

寻找用于跨数据库微表情识别的显著人脸区域*

江星洵 宗源[†] 郑文明[†] 刘佳腾 魏梦婷

生物科学与医学工程学院, 东南大学, 南京, 中国

{jiangxingxun,xhzhongyuan,wenming_zheng,jiateng_liu,weimengting}@seu.edu.cn

摘要

跨数据库微表情识别旨在开发出具有强领域适应能力的微表情识别方法, 即具备识别在不同场景下、由不同设备捕获的、不同被试的微表情的能力。跨数据库微表情识别技术的发展面临着两个关键难题: 1) 源域和目标域数据库间存在着严重的特征分布差异; 2) 现有的微表情表征方法存在瓶颈, 不能很好地表征微表情这种局部的、轻微的面部表情。为了解决这两个难题, 本文提出了一种新的组稀疏迁移回归方法, 简称 *TGSR*。*TGSR* 能够挑选出用于跨数据库微表情识别的显著面部区域, 通过特征层次上的稀疏化操作, 以便 1) 优化源域与目标域数据库间差异的衡量方法, 达到更好地缓解源域与目标域间差异的目的; 2) 突出已抽取特征中的有效部分, 使抽取的特征更有效、更具可解释性。我们在两个公开数据库上, 即 *CASME II* 和 *SMIC* 数据库上进行了实验, 来评估所提的 *TGSR* 方法。实验结果表明, 我们提出的 *TGSR* 方法能够学习到可解释的、判别性强的微表情特征, 且其识别性能能够超过目前大部分基于子空间学习和领域自适应技巧的跨数据库微表情识别方法。

1. 引言

微表情是一种运动幅度低、持续时间短的面部表情, 它能够反映被试者隐藏的真实情感 [2, 3]。微表情在诸如刑侦 [4]、临床诊断 [5]、人机交互 [6, 7, 8, 9]

等领域有着重要意义。

由于微表情巨大的潜在价值, 在过去数十年中, 研究者在自动微表情识别课题上做出了许多努力 [10, 11, 12, 13, 14, 15, 16, 17, 18], 推动了自动微表情识别技术的发展。在传统的模式识别方法中, 特征抽取是自动微表情识别中最为重要的步骤。研究者们聚焦于此, 提出了如 LBP-TOP [19]、LBP-SIP [11]、FDM [16] 和 FHOFO 在内诸多优秀的手工特征。这些特征能够促使分类器更好地识别微表情类别。在深度学习方法 [20, 21, 22, 23] 中, 微表情的特征提取与分类是一个端到端的过程。它通过层次化的特征学习方式, 促使模型学到更有效的特征, 提升模型整体的识别性能。过去的研究推动自动微表情识别技术取得了长足的发展。然而现在的方法多是基于单个微表情数据库进行估计的, 这使得现有的模型在识别在不同场景下、由不同设备捕获的、不同被试的微表情时, 即识别来自不同领域的微表情时, 可能无法表现出令人满意的性能。

为了学习到鲁棒的微表情识别模型, 近年来, 研究者们开始将研究兴趣转向领域自适应的微表情识别方法上。由此也诞生了一个颇具挑战性的新课题, 即跨数据库微表情识别 (CDMER)。跨数据库微表情识别通过在一个微表情数据库 (源数据库) 上训练模型, 在另一个微表情数据库 (目标数据库) 上测试该模型性能的方式, 模拟现实生活中微表情识别面临的领域变化问题, 评估所学模型的领域适应性能力。跨数据库微表情识别技术 [24, 25, 26] 面临着两个关键难题: 1) 源域数据库和目标域数据库间存在着严重的特征分布差异, 2) 现有的微表情表征

*本文为 ICPR'22 [1] 的中文翻译版。

[†]指通讯作者

约束项 $f_1(\mathbf{C}_i)$ 是领域间差异消除项。它通过最小化源域和目标域微表情特征分布差异估计量的方式，来缩小和缓解这种差异。如式2所示的最大平均差异 (MMD) 可以用来表示 $f_1(\mathbf{C}_i)$ 所表征的差异，

$$MMD(\mathbf{X}^s, \mathbf{X}^t) = \left\| \frac{1}{N_s} \sum_{i=1}^K \Phi(\mathbf{X}_i^s) \mathbf{1}_s - \frac{1}{N_t} \sum_{i=1}^K \Phi(\mathbf{X}_i^t) \mathbf{1}_t \right\|_{\mathcal{H}}, \quad (2)$$

其中， $\Phi(\cdot)$ 是用于将特征从原始空间映射到无穷维空间的核化操作， $\mathbf{1}_s \in \mathbb{R}^{N_s}$ 和 $\mathbf{1}_t \in \mathbb{R}^{N_t}$ 分别是将源域和目标域特征矩阵转换为标量值的全1向量（向量的每个元素都为1）。然而，带有核化操作的式子是不易求解的，所以我们将消除领域间差异的约束项 $f_1(\mathbf{C}_i)$ ，进一步修改为式3，

$$f_1(\mathbf{C}_i) = \left\| \frac{1}{N_s} \sum_{i=1}^K \mathbf{C}_i^T \mathbf{X}_i^s \mathbf{1}_s - \frac{1}{N_t} \sum_{i=1}^K \mathbf{C}_i^T \mathbf{X}_i^t \mathbf{1}_t \right\|_F^2. \quad (3)$$

式3通过将源域和目标域之间的核化特征差异放松为标签空间差异的方式，实现式2的求解。同时，通过面部区域选择的方式来增强微表情的特征表示，优化如式3所示的源域目标域间差异的衡量方式，从而达到更好地缓解该差异的目的。如式4所示的约束项 $f_2(\mathbf{C}_i)$ 是回归矩阵 \mathbf{C} 的 Frobenius 范数稀疏项，它用来改善所提取的微表情特征，使其更加有效、更具可解释性，

$$f_2(\mathbf{C}_i) = \lambda \sum_{i=1}^K \|\mathbf{C}_i\|_F. \quad (4)$$

通过将式3和式4带入式1的方式，我们将目标函数重写为式5，

$$\min_{\mathbf{C}_i} \left\| \mathbf{L}^s - \sum_{i=1}^K \mathbf{C}_i^T \mathbf{X}_i^s \right\|_F^2 + \lambda \sum_{i=1}^K \|\mathbf{C}_i\|_F + \xi \left\| \frac{1}{N_s} \sum_{i=1}^K \mathbf{C}_i^T \mathbf{X}_i^s \mathbf{1}_s - \frac{1}{N_t} \sum_{i=1}^K \mathbf{C}_i^T \mathbf{X}_i^t \mathbf{1}_t \right\|_F^2. \quad (5)$$

2.3. 优化过程

我们可以使用交替方向法 (ADM) [30] 和非精准增广拉格朗日乘子法 (IALM) [31] 来求解式5。我

们首先将式5重写为式6，

$$\min_{\mathbf{C}_i} \left\| \tilde{\mathbf{L}} - \sum_{i=1}^K \mathbf{C}_i^T \tilde{\mathbf{X}}_i \right\|_F^2 + \lambda \sum_{i=1}^K \|\mathbf{C}_i\|_F, \quad (6)$$

其中， $\tilde{\mathbf{L}} = [\mathbf{L}^s, \mathbf{0}]$ ， $\mathbf{0} \in \mathbb{R}^{c \times 1}$ ， $\tilde{\mathbf{X}}_i = [\mathbf{X}_i^s, \sqrt{\xi}(\frac{1}{N_s} \mathbf{X}_i^s \mathbf{1}_s - \frac{1}{N_t} \mathbf{X}_i^t \mathbf{1}_t)]$ 。然后，我们引入一个与变量 $\mathbf{C} = [\mathbf{C}_1^T, \dots, \mathbf{C}_K^T]^T$ 相等的新变量 $\mathbf{D} = [\mathbf{D}_1^T, \dots, \mathbf{D}_K^T]^T$ ，将优化问题式6转化为如式7所示受约束的优化形式，

$$\min_{\mathbf{C}, \mathbf{D}} \left\| \tilde{\mathbf{L}} - \sum_{i=1}^K \mathbf{D}_i^T \tilde{\mathbf{X}}_i \right\|_F^2 + \lambda \sum_{i=1}^K \|\mathbf{C}_i\|_F. \quad (7)$$

s. t. $\mathbf{D}_i = \mathbf{C}_i$.

随后，我们可以写出式7对应的增广拉格朗日函数，如式8所示，

$$\Gamma(\mathbf{C}_i, \mathbf{D}_i, \mathbf{P}_i, \mu) = \left\| \tilde{\mathbf{L}} - \sum_{i=1}^K \mathbf{D}_i^T \tilde{\mathbf{X}}_i \right\|_F^2 + \lambda \sum_{i=1}^K \|\mathbf{C}_i\|_F + \sum_{i=1}^K \text{tr}[\mathbf{P}_i^T (\mathbf{C}_i - \mathbf{D}_i)] + \frac{\mu}{2} \sum_{i=1}^K \|\mathbf{C}_i - \mathbf{D}_i\|_F^2. \quad (8)$$

其中， $\mathbf{P}_i \in \mathbb{R}^{d \times c}$ 表示第 i 个人脸局部序列对应的拉格朗日乘子矩阵， μ 是需要设置的超参权重系数。

我们可以通过迭代更新 \mathbf{C}_i 和 \mathbf{D}_i 的方式，来求得最小化式8时的回归矩阵 \mathbf{C}_i 的最优解 $\hat{\mathbf{C}}_i$ 。具体而言，我们需要重复以下四个步骤：

1) 固定 $\mathbf{C}, \mathbf{P}, \mu$ ，更新 \mathbf{D} ：

在本步骤中，与 \mathbf{D} 的子矩阵 \mathbf{D}_i 相关的优化问题可写为式9。它的闭式解如式12所示，

$$\min_{\mathbf{D}} \left\| \tilde{\mathbf{L}} - \mathbf{D}^T \tilde{\mathbf{X}} \right\|_F^2 + \text{tr}[\mathbf{P}^T (\mathbf{C} - \mathbf{D})] + \frac{\mu}{2} \|\mathbf{C} - \mathbf{D}\|_F^2, \quad (9)$$

其中， $\mathbf{P}^T = [\mathbf{P}_1^T, \dots, \mathbf{P}_K^T]$ ， $\mathbf{P} \in \mathbb{R}^{Kd \times c}$ ， $\mathbf{P}_j \in \mathbb{R}^{d \times c}$ 。

2) 固定 $\mathbf{D}, \mathbf{P}, \mu$ 更新 \mathbf{C} ：

在本步骤中，与 \mathbf{C} 的子矩阵 \mathbf{C}_i 相关的优化问题可以写为式10，

$$\min_{\mathbf{C}_i} \lambda \sum_{i=1}^K \|\mathbf{C}_i\|_F + \sum_{i=1}^K \text{tr}[\mathbf{P}_i^T (\mathbf{C}_i - \mathbf{D}_i)] + \frac{\mu}{2} \sum_{i=1}^K \|\mathbf{C}_i - \mathbf{D}_i\|_F^2. \quad (10)$$

我们可将式10转化为式11，并使用式13求解回归矩阵 \mathbf{C} 的最优解。

$$\min_{\mathbf{C}_i} \sum_{i=1}^K \left(\frac{\lambda}{\mu} \|\mathbf{C}_i\|_F + \frac{1}{2} \left\| \mathbf{C}_i - \left(\mathbf{D}_i - \frac{\mathbf{P}_i}{\mu} \right) \right\|_F^2 \right) \quad (11)$$

3) 更新 \mathbf{P} 和 μ 。

4) 检查 $\|\mathbf{C} - \mathbf{D}\|_\infty < \varepsilon$ 是否收敛。

更多的优化细节如算法 1 所示。

Algorithm 1 求解 TGSR 模型中回归矩阵 \mathbf{C} 最优解的算法

输入: 数据矩阵 $\tilde{\mathbf{L}}$ 和 $\tilde{\mathbf{X}} = [\tilde{\mathbf{X}}_1^T, \dots, \tilde{\mathbf{X}}_K^T]^T$, 显著的面部局部区域数 κ , 数值型参数 ρ, μ_{max} 。

- 初始化回归矩阵 $\mathbf{C} = [\mathbf{C}_1^T, \dots, \mathbf{C}_K^T]^T$
- 初始化朗格朗日乘子矩阵 $\mathbf{P} = [\mathbf{P}_1^T, \dots, \mathbf{P}_K^T]^T$ 和加权系数 μ 。

重复步骤 1) 到 4) 直到收敛。

1: 固定 $\mathbf{C}, \mathbf{P}, \mu$, 更新 \mathbf{D} :

$$\mathbf{D} = \left(\mu \mathbf{I}_{Kd} + 2\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T \right)^{-1} \left(2\tilde{\mathbf{X}}\tilde{\mathbf{L}}^T + \mathbf{P} + \mu\mathbf{C} \right), \quad (12)$$

其中, $\mathbf{I}_{Kd} \in \mathbb{R}^{Kd \times Kd}$ 是单位矩阵。

2: 固定 $\mathbf{D}, \mathbf{P}, \mu$, 更新 \mathbf{C} :

计算 $d_i = \left\| \mathbf{D}_i - \frac{\mathbf{P}_i}{\mu} \right\|_F$, 对 d_i 的值排序, 使得 $d_{i_1} \geq d_{i_2} \geq \dots \geq d_{i_\kappa}$, 令 $\lambda = \mu d_{i_{\kappa+1}}$, 然后按照下式更新 \mathbf{C}

$$\mathbf{C}_i = \begin{cases} \frac{d_i - \frac{\lambda}{\mu}}{d_i} \left(\mathbf{D}_i - \frac{\mathbf{P}_i}{\mu} \right), & \frac{\lambda}{\mu} < d_i, \\ \mathbf{0}, & \frac{\lambda}{\mu} \geq d_i. \end{cases} \quad (13)$$

3: 更新 \mathbf{P} 和 μ :

$$\mathbf{P} = \mathbf{P} + \mu (\mathbf{D} - \mathbf{C}), \mu = \min(\rho\mu, \mu_{max})$$

4: 检查收敛:

$$\|\mathbf{C} - \mathbf{D}\|_\infty < \varepsilon$$

输出: 回归矩阵 \mathbf{C} 的最优解 $\hat{\mathbf{C}}$ 。

2.4. 在跨数据库表情中的应用

基于已标注的源域和未标注的目标域数据库, 我们可以通过上述优化方法, 求解 TGSR 方法中回归矩阵 \mathbf{C} 的最优解 $\hat{\mathbf{C}}$ 。然后, 我们提取微表情样本的特征 $\mathbf{x}_i^{te} \in \mathbb{R}^{Kd}$, 通过求解如式14的优化问题, 来预测该样本标签向量 \mathbf{l}^{te} 。

$$\begin{aligned} \min_{\mathbf{l}^{te}} & \left\| \mathbf{l}^{te} - \sum_{i=1}^K \hat{\mathbf{C}}_i^T \mathbf{x}_i^{te} \right\|_F^2, \\ \text{s. t.} & \quad \mathbf{l}^{te} \geq \mathbf{0}, \mathbf{1}^T \mathbf{l}^{te} = 1, \end{aligned} \quad (14)$$

其中, $\hat{\mathbf{C}}_i \in \mathbb{R}^{d \times C}$ 是连接第 i 个人脸局部及其样本标签的回归矩阵 \mathbf{C}_i 的最优解, 且有 $\hat{\mathbf{C}}^T = [\hat{\mathbf{C}}_1^T, \dots, \hat{\mathbf{C}}_K^T]$, $\hat{\mathbf{C}}^T \in \mathbb{R}^{C \times Kd}$, $\mathbf{l}^{te} \in \mathbb{R}^C$ 。随后, 我们可以用 $\hat{c} = \arg \max_j \{l_j^{te}\}$, 来预测测试集样本的标签, 并将其标记为第 \hat{c} 类微表情。

表 1. 经过挑选的 CASME II (Selected CASME II) 和 SMIC 数据库的统计信息。

数据库	种类		
	正性	负性	惊讶
Selected CASME II	32	73	25
SMIC-HS	51	70	43
SMIC-VIS	23	28	20
SMIC-NIR	23	28	20

3. 实验

3.1. 实验设置

数据库. 我们在经过挑选的 CASME II 数据库 (Selected CASME II) 和 SMIC 数据库上进行了实验。CASME II 数据库 [27] 由来自 26 位被试的 255 个微表情样本组成, 包含了厌恶, 害怕, 开心, 其他, 沮丧, 伤心, 和惊讶共七类情感, 我们挑选其中标签样本为厌恶, 害怕, 开心, 惊讶的样本作为 Selected CASME II 数据集。SMIC 数据库 [28] 记录了来自 16 位被试 3 种模态的 306 个微表情样本。具体而言, SMIC 的 HS 子集包含了 164 个由 100 帧/秒的高速摄像头捕捉的微表情样本; VIS 子集包含了 71 个由 25 帧/秒的普通视觉相机捕捉的微表情样本; NIR

表 2. 第一类跨数据库微表情识别实验，即基于 SMIC 数据库中 HS、VIS、NIR 任意两个子集的实验结果。微表情种类包括：负性、正性和惊讶。每个实验中最好的实验结果加粗显示。我们用 M-F1 表示 Macro F1-score 指标，用 ACC 表示 Accuracy 指标。

方法	Exp.1(H→V)		Exp.2(V→H)		Exp.3(H→N)		Exp.4(N→H)		Exp.5(V→N)		Exp.6(N→V)	
	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC
SVM[32]	0.8002	80.28	0.5421	54.27	0.5455	53.52	0.4878	54.88	0.6186	63.38	0.6078	63.38
IW-SVM[33]	0.8868	88.73	0.5852	58.54	0.7469	74.65	0.5427	54.27	0.6620	69.01	0.7228	73.24
TCA[34]	0.8269	83.10	0.5477	54.88	0.5828	59.15	0.5443	57.32	0.5810	61.97	0.6598	67.61
GFK[35]	0.8448	84.51	0.5957	59.15	0.6977	70.42	0.6197	62.80	0.7619	76.06	0.8142	81.69
SA[36]	0.8037	80.28	0.5955	59.15	0.7465	74.65	0.5644	56.10	0.7004	71.83	0.7394	74.65
STM[37]	0.8253	83.10	0.5059	51.22	0.6628	66.20	0.5351	56.10	0.6427	67.61	0.6922	70.42
TKL[38]	0.7742	77.46	0.5738	57.32	0.7051	70.42	0.6116	62.20	0.7558	76.06	0.7580	76.06
TSRG[39]	0.8869	88.73	0.5652	56.71	0.6484	64.79	0.5770	57.93	0.7056	70.42	0.8116	81.69
DRLS[40]	0.8604	85.92	0.6120	60.98	0.6599	66.20	0.5599	55.49	0.6620	69.01	0.5771	61.97
我们的方法	0.9150	91.55	0.6226	62.20	0.5847	60.56	0.6272	61.59	0.6984	70.42	0.8403	84.51

表 3. 第二类跨数据库微表情识别实验，即基于 Selected CASME II 数据库和 SMIC 数据库某一个子集的实验结果。微表情种类包括：负性、正性和惊讶。每个实验中最好的实验结果加粗显示。我们用 M-F1 表示 Macro F1-score 指标，用 ACC 表示 Accuracy 指标。

方法	Exp.7(C→H)		Exp.8(H→C)		Exp.9(C→V)		Exp.10(V→C)		Exp.11(C→N)		Exp.12(N→C)	
	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC
SVM[32]	0.3697	45.12	0.3245	48.46	0.4701	50.70	0.5367	53.08	0.5295	52.11	0.2368	23.85
IW-SVM[33]	0.3541	41.46	0.5829	62.31	0.5778	59.15	0.5537	54.62	0.5117	50.70	0.3456	36.15
TCA[34]	0.4637	46.34	0.4870	53.08	0.6834	69.01	0.5789	59.23	0.4992	50.70	0.3937	42.31
GFK[35]	0.4126	46.95	0.4776	50.77	0.6361	66.20	0.6056	61.50	0.5180	53.52	0.4469	46.92
SA[36]	0.4302	47.56	0.5447	62.31	0.5939	59.15	0.5243	51.54	0.4738	47.89	0.3592	36.92
STM[37]	0.3640	43.90	0.6115	63.85	0.4051	52.11	0.2715	30.00	0.3523	42.25	0.3850	41.54
TKL[38]	0.4582	46.95	0.4661	54.62	0.6042	60.56	0.5378	53.08	0.5392	54.93	0.4248	43.85
TSRG[39]	0.5042	51.83	0.5171	60.77	0.5935	59.15	0.6208	63.08	0.5624	56.34	0.4105	46.15
DRLS[40]	0.4924	53.05	0.5267	59.23	0.5757	57.75	0.5942	60.00	0.4885	49.83	0.3838	42.37
我们的方法	0.5001	51.83	0.5061	56.92	0.5906	59.15	0.6403	63.85	0.5697	57.75	0.4474	48.46

子集捕捉了 71 个由近红外摄像头捕捉的微表情样本。为了使经过挑选的 CASME II 数据库 (Selected CASME II) 和 SMIC 数据库的标签类别相同，我们对经过挑选的 CASME II 数据库 (Selected CASME II) 的标签进行了转化：将开心的标签重新标注为正性，将厌恶和沮丧标签重新标注为负性，保持惊讶标签为惊讶。表 1 总结了经过挑选的 CASME II 数据库 (Selected CASME II) 和 SMIC 数据库的基本信息。

协议. 我们用 S→T 表示在源数据库 (S) 训练，在目标数据库 (T) 测试的跨数据库微表情识别任务。与

[29] 相同，我们设置了两类无监督的跨数据库微表情识别实验：第一类 (TYPE-I) 是在 SMIC 数据库的两个子集上进行的，第二类 (TYPE-II) 是在挑选的 CASME II 数据集 (Selected CASME II) 和 SMIC 数据库的某一个子集上进行的。为方便表达，我们将 SMIC-HS、SMIC-VIS、SMIC-NIR 和 Selected CASME II 分别简写做 H、V、N 和 C。那么具体而言，第一类 (TYPE-I) 跨数据库微表情识别实验包含了 H→V、V→H、H→N、N→H、V→N、N→V，一共 6 个任务；第二类 (TYPE-II) 跨数据库微表情识别实验包含了 C→H、H→C、C→V、V→C、C→N、N→C，一共 6 个任

务。

评价指标. 我们使用 macro F1-score (M-F1) 和 Accuracy (ACC) 来评估我们所提的 TGSR 方法。M-F1 由 $M-F1 = \frac{1}{C} \sum_{c=1}^C \frac{2p_c r_c}{p_c + r_c}$ 计算而得, 其中, p_c 和 r_c 分别代表第 c 类微表情的识别精度 (precision) 和召回率 (recall), C 是微表情的类别数。由于跨数据库微表情识别任务中的样本不平衡问题较为广泛, 所以使用 M-F1 作为评价指标是合适的。

实施细节. 在实验中, 我们使用微表情图片序列第一帧的人脸关键点信息来构建人脸边界框, 并利用该边界框裁剪出序列中的各帧人脸。然后, 在时序上, 我们使用时间插值模型 (TIM)[41, 42] 将微表情序列转换为固定的 16 帧, 在空间上, 我们将图片的大小调整 (resize) 为 112×112 平方像素。我们采用基于网格的多尺度空间划分方案, 将整个人脸划分为 $1 \times 1, 2 \times 2, 4 \times 4, 8 \times 8$ 四种尺度共计 85 个人脸局部序列, 抽取并将每个人脸局部序列的 LBP-TOP 特征 [19] 拼接起来, 作为微表情样本的特征表示。其中, LBP-TOP 特征的相邻半径取 $R = 3$, 相邻点个数取 $P = 8$ 。在求解我们所提 TGSR 方法的过程中, 涉及到两个超参的设置, 即挑选的人脸局部区域数 κ 和 MMD 项的加权系数 ξ 。这里, 挑选的人脸局部区域块数 κ 是一个整数, 而 MMD 项的加权系数 ξ 是一个连续的数。与工作 [29, 39] 相似, 我们使用网格搜索的方法来搜索模型取得最好 M-F1 性能时的超参, 并同时报告对应超参下模型的 ACC 指标。具体而言, 我们在预设的超参空间 [1:1:85] 中搜索超参 κ 的最优值, 在预设的超参空间 [0.001:0.0001:0.01:0.002:0.1 0.1:0.02:1:0.2:10 10:2:100 100:20:1000] 中搜索超参 ξ 的最优值。

3.2. 结果和分析

表2和表3分别展示了第一类 (TYPE-I) 和第二类 (TYPE-II) 跨数据库微表情识别实验的结果。我们计算并列出了两类实验六个任务平均的 M-F1 和 ACC 性能。在第一类实验中 (Exp.1 至 Exp.6), SVM[32]、IW-SVM[33]、TCA[34]、GFK[35]、SA[36]、

STM[37]、TKL[38]、TSRG[39]、DRLS[40], 在 M-F1 指标上分别提升到了 0.6003, 0.6911, 0.6238, 0.7223, 0.6917, 0.6440, 0.6964, 0.6991, 0.6552, 在 ACC 指标上分别提升到了 61.62%, 69.74%, 64.01%, 72.44%, 69.44%, 65.78%, 69.92%, 70.05%, 66.60%。在第二类实验中 (Exp.7 至 Exp.12), SVM[32]、IW-SVM[33]、TCA[34]、GFK[35]、SA[36]、STM[37]、TKL[38]、TSRG[39]、DRLS[40], 在 M-F1 指标上分别提升到了 0.4112, 0.4876, 0.5177, 0.5161, 0.4877, 0.3982, 0.5051, 0.5348, 0.5102, 在 ACC 指标上分别提升到了 45.55%, 50.73%, 53.45%, 54.31%, 50.90%, 45.61%, 52.33%, 56.22%, 53.71%。我们将表2和表3中各实验任务最好的结果加粗显示。从中我们可以看出, 我们所提的 TGSR 方法能够在 12 个中的 7 个任务中, 即超过一半的跨数据库微表情识别任务中, 取得当前最好的结果。在第一类 (TYPE-I) 跨数据库微表情识别任务中, 从 Exp.1 到 Exp.6, 我们所提的 TGSR 方法分别在超参 (κ, ξ) 取 (85, 0.0022), (46, 0.0036), (14, 4000), (85, 0.0044), (12, 44), (12, 280) 处, 取得最好的 M-F1 性能指标。在第二类 (TYPE-II) 跨数据库微表情识别任务中, 从 Exp.7 到 Exp.12, 我们所提的 TGSR 方法分别在超参 (κ, ξ) 取 (62, 0.0012), (28, 0.0980), (85, 0.0030), (85, 0.0280), (85, 0.0016), (75, 0.0220) 处, 取得最好的 M-F1 性能指标。

我们还可以从表2和表3中发现三个明显的特点。首先, 除 SVM 外的所有方法都使用了领域自适应技巧, 它们在所有任务的 M-F1 和 ACC 指标上, 都稳定地表现出优于 SVM 的性能。以 IW-SVM 为例, 它通过学习一组领域自适应加权系数, 有效改进了 SVM 的性能, 在第一类实验的 M-F1 和 ACC 指标上分别提高了 0.0908 和 8.12%, 在第二类实验的 M-F1 和 ACC 指标上分别提高了 0.0764 和 5.18%。因此我们推断, 领域自适应技巧能够有效地提高跨数据库微表情识别模型的性能。其次, 第一类实验 (TYPE-I) 的结果明显优于第二类实验 (TYPE-II)。我们认为, 这是由两类实验的设置本身造成的。第一类跨数据库微表情识别任务是从 SMIC 数据库中选择两种不同成像模态的样本分别作为源

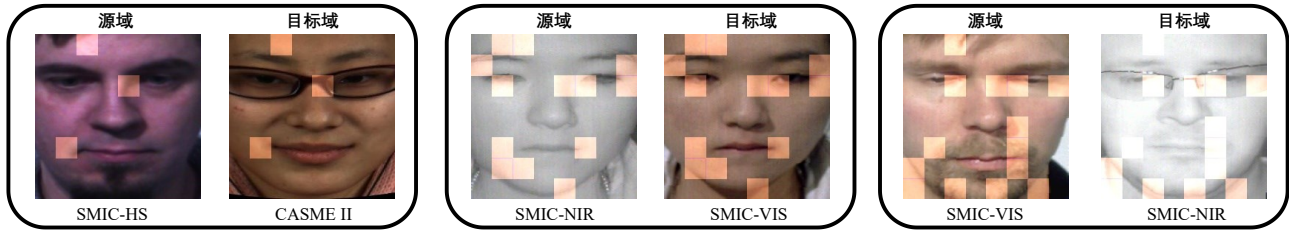


图 2. 所提的 TGSR 方法在三个跨数据库微表情识别任务中所挑选出的显著面部区域。该高亮区域对应的回归矩阵 \hat{C}_i 为全 1 矩阵，非高亮区域对应的回归矩阵 \hat{C}_i 为全 0 矩阵。

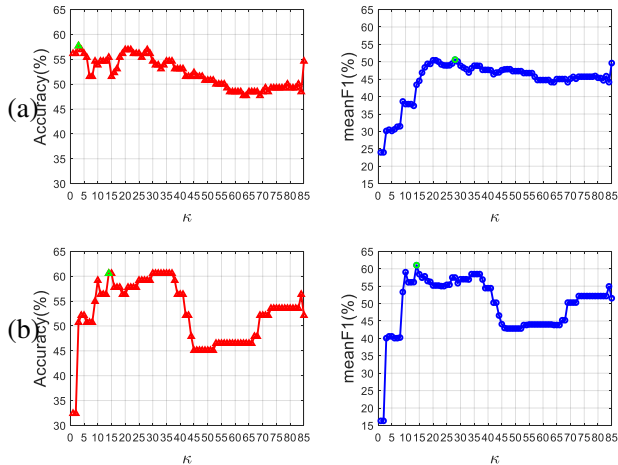


图 3. 模型在不同超参数 κ ，即不同的人脸局部区域选择块数，下的性能曲线图。(a) 展示的是 Exp. 8(H→C) 的实验结果，(b) 展示的是 Exp. 4(H→N) 的实验结果。

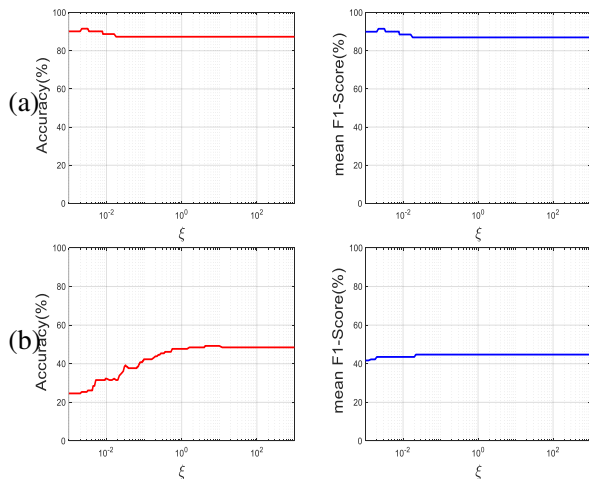


图 4. 模型在不同超参数 ξ ，即不同的 MMD 项加权系数，下的性能曲线图。(a) 展示的是 Exp. 1(H→V) 的实验结果，(b) 展示的是 Exp. 12(N→C) 的实验结果。

域和目标域数据库；第二类跨数据库微表情识别任务是将挑选的 CASME II 数据库 (Selected CASME II) 和 SMIC 数据库中的某一子集分别作为源域和目标域数据库。这样的实验设置中，第一类实验 (TYPE-I) 中源域和目标域的差异明显大于第二类实验 (TYPE-II) 中源域和目标域之间的差异。第三，我们发现模型在交换源域和目标域数据库的实验上存在着一定的性能差异：各方法在 Exp. 1(H→V) 上的实验结果普遍优于 Exp. 2(V→H) 上的实验结果；在 Exp. 3(H→N) 上的实验结果普遍优于 Exp. 4(N→H) 上的实验结果；在 Exp. 11(C→N) 上的实验结果普遍优于 Exp. 12(N→C) 上的实验结果。Exp. 1(H→V) 使用高速摄像头捕获的微表情样本作为源域数据库，使用普通视觉相机捕获的微表情样本作为目标域数据库；Exp. 2(V→H) 与 Exp. 1(H→V) 相反。Exp. 3(H→N) 使用高速摄像头捕获的微表情样本作为源域数据库，使用不带颜色信息的近红外摄像头捕获的微表情样本作为目标域数据库；Exp. 4(N→H) 与 Exp. 3(H→N) 相反。Exp. 11(C→N) 使用的是带有颜色信息的 Selected CASME II 数据库作为源域数据库，使用不带颜色信息的 SMIC 数据库 NIR 子集作为目标域数据库；Exp. 12(N→C) 与 Exp. 11(C→N) 相反。我们认为，Exp. 1(H→V) 的结果优于 Exp. 2(V→H)、Exp. 3(H→N) 的结果优于 Exp. 4(N→H) 的原因是高帧率的设备能够更加有效地捕捉到巧妙的面部肌肉运动变化；而 Exp. 11(C→N) 的结果优于 Exp. 12(N→C) 的原因是因为 Selected CASME II 数据库保留了对于情感理解重要的颜色信息 [43]。另外，虽然 Exp. 3(N→H)-Exp. 4(N→H) 组合和 Exp. 11(C→N)-Exp. 12(N→C) 组合的源域和目标域数据库同样存在颜色差异，但

模型在 Exp.11(C→N) 与 Exp.12(N→C) 之间的性能差是明显大于 Exp.3(N→H) 和 Exp.4(N→H) 之间的性能差的。我们猜测, 这是因为 Exp.11(C→N)-Exp.12(N→C) 组合较之 Exp.3(N→H)-Exp.4(N→H) 组合, 除了颜色信息外, 还包含其他的领域差异信息。

3.3. 超参讨论和可视化

在求解文章所提 TGSR 模型优化矩阵 C 的过程中, 涉及两个超参的设置, 即挑选的显著人脸局部区域块数 κ 和 MMD 项的加权系数 ξ 。显而易见, 这两个超参的设置会影响模型性能的发挥。因此, 我们做了两个实验来调研所提的 TGSR 模型对于超参 κ 和 ξ 的敏感性, 观察模型是否会在较大的超参范围内表现出稳定的性能。

在第一个实验中, 我们将超参 ξ 固定, 让挑选的人脸局部区域块数 κ 在 1 到 85 的范围内变化, 记录下模型在不同超参下对应的 M-F1 和 ACC 性能指标, 用来探索模型对于超参 κ 的敏感性。我们选择选 Exp.4(H→N) 和 Exp.8(H→C), 分别作为第一类 (TYPE-I) 和第二类 (TYPE-II) 跨数据库微表情识别实验的典型任务, 来探讨模型对于超参 κ 的敏感性, 其 M-F1 和 ACC 的性能曲线如图3所示。我们可以清楚地看到, 模型的性能先是随着 κ 的上升而上升, 在较低 κ 值时便达到了性能的顶峰, 然后随着 κ 值的上升反而呈现出下降的趋势。这表明, 用于跨数据库微表情识别的显著面部区域块数是稀少的, 且无用的噪声数据会淹没有效特征。在第二个实验中, 我们将超参 κ 固定, 让 MMD 项的加权系数在 10^{-3} 到 10^3 范围内变化, 记录下模型在不同超参下对应的 M-F1 和 ACC 性能指标, 用来探索模型对于超参 ξ 的敏感性。我们选择 Exp.1(H→V) 和 Exp.12(N→C), 分别作为第一类 (TYPE-I) 和第二类 (TYPE-II) 跨数据库微表情识别实验的典型任务, 来探讨模型关于超参 ξ 的敏感性, 其 M-F1 和 ACC 的性能曲线如图4所示。从图4中我们可以看到, 挑选合适的 ξ 值有助于模型表现出更好的性能, 且 MMD 项可以在较大的 ξ 范围内稳定有效地改善模型性能。为了验证我们所提的 TGSR 方法是否挑选到了合适、正确的人

脸局部区域, 我们选择了 Exp.5(V→N), Exp.6(N→V) 和 Exp.8(H→C) 作为典型, 将模型所挑选出用于跨数据库微表情识别的显著人脸局部区域可视化, 如图2所示。通过图2分析, 我们可以认为, 我们所提的 TGSR 方法可解释性强、性能优越。

4. 总结

本文提出了一种新的组稀疏迁移回归方法, 即 TGSR, 来处理跨数据库微表情识别问题。TGSR 通过一个在源域和目标域数据库间共享的、可学习的二进制回归矩阵来挑选用于跨数据库微表情识别的显著人脸局部区域, 通过特征层面的操作, 以便 1) 在标签空间上优化源域和目标域数据库间差异的衡量方法, 从而更好地缓解源域和目标域间的差异, 2) 突出已抽取特征中的有效部分, 使抽取的特征更有效、更具可解释性。实验和可视化结果表明, TGSR 能够学习到可解释的、判别力强的微表情识别特征, 且 TGSR 模型的识别性能能够超过目前大部分的基于子空间学习和领域自适应技巧的跨数据库微表情识别方法。

致谢. 本论文受到国家自然科学基金委 (编号 U2003207, 编号 61921004, 和编号 61902064)、中央高校基础科研经费的资助 (编号: 2242022k30036) 和东南大学至善青年学者经费的资助。

参考文献

- [1] Xingxun Jiang, Yuan Zong, Wenming Zheng, Jiateng Liu, and Mengting Wei. Seeking salient facial regions for cross-database micro-expression recognition. In *2022 26th International Conference on Pattern Recognition (ICPR)*, pages xxxx–xxxx. IEEE, 2022. 1
- [2] Paul Ekman and Wallace V Friesen. Nonverbal leakage and clues to deception. *Psychiatry*, 32(1):88–106, 1969. 1
- [3] Paul Ekman. *Telling lies: Clues to deceit in the marketplace, politics, and marriage (revised edition)*. WW Norton & Company, 2009. 1
- [4] Wen-Jing Yan, Qi Wu, Jing Liang, Yu-Hsin Chen, and Xiaolan Fu. How fast are the leaked facial expressions: The duration of micro-expressions. *Journal of Nonverbal Behavior*, 37(4):217–230, 2013. 1
- [5] Mark Frank, Malgorzata Herbasz, Kang Sinuk, A Keller, and Courtney Nolan. I see how you feel: Training laypeople and professionals to recognize fleeting emotions. In *The Annual Meeting of the International Communication Association. Sheraton New York, New York City*, pages 1–35, 2009. 1
- [6] Xingxun Jiang, Yuan Zong, Wenming Zheng, Chuangao Tang, Wanchuang Xia, Cheng Lu, and Jiateng Liu. Dfew: A large-scale database for recognizing dynamic facial expressions in the wild. In *Proc. ACM MM*, pages 2881–2889, 2020. 1
- [7] Sunan Li, Wenming Zheng, Yuan Zong, Cheng Lu, Chuangao Tang, Xingxun Jiang, Jiateng Liu, and Wanchuang Xia. Bi-modality fusion for emotion recognition in the wild. In *Proc. ICMI*, pages 589–594, 2019. 1
- [8] Chenguang Yang, Huaiwei Wu, Zhijun Li, Wei He, Ning Wang, and Chun-Yi Su. Mind control of a robotic arm with visual fusion technology. *IEEE Trans. Ind. Informat.*, 14(9):3822–3830, 2017. 1
- [9] Chenguang Yang, Jing Luo, Yongping Pan, Zhi Liu, and Chun-Yi Su. Personalized variable gain control with tremor attenuation for robot teleoperation. *IEEE Trans. Syst., Man, Cybern.*, 48(10):1759–1770, 2017. 1
- [10] Tomas Pfister, Xiaobai Li, Guoying Zhao, and Matti Pietikäinen. Recognising spontaneous facial micro-expressions. In *Proc. ICCV*, pages 1449–1456. IEEE, 2011. 1
- [11] Yandan Wang, John See, Raphael C-W Phan, and Yee-Hui Oh. Lbp with six intersection points: Reducing redundant information in lbp-top for micro-expression recognition. In *Proc. ACCV*, pages 525–537. Springer, 2014. 1
- [12] Su-Jing Wang, Wen-Jing Yan, Xiaobai Li, Guoying Zhao, Chun-Guang Zhou, Xiaolan Fu, Minghao Yang, and Jianhua Tao. Micro-expression recognition using color spaces. *IEEE Trans. Image Process.*, 24(12):6034–6047, 2015. 1
- [13] Yong-Jin Liu, Jin-Kai Zhang, Wen-Jing Yan, Su-Jing Wang, Guoying Zhao, and Xiaolan Fu. A main directional mean optical flow feature for spontaneous micro-expression recognition. *IEEE Trans. Affect. Comput.*, 7(4):299–310, 2015. 1
- [14] Dae Hoe Kim, Wissam J Baddar, and Yong Man Ro. Micro-expression recognition with expression-state constrained spatio-temporal feature representations. In *Proc. ACM MM*, pages 382–386, 2016. 1
- [15] Ping Lu, Wenming Zheng, Ziyang Wang, Qiang Li, Yuan Zong, Minghai Xin, and Lenan Wu. Micro-expression recognition by regression model and group sparse spatio-temporal feature learning. *IEICE Transactions on Information and Systems*, 99(6):1694–1697, 2016. 1
- [16] Feng Xu, Junping Zhang, and James Z Wang. Microexpression identification and categorization using a facial dynamics map. *IEEE Trans. Affect. Comput.*, 8(2):254–267, 2017. 1
- [17] SL Happy and Aurobinda Routray. Fuzzy histogram of optical flow orientations for micro-expression recognition. *IEEE Trans. Affect. Comput.*, 10(3):394–406, 2017. 1
- [18] Yuan Zong, Xiaohua Huang, Wenming Zheng, Zhen Cui, and Guoying Zhao. Learning from hierarchical spatiotemporal descriptors for micro-expression recognition. *IEEE Trans. Multimedia*, 20(11):3160–3172, 2018. 1
- [19] Guoying Zhao and Matti Pietikainen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(6):915–928, 2007. 1, 6
- [20] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Proc. NeurIPS*, 25:1097–1105, 2012. 1

- [21] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997. 1
- [22] Mengting Wei, Wenming Zheng, Yuan Zong, Xingxun Jiang, Cheng Lu, and Jiateng Liu. A novel micro-expression recognition approach using attention-based magnification-adaptive networks. In *Proc. ICASSP*, pages 2420–2424. IEEE, 2022. 1
- [23] Wanchuang Xia, Wenming Zheng, Yuan Zong, and Xingxun Jiang. Motion attention deep transfer network for cross-database micro-expression recognition. In *Proc. ICPR*, pages 679–693. Springer, 2021. 1
- [24] Yuan Zong, Wenming Zheng, Zhen Cui, Guoying Zhao, and Bin Hu. Toward bridging microexpressions from different domains. *IEEE Transactions on Cybernetics*, 50(12):5047–5060, 2019. 1
- [25] Lingyan Li, Xiaoyan Zhou, Yuan Zong, Wenming Zheng, Xiuzhen Chen, Jingang Shi, and Peng Song. Unsupervised cross-database micro-expression recognition using target-adapted least-squares regression. *IEICE Transactions on Information and Systems*, 102(7):1417–1421, 2019. 1
- [26] Xiuzhen Chen, Xiaoyan Zhou, Cheng Lu, Yuan Zong, Wenming Zheng, and Chuangao Tang. Target-adapted subspace learning for cross-corpus speech emotion recognition. *IEICE Transactions on Information and Systems*, 102(12):2632–2636, 2019. 1
- [27] Wen-Jing Yan, Xiaobai Li, Su-Jing Wang, Guoying Zhao, Yong-Jin Liu, Yu-Hsin Chen, and Xiaolan Fu. Casme ii: An improved spontaneous micro-expression database and the baseline evaluation. *PLOS One*, 9(1):e86041, 2014. 2, 4
- [28] Xiaobai Li, Tomas Pfister, Xiaohua Huang, Guoying Zhao, and Matti Pietikäinen. A spontaneous micro-expression database: Inducement, collection and baseline. In *Proc. FG*, pages 1–6. IEEE, 2013. 2, 4
- [29] Tong Zhang, Yuan Zong, Wenming Zheng, CL Philip Chen, Xiaopeng Hong, Chuangao Tang, Zhen Cui, and Guoying Zhao. Cross-database micro-expression recognition: A benchmark. *IEEE Trans. Knowl. Data Eng.*, 2020. 2, 5, 6
- [30] Zhiwei Tony Qin and Donald Goldfarb. Structured sparsity via alternating direction methods. *Journal of Machine Learning Research*, 13(5), 2012. 3
- [31] Zhouchen Lin, Minming Chen, and Yi Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055*, 2010. 3
- [32] Chih-Chung Chang and Chih-Jen Lin. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3):1–27, 2011. 5, 6
- [33] Ali Hassan, Robert Damper, and Mahesan Niranjan. On acoustic emotion recognition: compensating for covariate shift. *IEEE Trans. Speech Audio Process.*, 21(7):1458–1468, 2013. 5, 6
- [34] Sinno Jialin Pan, Ivor W Tsang, James T Kwok, and Qiang Yang. Domain adaptation via transfer component analysis. *IEEE Trans. Neural Netw.*, 22(2):199–210, 2010. 5, 6
- [35] Boqing Gong, Yuan Shi, Fei Sha, and Kristen Grauman. Geodesic flow kernel for unsupervised domain adaptation. In *Proc. CVPR*, pages 2066–2073. IEEE, 2012. 5, 6
- [36] Basura Fernando, Amaury Habrard, Marc Sebban, and Tinne Tuytelaars. Unsupervised visual domain adaptation using subspace alignment. In *Proc. ICCV*, pages 2960–2967, 2013. 5, 6
- [37] Wen-Sheng Chu, Fernando De la Torre, and Jeffery F Cohn. Selective transfer machine for personalized facial action unit detection. In *Proc. CVPR*, pages 3515–3522, 2013. 5, 6
- [38] Mingsheng Long, Jianmin Wang, Jianguang Sun, and S Yu Philip. Domain invariant transfer kernel learning. *IEEE Trans. Knowl. Data Eng.*, 27(6):1519–1532, 2014. 5, 6
- [39] Yuan Zong, Xiaohua Huang, Wenming Zheng, Zhen Cui, and Guoying Zhao. Learning a target sample re-generator for cross-database micro-expression recognition. In *Proc. ACM MM*, pages 872–880, 2017. 5, 6
- [40] Yuan Zong, Wenming Zheng, Xiaohua Huang, Jingang Shi, Zhen Cui, and Guoying Zhao. Domain regeneration for cross-database micro-expression recognition. *IEEE Trans. Image Process.*, 27(5):2484–2498, 2018. 5, 6
- [41] Ziheng Zhou, Guoying Zhao, and Matti Pietikäinen. Towards a practical lipreading system. In *Proc. CVPR*, pages 137–144. IEEE, 2011. 6
- [42] Ziheng Zhou, Xiaopeng Hong, Guoying Zhao, and Matti Pietikäinen. A compact representation of visual speech data using latent variables. *IEEE Trans. Pattern Anal. Mach. Intell.*, 36(1):1–1, 2013. 6

[43] Carlos F Benitez-Quiroz, Ramprakash Srinivasan, and Aleix M Martinez. Facial color is an efficient mechanism

to visually transmit emotion. *Proceedings of the National Academy of Sciences*, 115(14):3581–3586, 2018. 7