从而使得模型既不偏向头部类别也不偏向尾部类别,实现模型的无偏性.

本文的主要贡献包括:

(1) 本文指出了 AU 标签不平衡问题中被忽视的无偏性:模型应该公平对待头部类别和尾部类别,从而在不平衡和平衡的数据分布上都表现良好.

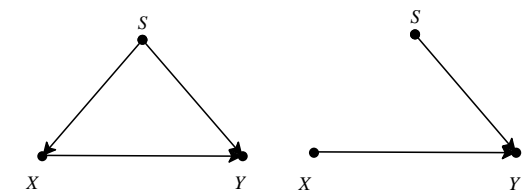
(2) 本文提出一个无偏化的面部动作单元识别方法 CIU,在 BP4D 数据集上取得优秀结果,并且在 DISFA 数据集上优于先前最好的结果,表明 CIU 可以适用于不同的数据分布.

(3) 本文引入因果关系理论,对 CIU 进行了理论分析,消除了混淆因子.

2 算法框架设计

2.1 因果图的构建

为对人脸 AU 识别中的标签不平衡问题进行因果分析,本文首先构建 AU 识别的因果图,如图 1 所示.由于域选择变量 S 引入了伪相关后门路径 $X \leftarrow S \rightarrow Y$,为了实现因果干预,通过 $do(X)$ 切断指向 X 的虚假关联,建模变量之间的因果关系.因果图包含三个节点:图像



(a) 人脸 AU 识别的因果图 (b) 因果干预后的模型

图 1 因果图及因果干预

X 、标签 Y 、域选择变量 S ，其中， S 的每个值代表一个特定的域。通过调整 S 的值，可以在不同的域之间进行切换。我们令 $S=0$ 表示平衡域， $S=1$ 表示不平衡域。图中具体的因果关系如下：

$X \rightarrow Y$: 该路径表示模型根据图像内容预测标签。

$S \rightarrow X$: 该路径表示根据选定的域对图像进行采样。例如，不平衡域容易包含出现频率高的 AU 类别图像。

$S \rightarrow Y$: 该路径表明预测的标签分布遵循其自己的训练域先验。例如，不平衡模型在预测时偏向头部类别，而平衡模型在不同的测试分布中表现出对尾部类别的明显偏好。

这里，后门路径 $X \leftarrow S \rightarrow Y$ 造成了 X 和 Y 之间的虚假关联性^[20]，其中 S 充当混淆因子。后门路径的存在是模型实现无偏识别的关键挑战。简单地学习 $P(Y|X=x)$ 将不可避免地包含这种虚假关联性，并导致模型产生偏置。因此，我们改为估计 $P(Y|\text{do}(X=x))$ ，它可以消除混淆因子带来的偏置以获得无偏模型。

2.2 经验风险的因果干预

本文从因果干预的角度来实现经验风险最小化。施加了因果干预的经验风险 R 被定义为

$$\begin{aligned} R(f) &= \mathbb{E}_{x \sim P(X), y \sim P(Y|\text{do}(X=x))} [L(y, f(x))] \\ &= \sum_x \sum_y L(y, f(x)) P(y|\text{do}(x)) P(x) \end{aligned} \quad (1)$$

其中， $L(y, f(x))$ 被形式化为标准交叉熵。通过因果干预 do 运算，我们切断了指向 x 的路径，并将图 1(a) 中的因果图转换为图 1(b)。与从 $P(X, Y)$ 中采样 (x, y) 不同，我们从干预分布 $P(\text{do}(X), Y)$ 中采样 (x, y) 。根据 do 运算的第二条规则^[20]，即如果干预只影响该变量本身，那么该变量的干预和观察的概率分布也会相同，我们有 $P(\text{do}(X=x)) = P(X=x)$ 。因此，我们可以分两步从干预分布中采样。首先，我们从 $P(x)$ 中采样 x 。其次，我们从干预条件概率 $P(Y|\text{do}(x))$ 对 y 进行采样。为方便起见，我们将 $P(X=x)$ 表示为 $P(x)$ ， $P(Y=y|\text{do}(X=x))$ 表示为 $P(y|\text{do}(x))$ 。

根据后门调整方法^[20]， $P(y|\text{do}(x))$ 可以表示为针对所有混淆变量 S 的求和，而 $\text{do}(x)$ 不影响 S 的分布，且在给定混淆变量 S 的条件下，干预 $\text{do}(x)$ 和观察 x 对 y 的影响是等价的，因此我们将 $P(y|\text{do}(x))$ 重写为

$$\begin{aligned} P(y|\text{do}(x)) &= \sum_{S=s \in \{0,1\}} P(y|\text{do}(x), S=s) P(S=s|\text{do}(x)) \\ &= \sum_{S=s \in \{0,1\}} P(y|x, S=s) P(S=s) \\ &= \frac{P(x, y, S=1)}{P(x|S=1)} + \frac{P(x, y, S=0)}{P(x|S=0)} \end{aligned} \quad (2)$$

这里 $S=1$ 表示可见的不平衡域，即训练样本， $S=0$

表示不可见的平衡域。结合式(1)和式(2)，可以得到

$$\begin{aligned} R(f) &= \frac{1}{N} \sum_{(x,y)} [L(y_{s=1}, f(x)) \frac{P(x)}{P(x|S=1)} \\ &\quad + L(y_{s=0}, f(x)) \frac{P(x)}{P(x|S=0)}] \end{aligned} \quad (3)$$

其中， $y_{s=1}$ 表示来自可见的不平衡域的真实经验标签， $y_{s=0}$ 表示来自不可见的平衡域的样本标签。正如前文所述，平衡域的样本不能直接得到，因此 $y_{s=0}$ 是根据平衡模型 $p^{\text{imba}(y|x)}$ 估计的，并表示为 \hat{y}_i (对于第 i 个类别则表示为 \hat{y}_i)。

为最小化式(3)中跨域经验风险 R ，还需要计算样本权重 $\frac{P(x)}{P(x|s)}$ 。由于 $\frac{P(x)}{P(x|s)} = \frac{P(s)}{P(s|x)}$ ，可以假设 $P(S=1) = P(S=0)$ ，进而可以得到 $\frac{P(x)}{P(x|s)} \propto \frac{1}{P(s|x)}$ 。本质上，它们是样本 x 来自不平衡域 ($S=1$) 和平衡域 ($S=0$) 的可能性的逆概率。受此启发，本文提出使用交叉熵损失 XE 来描述不平衡域和平衡域之间的差异并估计权重：

$$\begin{aligned} w^{\text{imba}} &\propto \frac{1}{P(S=1|x)} \propto (\text{XE}^{\text{imba}})^{\gamma} \\ w^{\text{ba}} &\propto \frac{1}{P(S=0|x)} \propto (\text{XE}^{\text{ba}})^{\gamma} \end{aligned} \quad (4)$$

式(4)表明，如果 x 不适合不平衡模型 p^{imba} ，即 XE^{imba} 较大，则来自不平衡域的可能性较低，即 $P(S=1|x)$ 较小，反之亦然。而且，损失 XE 越大，相应的权重 w 也越大。通过将权重归一化为 1，我们可以计算出自适应权重。

2.3 算法流程

根据式(4)，可以利用自适应权重实现式(1)中经验风险最小化，达到消除混淆因子的目的。我们的方法并没有对测试集的分布作特定假设，因此在理论上适用于任意分布的测试集。算法具体流程如下：

算法 1 基于因果干预的无偏化方法

输入: 训练样本 $\{(x, y)\}$ ，其中 x 为图像， y 为标签。

输出: 无偏分类模型 f 。

- 1: 训练一个平衡模型 p^{imba} 和一个不平衡模型 p^{ba} 。
- 2: 计算不平衡域经验风险权重 w^{imba} 。
- 3: 计算平衡域经验风险权重 w^{ba} 。
- 4: 计算不平衡域经验风险 $R^{\text{imba}}(f)$ 。
- 5: 计算平衡域经验风险 $R^{\text{ba}}(f)$ 。
- 6: 计算总经验风险 $R(f)$ 以训练无偏模型 f 。

如算法 1 所示，不平衡域经验风险权重 w^{imba} 的定义为

$$w^{\text{imba}} = \frac{(\text{XE}^{\text{imba}})^{\gamma}}{(\text{XE}^{\text{imba}})^{\gamma} + (\text{XE}^{\text{ba}})^{\gamma}} \quad (5)$$

平衡域经验风险权重 w^{ba} 的定义为

$$w^{ba} = 1 - w^{imba} = \frac{(XE^{ba})^\gamma}{(XE^{imba})^\gamma + (XE^{ba})^\gamma} \quad (6)$$

其中, $XE^{imba} = -\sum_i y_i \log p^{imba}(y_i|x)$ 是不平衡模型预测结果的交叉熵损失 (XE^{ba} 与之类似), 下标 i 代表第 i 类标签, $\gamma > 0$ 是缩放参数. 这里, 若不平衡模型的预测置信度较高, 即 XE^{imba} 较低, 则它的预测偏置会占主导地位, 我们希望降低其偏置的影响, 因此其所在的域对应的调整权重应自适应地减小, 即与 XE^{imba} 成正相关, 对于平衡模型亦是如此.

在此基础上, 不平衡域经验风险 $R^{imba}(f)$ 的定义为

$$R^{imba}(f) = -w^{imba} \sum_{i=1}^N y_i \log f_i(x) \quad (7)$$

其中, N 是类别的数量, y_i 是第 i 个类别的真实概率标签, 取值为 0 或 1, $f_i(x)$ 是无偏模型 f 对第 i 个类别的预测概率, 其值在 0 和 1 之间. 由于训练数据是不平衡的, 因此我们将不平衡域经验风险权重 w^{imba} 施加在无偏模型的训练过程中, 从而调整不平衡域经验风险.

平衡域经验风险 $R^{ba}(f)$ 的定义为

$$R^{ba}(f) = -w^{ba} \sum_{i=1}^N \hat{y}_i \log f_i(x) \quad (8)$$

其中, \hat{y}_i 是平衡模型对第 i 个类别的预测概率. 由于平衡域实质是不可见的, 我们无法直接从平衡域抽取样本. 幸运的是, 根据因果关系框架^[21], 我们可以通过平衡模型对样本的预测结果这一反事实推理, 来模拟平衡域的样本. 最后我们将平衡域权重施加在最小化无偏模型的预测分布与平衡模型预测分布之间损失的训练过程中, 从而调整平衡域经验风险.

通过融合不平衡域经验风险和平衡域经验风险, 可以得到总经验风险 $R(f)$:

$$R(f) = R^{imba}(f) + R^{ba}(f) \quad (9)$$

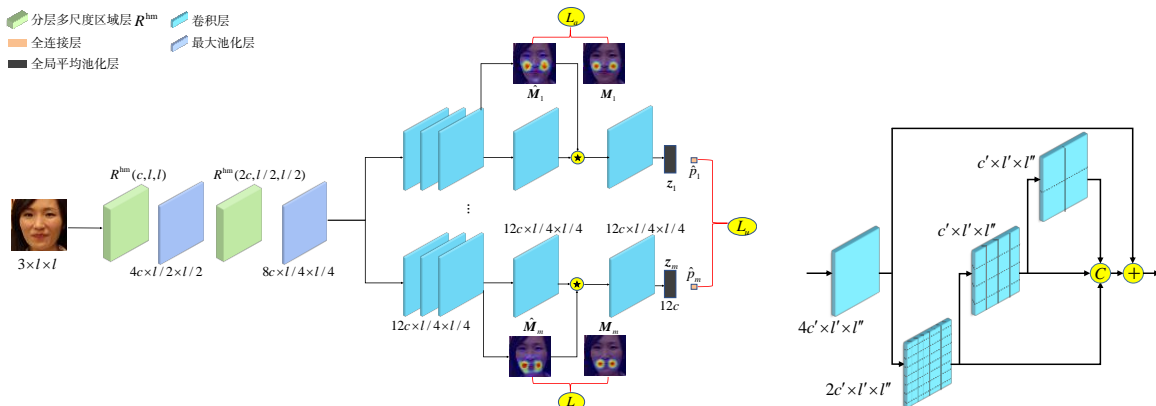
总经验风险将被用来指导无偏模型 f 的训练过程, 以消除域选择混淆因子带来的偏置影响.

3 网络架构设计

根据上述算法, 首先需要训练一个不平衡模型和一个平衡模型, 然后根据它们的预测损失, 利用式(5)和式(6)计算不平衡域权重 w^{imba} 和平衡域权重 w^{ba} . 接着, 通过权重调整各个域的经验风险 $R^{imba}(f)$ 和 $R^{ba}(f)$, 消除域选择混淆因子带来的偏置. 最后将两者融合得到总经验风险 $R(f)$, 以指导无偏模型的训练过程.

本文方法 CIU 所需的平衡模型、不平衡模型, 以及最终生成的无偏模型均采用自适应注意力回归网络, 其结构如图 2(a) 所示. 与 ARL^[9] 方法的结构类似, 给定一张输入图像, 首先经过多个 AU 共享的特征提取模块, 由两个分层多尺度区域层 R^{hm} ^[11] 构成, 用于提取多尺度面部特征, 以适应处于面部不同局部区域的不同大小的 AU, 其中, 每一分层多尺度区域层之后有一个最大池化层.

由于在面部不同局部区域的 AU 具有多样的结构和纹理信息, 不同局部区域应该用不同的滤波器进行处理. 基于此, 分层多尺度区域层的详细结构如图 2(b) 所示, 其由一个普通卷积层和另三个层次化的分块卷积层组成. 具体地, 普通卷积层的特征图被等分为 8×8 个块, 然后每个块分别被第一个分块卷积层的一个独立的卷积滤波器所处理. 以同样的方式, 第二和第三个分块卷积层分别对它们的前一层特征图被等分的 4×4 和 2×2 块采用独立的卷积滤波器. 通过串联第一、第二和第三个分块卷积层的特征图, 我们可以提取一个分层多尺度特征图, 其通道数 $4c'$ 与第一个普通卷积层的特征图通道数相等. 接下来一个残差结构对这两个特征图进行元素级求和, 从而学习完备的特征, 并避免梯度消失问题.



(a) 自适应注意力回归网络模型架构

(b) 分层多尺度区域层 $R^{hm}(c',l',l'')$ 的结构

图2 网络架构

3.1 注意预定义约束

在特征提取模块后,共有 m 个 AU 分支,每一分支同时进行注意图回归和 AU 预测. 具体地,在三个卷积层之后采用一个单通道卷积层来得到输入图像第 j 个 AU 大小为 $l/4 \times l/4$ 的预测注意力图 \hat{M}_j ,它反映了模型对第 j 个 AU 相关联区域的捕捉能力,其中 $j=1, \dots, m$. 考虑到相关的面部特征点可以精确地指定细微 AU 的位置^[11],我们为每个 AU 预定义了真实注意力图 M_j ,其由先验知识定义的 AU 中心生成. 取一个 AU 中心,坐标为 $(\bar{a}_j^{\text{sub1}}, \bar{b}_j^{\text{sub1}})$ 为例,我们采用以其位置为中心的高斯分布来生成一个子注意力图 M_j^{sub1} ,其在位置 (a, b) 的元素为

$$M_{jab}^{\text{sub1}} = \exp\left(-\frac{(a - \bar{a}_j^{\text{sub1}})^2 + (b - \bar{b}_j^{\text{sub1}})^2}{2\delta^2}\right) \quad (10)$$

其中, δ 表示标准差, $M_{jab}^{\text{sub1}} \in (0, 1]$. 然后,通过在每个位置 (a, b) 选择两个注意力权重中的较大者,来合并两个 AU 中心的预定义子注意力图 M_j^{sub1} 和 M_j^{sub2} :

$$M_{jab} = \max(M_{jab}^{\text{sub1}}, M_{jab}^{\text{sub2}}) \quad (11)$$

式(10)和式(11)本质上是将较低的注意力权重分配给离两个 AU 中心较远的位置,如图 2(a)中 M_j 的例子所示. 为使得 \hat{M}_j 保留由特征点指定的 AU 中心附近的强相关区域,采用注意力回归损失来鼓励 \hat{M}_j 接近 M_j :

$$L_a = \frac{\sum_{j=1}^m \sum_{a=1}^{l/4} \sum_{b=1}^{l/4} (\hat{M}_{jab} - M_{jab})^2}{m(l/4 \times l/4)} \quad (12)$$

3.2 AU 识别的引导

为使得 \hat{M}_j 同时能够自适应地捕捉在人脸中全局分布的弱相关区域,且这些区域在不同的人和表情之间有不同的变化,我们将 \hat{M}_j 与第四个卷积层特征图的每个通道按元素相乘,以强调具有较高注意力权重的区域,并通过进一步应用卷积层和全局平均池化层^[22]来获得具有 $12c$ 大小的学习的 AU 特征 z_j . 为了监督每个全局注意力图 \hat{M}_j 的学习,我们最后使用一个一维全连接层,后面跟着 Sigmoid 函数,以获得预测的 AU 出现概率 \hat{p}_j .

由于存在 m 个 AU,每个 AU 有出现和不出现两个类别 ($N=2$),故针对不平衡模型采用多标签二元交叉熵损失:

$$L_u^{\text{imba}} = -\sum_{j=1}^m [p_j \log \hat{p}_j + (1-p_j) \log(1-\hat{p}_j)] \quad (13)$$

针对平衡模型,采用重加权策略,则 AU 识别损失为加权多标签二元交叉熵损失:

$$L_u^{\text{ba}} = -\sum_{j=1}^m w_j [v_j p_j \log \hat{p}_j + (1-p_j) \log(1-\hat{p}_j)] \quad (14)$$

其中, w_j 表示第 j 个 AU 的权重, v_j 表示第 j 个 AU 对应于出现情况的权重,而 p_j 表示输入图像中第 j 个 AU 的真实出现概率. 在大多数 AU 数据集^[23,24]的训练集中,不同的 AU 出现频率显著不同,大多数 AU 的出现频率远低于未出现频率. 为抑制两种类型的数据不平衡问题,将 w_j 和 v_j 定义为

$$w_j = \frac{n}{n_j^{\text{occ}}} / \sum_{k=1}^m \frac{n}{n_k^{\text{occ}}}, v_j = \frac{n - n_j^{\text{occ}}}{n_j^{\text{occ}}} \quad (15)$$

其中, n 和 n_j^{occ} 分别是样本总数和在训练集中第 j 个 AU 出现的样本数,第 j 个 AU 的出现频率为 n_j^{occ}/n .

通过结合式(12)~(14),我们得到了不平衡模型和平衡模型各自的总体损失:

$$\begin{aligned} L_{\text{AA}}^{\text{imba}} &= L_u^{\text{imba}} + \lambda_a L_a \\ L_{\text{AA}}^{\text{ba}} &= L_u^{\text{ba}} + \lambda_a L_a \end{aligned} \quad (16)$$

其中, λ_a 用于衡量注意力回归损失 L_a 的重要性. 通过注意力预定义的约束和 AU 识别的指导,自适应回归的 \hat{M}_j 可以捕捉由特征点指定的 AU 中心附近的强相关区域和离 AU 中心较远的全局分布的弱相关区域,从而适应细微和多样的 AU. 得益于 \hat{M}_j 学习的区域相关性分布,特征 z_j 可以准确地提取每个 AU 的有用信息.

基于式(16)中的不平衡模型总体损失和平衡模型总体损失,可以分别训练一个自适应注意力回归网络,得到不平衡模型和平衡模型. 通过式(5)和式(6),可以得到自适应的不平衡域权重 w^{imba} 和平衡域权重 w^{ba} ; 然后通过式(7)和式(8),以 w^{imba} 和 w^{ba} 分别调整两个域的经验风险 $R^{\text{imba}}(f)$ 和 $R^{\text{ba}}(f)$. 最后利用式(9)将两者融合,得到总经验风险 $R(f)$ 以指导无偏模型进行训练. 无偏模型的总体损失为

$$L_{\text{AA}} = R(f) + \lambda_a L_a \quad (17)$$

由于需要识别多个 AU,因此上述过程中计算 XE^{imba} 、 XE^{ba} 、 $R^{\text{imba}}(f)$ 和 $R^{\text{ba}}(f)$ 时交叉熵损失需扩展为多标签二元交叉熵损失. 基于式(17),我们可以训练一个自适应注意力回归网络,得到既不偏向头部类别、也不偏向尾部类别的无偏模型.

4 实验结果和分析

4.1 实验设置

4.1.1 数据集和评价指标

本文在面部动作单元数据集上进行实验,包括 BP4D^[23] 和 DISFA^[24]. 采用 3-折交叉验证 (3-fold cross-validation), 每次实验中选择两折进行训练,第三折用作测试,每一折中的人脸身份是独立的. BP4D 在 12 个 AU (1、2、4、6、7、10、12、14、15、17、23 和 24) 上评估, DISFA 在 8 个 AU (1、2、4、6、9、12、25 和 26) 上评估. 评估指标采用基于帧的 F_1 分数 (frame-based F_1 -score), 其定义为

$$F_1 = \frac{2PR}{P+R} \quad (18)$$

其中 P 指精确率(precision), R 指召回率(recall). 在 AU 标签出现与不出现频率不均衡的情境下, F_1 分数是一个可信的模型性能度量标准. 为进行更全面的评估, 还计算所有 AU 上的平均 F_1 分数, 用 Avg 表示.

4.1.2 实验参数设置

为了验证本文提出 CIU 方法的有效性, 本文在 NVIDIA GeForce RTX 3090 GPU 上进行实验, 并基于 PyTorch^[25] 深度学习框架实现. 裁剪大小 l 设置为 176, 标准差 δ 设置为 3, 权衡参数 λ_a 设置为 2, 输入训练序列的帧数 t 和中间维数 c^e 分别设置为 48 和 8. 数据集 BP4D, DISFA 的 AU 数分别为 12 和 8. 我们采用随机梯

度下降(Stochastic Gradient Descent, SGD)求解器, 动量^[26]设置为 0.9, 权重衰减为 0.005, 学习率为 0.006. 在 BP4D 数据集上参数 γ 设置为 1, DISFA 数据集上参数 γ 设置为 2.

4.2 实验对比

在相同的评估设置下, 我们将 CIU 与先进的 AU 识别方法进行比较, 一部分采用了缓解标签不均衡的策略, 这些方法包括 DSIN^[27]、EAC-Net^[6]、ARL^[9]、JAA-Net^[11]、SRERL^[28]、AAR^[4]、CMS^[29]、LP-Net^[30]、HRR^[31] 和 GeoCNN^[32]. 另一部分未采用缓解标签不均衡策略, 包括 AU R-CNN^[33]、HMP-PS^[34]、RTATL^[35]. 表 1 和表 2 分别展示了 BP4D 数据集和 DISFA 数据集中 F_1 分数比较, 其中加粗代表最优结果.

表 1 BP4D 数据集上 F_1 分数比较

单位: %

AU		1	2	4	6	7	10	12	14	15	17	23	24	Avg
未缓解 AU 不均衡	AU R-CNN ^[33]	50.2	43.7	57.0	78.5	78.5	82.6	87.0	67.7	49.1	62.4	50.4	49.3	63.0
	HMP-PS ^[34]	53.1	46.1	56.0	76.5	76.9	82.1	86.4	64.8	51.5	63.0	49.9	54.5	63.4
	RTATL ^[35]	57.1	49.7	60.5	77.9	76.1	84.4	87.2	64.3	53.5	67.0	48.9	48.6	64.6
缓解 AU 不均衡	EAC-Net ^[6]	39.0	35.2	48.6	76.1	72.9	81.9	86.2	58.8	37.5	59.1	35.9	35.8	55.9
	DSIN ^[27]	51.7	40.4	56.0	76.1	73.5	79.9	85.4	62.7	37.3	62.9	38.8	41.6	58.9
	CMS ^[29]	49.1	44.1	50.3	79.2	74.7	80.9	88.3	63.9	44.4	60.3	41.4	51.2	60.6
	LP-Net ^[30]	43.4	38.0	54.2	77.1	76.7	83.8	87.2	63.3	45.3	60.5	48.1	54.2	61.0
	ARL ^[9]	45.8	39.8	55.1	75.7	77.2	82.3	86.6	58.8	47.6	62.1	47.7	55.4	61.1
	JAA-Net ^[11]	53.8	47.8	58.2	78.5	75.8	82.7	88.2	63.7	43.3	61.8	45.6	49.9	62.4
	HRR ^[31]	49.3	50.6	51.5	77.5	77.2	81.1	88.4	66.6	50.7	63.1	49.0	55.8	63.4
	GeoCNN ^[32]	48.4	44.2	59.9	78.4	75.6	83.6	86.7	65.0	53.0	64.7	49.5	54.1	63.6
	AAR ^[4]	53.2	47.7	56.7	75.9	79.1	82.9	88.6	60.5	51.5	61.9	51.0	56.8	63.8
	SRERL ^[28]	49.4	42.1	55.5	79.4	78.9	84.5	88.2	67.3	50.5	65.1	50.0	56.5	64.0
	CIU	54.4	50.1	58.4	77.0	78.3	84.3	89.2	61.3	53.3	61.8	52.4	55.0	64.6

通过分析可以发现, CIU 在 AU 识别任务上既优于用了缓解标签不均衡策略, 也优于没用缓解标签不均衡策略的方法. 缓解标签不均衡的 AU 识别方法使用频率加权损失函数或是采用重采样方法, 而这会不可避免地牺牲出现频率高的 AU 的识别性能以提升出现频率低的 AU 的识别性能. 未缓解标签不均衡的方法则会因为数据不均衡现象影响模型性能.

RTATL^[35] 方法虽然与本文的模型 CIU 在 BP4D 上同样取得最佳性能, 但在 DISFA 数据集上的表现有明显的下降. 这是因为 RTATL^[35] 方法并未采用缓解标签不均衡的策略, 而 DISFA 数据集的数据不均衡问题更加严重, 因此性能有所下降. CIU 在 DISFA 数据集上仍然表现优异, 并且与其他方法相比性能有很大提升, 这表明 CIU 有很强的泛化能力, 进一步缓解了标签不均衡. 此外, 从表 1 和表 2 可以看出, CIU 对绝大部分 AU 的识别都取得了最优或次优的结果, 这体现出了模型的无偏性, 即不论是出现频率高的 AU 还是出现频率低

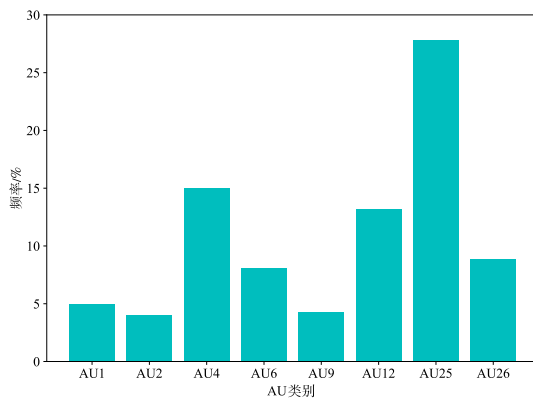
的 AU, CIU 都有很高的识别性能. CIU 作为一个简单而有效的办法, 可以进一步缓解标签不均衡问题, 未来可以将其集成到具有更加复杂网络结构的模型上, 在不增加额外复杂的网络结构的情况下提高模型性能.

为探究 CIU 的无偏性能, 我们比较不平衡模型、平衡模型与 CIU 的真阳性和假阳性结果. 由于 DISFA 数据集的不平衡现象更为严重, 因此我们以 DISFA 数据集为例. 图 3(a) 展示了 DISFA 数据集上的 AU 出现频率. 如图 3(b) 所示, 可以发现不平衡模型在 AU12、AU25 等高频率 AU 的预测上具有较高的假阳性结果, 这是因为数据集中存在数据不平衡的现象, 高频率 AU 的权重贡献会更大, 因此不平衡模型偏向预测高频率 AU 出现. 而平衡模型因为采用了 AU 加权平衡损失函数, 对 AU12、AU25 等高频率 AU 的预测结果并没有显著的假阳性. 然而, 平衡模型会对低频率 AU 的出现更有信心, 比如在 AU9 的预测结果上, 平衡模型有十分突出的假阳性结果. 相比之下, 本文提出的 CIU 在借助因

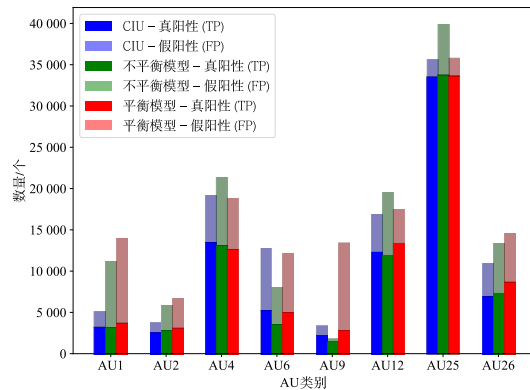
表2 DISFA数据集上 F_1 分数比较

单位: %

AU		1	2	4	6	9	12	25	26	Avg
未缓解 AU 不平衡	AU R-CNN ^[33]	32.1	25.9	59.8	55.3	39.8	67.7	77.4	52.6	51.3
	HMP-PS ^[34]	38.0	45.9	65.2	50.9	50.8	76.0	93.3	67.6	61.0
	RTATL ^[35]	57.8	52.8	70.8	53.2	52.7	74.5	91.5	51.9	63.1
缓解 AU 不平衡	EAC-Net ^[6]	41.5	26.4	66.4	50.7	8.5	89.3	88.9	15.6	48.5
	DSIN ^[27]	42.4	39.0	68.4	28.6	46.8	70.8	90.4	42.2	53.6
	LP-Net ^[30]	29.9	24.7	72.7	46.8	49.6	72.9	93.8	65.0	56.9
	CMS ^[29]	40.2	44.3	53.2	57.1	50.3	73.5	81.1	59.7	57.4
	SRERL ^[28]	43.8	46.2	67.3	50.1	42.4	71.2	93.5	54.3	58.6
	ARL ^[9]	43.9	42.1	63.6	41.8	40.0	76.2	95.2	66.8	58.7
	GeoCNN ^[32]	65.5	65.8	67.2	48.6	51.4	72.6	80.9	44.9	62.1
	HRR ^[31]	57.1	61.8	69.6	35.1	48.6	73.8	95.4	64.0	63.2
	JAA-Net ^[11]	62.4	60.7	67.1	41.1	45.1	73.5	90.9	67.4	63.5
	AAR ^[4]	62.4	53.6	71.5	39.0	48.8	76.1	91.3	70.6	64.2
	CIU	56.0	58.3	72.5	45.6	56.8	73.8	92.3	63.9	64.9



(a) DISFA 中 AU 出现频率



(b) DISFA 数据集上真阳性和假阳性比较

图3 DISFA 数据集的 AU 出现频率以及不同模型的预测结果

果干预消除混淆因子后,不论在低频率 AU(AU1、AU2、AU9)还是在高频率 AU(AU12、AU25)上假阳性的比例并不高,实现了高精度预测,体现了 CIU 的无偏性.

4.3 消融实验

在本节中,在 BP4D 数据集上验证本文提出的 CIU 框架中主要组件的合理性和有效性,包括自适应权重模块和跨域经验风险最小化模块.

4.3.1 自适应权重模块

在自适应权重模块中,权重 w^{imba} 和 w^{ba} 控制两个交叉熵损失之间的平衡,而经验参数 $\gamma=0$ 则调节 w^{imba} 和 w^{ba} 的影响力. 当 $\gamma=20$ 的绝对值很大时,权重 w^{imba} 和 w^{ba} 会更容易倾向于 0 或 1. 当 $\gamma=0$ 时, w^{imba} 和 w^{ba} 恒定为 0.5,不再产生自适应变化. 通过调节 γ ,我们可以探究自适应权重模块的重要性.

如图 4 所示,当 $\gamma=1$ 时, CIU 的预测精度最高. 当 $\gamma=0$ 时,由于 w^{imba} 和 w^{ba} 不会随式(5)和式(6)的计算结

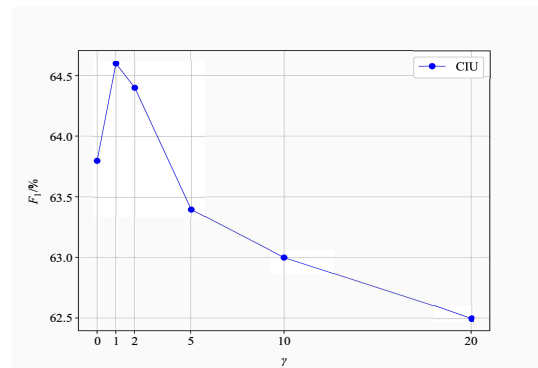


图4 BP4D 数据集上不同的 γ 值模型预测精度比较

果而发生变化,因而预测精度与 $\gamma=1$ 相比下降了 1.2%. 当进一步增大 γ , w^{ba} 和 w^{imba} 会由于影响力的增大而易于极端偏向 0 或 1,即一个权重过分大而另一个权重过分小,具体地,当 $\gamma=5$ 、 $\gamma=10$ 、 $\gamma=20$ 时,预测精度分别下降了 1.8%、2.5%、3.3%.

4.3.2 跨域经验风险最小化模块

跨域经验风险包括不平衡域经验风险和平衡域经验风险. 为了验证跨域经验风险最小化模块的有效性和重要性,我们将 CIU 拆分成 TE(只包含不平衡域经验风险)和 TDE(只包含平衡域经验风险). 如表 3 所示, CIU 较 TDE 在 BP4D 数据集上的预测精度提升

了 34.9%,这是因为 TDE 只是让无偏模型从平衡模型的预测结果学习,而不是学习真实标签,因此没有很好地学习到面部动作单元的特征. CIU 较 TE 的预测精度提升了 3.1%,这是因为 TE 没有让模型采取缓解标签不平衡策略,受面部动作单元标签不均衡的影响,预测精度有所下降.

表 3 BP4D 数据集上 AU 识别消融实验结果

单位:%

AU	1	2	4	6	7	10	12	14	15	17	23	24	Avg
TDE	44.6	38.5	50.3	60.5	57.2	67.6	78.9	31.1	26.9	52.5	28.4	38.6	47.9
TE	44.5	41.3	56.2	77.6	78.5	83.6	89	63.3	53.2	64.2	45.0	54.6	62.6
CIU	54.4	50.1	58.4	77.0	78.3	84.3	89.2	61.3	53.3	61.8	52.4	55.0	64.6

4.4 可解释性分析

为进一步探究模型的特征学习能力,本文尝试对模型的注意力机制进行可视化,使用热力图来显示面部动作单元相应的激活区域,如图 5 所示. 从图中可以看出, EAC-Net 中对每个 AU 的注意力具有固定的分布,且只分布在具有固定大小的预定义感兴趣区域(Region of Interest, ROI),这会忽略 ROI 之外的潜在相关区域. 对于 JAA-Net,它对不同 AU 的注意力显得过于宽泛,同时对于某些 AU,比如 AU4、AU10 的特征提取并不准确. 与 EAC-Net 相比, CIU 不仅正确的注意到每个 AU 对应的 ROI 区域,同时还会注意到一些相关联的区域.

例如 AU1(内眉毛提升)和 AU7(眼轮匝肌内圈放大)经常同时出现,因此 CIU 在识别 AU1 的时候,除了会关注 AU1 的区域,还会关注上眼眶的区域. 与 JAA-Net 相比, CIU 的注意力热图显得更为紧凑,这表明 CIU 有更好的特征学习能力. 同时, AU7、AU14 是出现频率较高的 AU,但 JAA-Net 对其施加的注意力权重并不高,这是因为 JAA-Net 采取了缓解标签不均衡的方法,但这会让模型预测时偏向出现频率不高的 AU 而抑制出现频率高的 AU. 而 CIU 对每个 AU 都施加同样重要的注意力权重,对出现频率不同的 AU 同等对待,体现了 CIU 的无偏性.

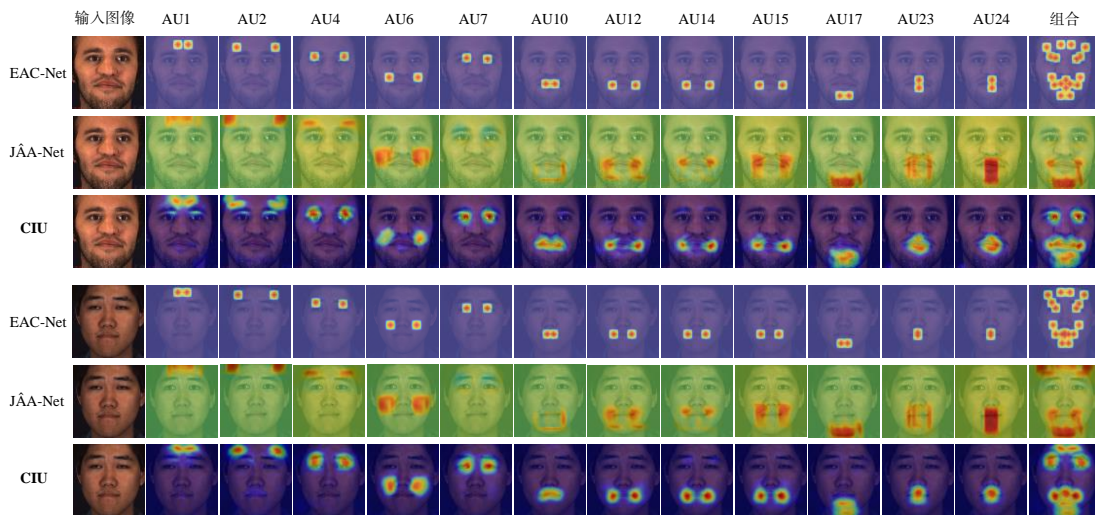


图 5 注意力机制可视化示意图

5 总结与展望

本文针对面部 AU 识别中的标签不均衡问题,提出了一种基于因果干预的无偏多 AU 识别方法 CIU. 研究强调了无偏性的重要性,即模型应当公平对待出现频率高的 AU 和频率低的 AU. CIU 方法在两个基准数据集 BP4D 和 DISFA 上的测试表明,它能有效提升面部 AU 识别的准确性,并超越现有的先进方法. 此外,通过消融实验,验证了

自适应权重模块和跨域经验风险最小化模块的重要性,进一步支持了 CIU 方法的合理性和有效性.

虽然本研究在基准数据集上对面部 AU 的识别取得了超越以往方法的性能提升,然而,其准确度仍有较大的进步空间,尚未达到实际应用所需的水平. 未来可以从数据增强技术和利用大量未标注的真实场景数据等方面进一步探索,增强模型的鲁棒性和准确性.

参考文献

- [1] 张晶, 王翌歆, 任永功. 统一全局空间表达的脑电信号跨被试情感识别[J]. 电子学报, 2023, 51(5): 1396-1404.
ZHANG J, WANG Y X, REN Y G. Unified global spatial representation for EEG subject-independent emotion recognition[J]. *Acta Electronica Sinica*, 2023, 51(5): 1396-1404. (in Chinese)
- [2] FRANK M G, EKMAN P. The ability to detect deceit generalizes across different types of high-stake lies[J]. *Journal of Personality and Social Psychology*, 1997, 72(6): 1429-1439.
- [3] EKMAN P, FRIESEN W V. *Facial Action Coding System: A Technique for the Measurement of Facial Movement* [M]. Palo Alto: Consulting Psychologists Press, 1978.
- [4] SHAO Z W, ZHOU Y, CAI J F, et al. Facial action unit detection via adaptive attention and relation[J]. *IEEE Transactions on Image Processing (TIP)*, 2023, 32: 3354-3366.
- [5] 邵志文, 周勇, 谭鑫, 等. 基于深度学习的表情动作单元识别综述[J]. 电子学报, 2022, 50(8): 2003-2017.
SHAO Z W, ZHOU Y, TAN X, et al. Survey of expression action unit recognition based on deep learning[J]. *Acta Electronica Sinica*, 2022, 50(8): 2003-2017. (in Chinese)
- [6] LI W, ABTAHI F, ZHU Z G, et al. EAC-net: Deep nets with enhancing and cropping for facial action unit detection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2018, 40(11): 2583-2596.
- [7] LI Y Q, WU B Y, ZHAO Y P, et al. Handling missing labels and class imbalance challenges simultaneously for facial action unit recognition[J]. *Multimedia Tools and Applications*, 2019, 78(14): 20309-20332.
- [8] WANG S J, LIN B, WANG Y, et al. Action units recognition based on deep spatial-convolutional and multi-label residual network[J]. *Neurocomputing*, 2019, 359: 130-138.
- [9] SHAO Z W, LIU Z L, CAI J F, et al. Facial action unit detection using attention and relation learning[J]. *IEEE Transactions on Affective Computing (TAFFC)*, 2022, 13(3): 1274-1289.
- [10] TALLEC G, DAPOGNY A, BAILLY K. Fighting noise and imbalance in action unit detection problems[EB/OL]. (2023-03-06)[2024-07-13]. <https://arxiv.org/abs/2303.02994>.
- [11] SHAO Z W, LIU Z L, CAI J F, et al. JAA-net: Joint facial action unit detection and face alignment via adaptive attention[J]. *International Journal of Computer Vision (IJCV)*, 2021, 129(2): 321-340.
- [12] SONG T F, CHEN L S, ZHENG W M, et al. Uncertain graph neural networks for facial action unit detection[C]// *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*. Washington: AAAI, 2021: 5993-6001.
- [13] ZHU B E, NIU Y L, HUA X S, et al. Cross-domain empirical risk minimization for unbiased long-tailed classification[C]// *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*. Washington: AAAI, 2022: 3589-3597.
- [14] 刘兵, 李穗, 刘明明, 等. 基于全局与序列混合变分Transformer的多样化图像描述生成方法[J]. 电子学报, 2024, 52(4): 1305-1314.
LIU B, LI S, LIU M M, et al. Diverse image captioning based on hybrid global and sequential variational transformer[J]. *Acta Electronica Sinica*, 2024, 52(4): 1305-1314. (in Chinese)
- [15] 卓亚琦, 魏家辉, 李志欣. 基于双注意模型的图像描述生成方法研究[J]. 电子学报, 2022, 50(5): 1123-1130.
ZHUO Y Q, WEI J H, LI Z X. Research on image captioning based on double attention model[J]. *Acta Electronica Sinica*, 2022, 50(5): 1123-1130. (in Chinese)
- [16] WANG T, HUANG J Q, ZHANG H W, et al. Visual commonsense R-CNN[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway: IEEE, 2020: 10760-10770.
- [17] YANG X, ZHANG H W, QI G J, et al. Causal attention for vision-language tasks[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway: IEEE, 2021: 9847-9857.
- [18] JUNG Y H, TIAN J, BAREINBOIM E. Learning causal effects via weighted empirical risk minimization[C]// *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*. Cambridge: MIT Press, 2020: 12697-12709.
- [19] PEARL J. *Causality: Models, Reasoning, and Inference* [M]. 2nd ed. Cambridge: Cambridge University Press, 2009.
- [20] PEARL J, GLYMOUR M, JEWELL N P. *Causal Inference in Statistics: A Primer* [M]. Reprinted ed. with revisions. Chichester: Wiley, 2021.
- [21] TANG K H, HUANG J Q, ZHANG H W. Long-tailed classification by keeping the good and removing the bad momentum causal effect[C]// *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*. Cambridge: MIT Press, 2020: 1513-1524.
- [22] LIN M, Chen Q, Yan S C. Network in network[C]// *Proceedings of the International Conference on Learning Representations (ICLR)*. Alameda: OpenReview, 2014: 1-10.
- [23] ZHANG X, YIN L J, COHN J F, et al. BP4D-Spontaneous: A high-resolution spontaneous 3D dynamic facial expression database[J]. *Image and Vision Computing (IVC)*, 2014, 32(10): 692-706.
- [24] MAVADATI S M, MAHOOR M H, BARTLETT K, et al. DISFA: A spontaneous facial action intensity database[J].

IEEE Transactions on Affective Computing (TAFFC), 2013, 4(2): 151-160.

- [25] PASZKE A, GROSS S, MASSA F, et al. Pytorch: An imperative style, high-performance deep learning library [C]//Proceedings of the Advances in Neural Information Processing Systems (NeurIPS). Cambridge: MIT Press, 2019: 8024-8035.
- [26] SUTSKEVER I, MARTENS J, DAHL G, et al. On the importance of initialization and momentum in deep learning[C]//Proceedings of the International Conference on Machine Learning (ICML). Stroudsburg: IMLS, 2013: 1139-1147.
- [27] CORNEANU C, MADADI M, ESCALERA S. Deep structure inference network for facial action unit recognition [C]//Proceedings of the European Conference on Computer Vision (ECCV). Berlin: Springer, 2018: 309-324.
- [28] 李冠彬, 张锐斐, 朱鑫, 等. 语义关系引导的面部动作单元分析[J]. 软件学报, 2023, 34(6): 2922-2941.
LI G B, ZHANG R F, ZHU X, et al. Semantic relationships guided facial action unit analysis[J]. Journal of Software, 2023, 34(6): 2922-2941. (in Chinese)
- [29] SANKARAN N, MOHAN D D, SETLUR S, et al. Representation learning through cross-modality supervision[C]//Proceedings of the IEEE International Conference on Automatic Face & Gesture Recognition (FG). Piscataway: IEEE, 2019: 1-8.
- [30] NIU X S, HAN H, YANG S F, et al. Local relationship learning with person-specific shape regularization for facial action unit detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019: 11917-11926.
- [31] SHAO Z W, ZHOU Y, LIU B, et al. Facial action unit detection via hybrid relational reasoning[J]. The Visual Computer (TVC), 2022, 38(9): 3045-3057.
- [32] CHEN Y D, SONG G X, SHAO Z W, et al. GeoConv: Geodesic guided convolution for facial action unit recognition[J]. Pattern Recognition (PR), 2022, 122: 108355.
- [33] MA C, CHEN L, YONG J H. AU R-CNN: Encoding expert prior knowledge into R-CNN for action unit detection [J]. Neurocomputing, 2019, 355: 35-47.
- [34] SONG T F, CUI Z J, ZHENG W M, et al. Hybrid message passing with performance-driven structures for facial action unit detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 6263-6272.
- [35] YAN J W, WANG J J, LI Q, et al. Self-supervised regional and temporal auxiliary tasks for facial action unit recognition [C]//Proceedings of the ACM International Conference on

Multimedia (MM). New York: ACM, 2021: 1038-1046.

作者简介



邵志文 男, 1994年12月出生于安徽省马鞍山市. 现为中国矿业大学计算机科学与技术学院副教授、硕士生导师. 主要研究方向为情感计算、计算机视觉和人工智能.

E-mail: zhiwen_shao@cumt.edu.cn



陈必宽 男, 2003年5月出生于广东省江门市. 现为中国矿业大学计算机科学与技术学院本科生. 主要研究方向为计算机视觉.

E-mail: bikuan_chen@cumt.edu.cn



祝汉城 男, 1989年12月出生于江苏省徐州市. 现为中国矿业大学计算机科学与技术学院副教授、硕士生导师. 主要研究方向为情感计算、计算机视觉和人工智能.

E-mail: zhuhancheng@cumt.edu.cn



周勇 男, 1974年9月出生于江苏省徐州市. 现为中国矿业大学计算机科学与技术学院教授、博士生导师. 主要研究方向为计算机视觉、机器学习和人工智能.

E-mail: yzhou@cumt.edu.cn



姚睿 男, 1982年7月出生于河南省南阳市. 现为中国矿业大学计算机科学与技术学院教授、博士生导师. 主要研究方向为计算机视觉和人工智能.

E-mail: ruiyao@cumt.edu.cn



马利庄 男, 1963年2月出生于浙江省宁波市. 现为上海交通大学计算机科学与工程系教授、博士生导师. 主要研究方向为计算机视觉、计算机图形学和人工智能.

E-mail: ma-lz@cs.sjtu.edu.cn