

寻找用于跨库微表情识别的面部显著局部区域

江星洵 宗源 郑文明

东南大学

{jiangxingxun,xhzyuan,wenming_zheng}@seu.edu.cn

摘要

本文聚焦于跨库微表情研究：用于训练和测试的微表情样本来自不同数据集。训练集和测试集的微表情样本存在特征分布不匹配的问题，这使得将现有的微表情识别算法用于跨库微表情识别性能有所下降。为了更好地处理跨库微表情识别问题，我们提出了一种叫做迁移组稀疏回归方法(*TGSR*)的新的领域自适应方法。*TGSR*通过学习一个稀疏的共享回归矩阵，消除了训练集和测试集的特征分布差异，同时能够挑选出用于跨库微表情识别的面部局部区域。我们在*CASME II*数据库和*SMIC*数据库上对*TGSR*方法进行了测试。实验结果表明，我们所提的*TGSR*方法，其性能超过了目前大部分基于子空间学习的方法。

1. 引言

微表情是一种无意识下的面部表情，它能够反映被试者真实的情感[4]。自1969年Paul Ekman教授发现微表情以来，微表情在刑事调查[27]、临床诊断[7]、人机交互[29, 28]等诸多领域中，逐渐占据不可替代的地位。微表情是一种幅度小、持续时间短的微妙面部动作[5]，这使得识别微表情需要的专业门槛较高。人工微表情的识别，除了需要较高的专业门槛外，还需要心理学专业花费大量宝贵的时间，这使得微表情识别无法在现实生活中得到广泛而有效的运用。

为了解决这个问题，过去十年中，研究者们

在研究和设计微表情识别系统上进行了诸多尝试[21, 24, 23, 16, 12, 18, 25, 9, 36]。一些研究者建立了如*CASME II*[26]和*SMIC*[14]的微表情数据库，为研究微表情识别算法提供测试平台和性能基线标准。一些研究者设计了如LBP-TOP[31]、LBP-SIP[24]、面部动态映射(FDM)[25]、光流方向模糊直方图(FHOFO)[9]等优秀的手工时空特征，用于抽取微表情特征。另一些研究者[24, 16]结合了微表情特点及人类认识的处理过程，进一步提升了微表情识别性能。另外，还有一些研究者们使用深度卷积神经网络(CNNs)[13]和长短时记忆网络(LSTM)[11]，探索基于深度学习的自动微表情识别系统。然而即使是在精心训练的前提下，在某数据来源的数据库上训练的模型，仍然无法在另一数据来源的数据集上取得良好的效果，也即模型无法在不同来源的数据(被试者、摄像头和场景)上获得好的、鲁棒的性能。

为了增强微表情识别模型的鲁棒性，研究者们将领域自适应话题引入微表情识别，并提出跨数据库微表情识别问题：在一个微表情数据库（源域）上训练，在另外一个微表情数据库（目标域）上测试。跨数据库微表情识别是领域自适应问题的代表之一，能够较好地模仿现实生活中由于数据来源不同造成现有微表情模型不鲁棒的问题。事实上，由于跨数据库情感识别广泛价值的存在，研究者们自然图像[20]、语音[22]、面部表情[33, 32]、脑电[34, 15]等领域已进行了诸多研究。这些研究为跨库微表情识别的方法研究提供了基线性能结果。

2. 方法

2.1. 微表情特征的获取

微表情识别的第一步是抽取特征。如图1所示，我们采用多尺度网格空间划分[30]的方法，将人脸划分为 K 个块并抽取对应的手工特征。将特征抽取后得到的 K 个特征向量 \mathbf{x}_k 逐一连接起来，我们便得到描述微表情的层级特征 $\mathbf{x}^\nu = [\mathbf{x}_1^T, \dots, \mathbf{x}_K^T]^T \in \mathbb{R}^{Kd}$ 。假设有 N_s 个源域的微表情样本和 N_t 个目标域的微表情样本。故源域和目标域微表情样本的第 i 个人脸局部区域特征，可以分别用 $\mathbf{X}_i^s \in \mathbb{R}^{d \times N_s}$ 和 $\mathbf{X}_i^t \in \mathbb{R}^{d \times N_t}$ 表示；源域的特征矩阵可以用 $\mathbf{X}^s = [\mathbf{X}_1^{sT}, \dots, \mathbf{X}_K^{sT}]^T \in \mathbb{R}^{Kd \times N_s}$ 表示，目标域的特征矩阵可以用 $\mathbf{X}^t = [\mathbf{X}_1^{tT}, \dots, \mathbf{X}_K^{tT}]^T \in \mathbb{R}^{Kd \times N_t}$ 表示，矩阵的每列是类似 \mathbf{x}^ν 的向量。这里， d 是特征维度。本文用 $\mathbf{L}_s = [\mathbf{l}_1, \dots, \mathbf{l}_{N_s}] \in \mathbb{R}^{C \times N_s}$ 表示源域的标签。这里， \mathbf{L}_s 的第 i 列 $\mathbf{l}_i = [l_{i,1}, \dots, l_{i,C}]^T$ 代表源域第 i 个微表情样本。当第 i 个微表情样本表现的是第 j 种微表情时， \mathbf{l}_i 的第 j 个元素为1，其他元素为0。

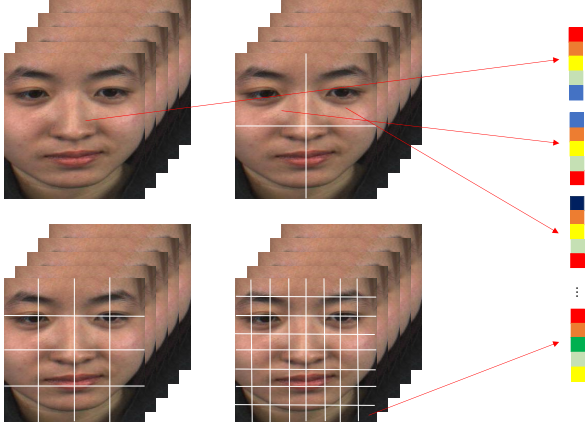


图 1. 实验中用于抽取多尺度特征的多尺度网格空间划分方法。

2.2. 建立TGSR模型

为了让学到的模型在不同数据来源的数据集上鲁棒，TGSR通过训练得到源域和目标域共享的标签-特征回归矩阵 $\mathbf{C} = [\mathbf{C}_1^T, \dots, \mathbf{C}_K^T] \in \mathbb{R}^{Kd \times C}$ ，

消除源域和目标域微表情特征的差异，这里 $\mathbf{C}_i^T \in \mathbb{R}^{d \times c}$ ，是第 i 个面部局部区域对应特征 \mathbf{X}_i 的回归矩阵。首先，我们将TGSR设计为如式1所示的正则化线性回归问题。

$$\min_{\mathbf{C}_i} \left\| \mathbf{L}^s - \sum_{i=1}^K \mathbf{C}_i^T \mathbf{X}_i^s \right\|_F^2 + \xi f_1(\mathbf{C}_i) \quad (1)$$

正则项 $f_1(\mathbf{C}_i)$ 表示源域特征投影和目标域特征投影的距离。在回归矩阵的求解过程中最小化正则项 $f_1(\mathbf{C}_i)$ ，可以减小源域和目标域特征分布间的差异。本文选用最大平均差异(MMD)作为 $f_1(\mathbf{C}_i)$ ，如式2所示。

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

这里， $\Phi(\cdot)$ 是用来将样本从原始特征空间映射到无穷维空间的核映射算子。然而，加入核映射算子的式子是较难优化的，所以本文将正则项 $f_1(\mathbf{C}_i)$ 修改为式3。

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

本文用 $s_i \in \{0, 1\}$ ($i = 1, \dots, K$)表示 K 个局部区域的重要性程度，并用 $s_i \mathbf{X}_i$ 代替 \mathbf{X}_i 。 $s_i = 1$ 意味着第 i 个面部局部区域和对应特征矩阵 \mathbf{X}_i 对于微表情识别而言，是显著的区域；而 $s_i = 0$ 意味着其不重要，是可以删去的。可以注意到， $\sum_{i=1}^K s_i$ 的值等于面部显著区域的块数。所以挑选显著区域的问题等价于 $\sum_{i=1}^K s_i$ 的最小化问题。结合式3，我们可以将式1重写为式4。

$$\begin{aligned} \min_{\mathbf{C}_i} & \left\| \mathbf{L}^s - \sum_{i=1}^K s_i \mathbf{C}_i^T \mathbf{X}_i^s \right\|_F^2 + \lambda \sum_{i=1}^K s_i \\ & + \xi \left\| \frac{1}{N_s} \sum_{i=1}^K s_i \mathbf{C}_i^T \mathbf{X}_i^s \mathbf{1}_s - \frac{1}{N_t} \sum_{i=1}^K s_i \mathbf{C}_i^T \mathbf{X}_i^t \mathbf{1}_t \right\|_F^2 \quad (4) \\ \text{s.t. } & s_i \in \{0, 1\} \end{aligned}$$

由于 $s_i = 0$ 使得 $\|s_i \mathbf{C}_i\|_F = 0$ ， $\|s_i \mathbf{C}_i\|_F$ 等价于 $s_i = 0$ 或 $\mathbf{C}_i = \mathbf{0}$ 。 \mathbf{C}_i 是第 i 个面部局部区域对应特征 \mathbf{X}_i 的回归矩阵，当投影矩阵 $\mathbf{C} = \mathbf{0}$ 时，由于其

数学含义等价于 $s_i = 0$ ，所以，我们可以将 s_i 和 \mathbf{C}_i 相结合，把式4重写为式5。

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

2.3. TGSr模型的优化求解

令 $\tilde{\mathbf{X}}_i = \left[\mathbf{X}_i^s, \sqrt{\xi} \left(\frac{1}{N_s} \mathbf{X}_i^s \mathbf{1}_s - \frac{1}{N_t} \mathbf{X}_i^t \mathbf{1}_t \right) \right]$, $\tilde{\mathbf{L}} = [\mathbf{L}^s, \mathbf{0}]$, $\|\mathbf{C}\|_F = \sum_{i=1}^K \|\mathbf{C}_i\|_F$ 。从而，可以将关于矩阵 \mathbf{C} 的优化问题式5整理为式6。

$$\min_{\mathbf{C}} \left\| \tilde{\mathbf{L}} - \mathbf{C}^T \tilde{\mathbf{X}} \right\|_F^2 + \lambda \|\mathbf{C}\|_F \quad (6)$$

并进一步将式6转化为式7。

$$\min_{\mathbf{C}, \mathbf{D}} \left\| \tilde{\mathbf{L}} - \mathbf{D}^T \tilde{\mathbf{X}} \right\|_F^2 + \lambda \|\mathbf{C}\|_F \quad (7)$$

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

我们可以求得式7对应的拉格朗日增广函数如式8

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

式8可以用交错方向法(ADM)和非精准拉格朗日乘子法(IALM)求解。我们将完整的求解算法总结如算法1所示。

2.4. 将TGSr模型用于跨库微表情识别

当TGSr模型在标记过的源数据库微表情样本和未标记的微表情样本上训练，并得到模型的最佳参数后，我们便可使用TGSr模型预测测试集微表情样本的类别。当给定测试集微表情样本特征 \mathbf{x}_i^{te} 时，其微表情类别预测问题便可转化为如式10所示的优化问题，用其估计对应的微表情标签 \mathbf{l}^{te} 。

$$\min_{\mathbf{l}^{te}} \left\| \mathbf{l}^{te} - \sum_{i=1}^K \hat{\mathbf{C}}_i^T \mathbf{x}_i^{te} \right\|_F^2 \quad (10)$$

$s.t. \mathbf{l}^{te} \geq 0, \mathbf{1}^T \mathbf{l}^{te} = 1$

这里， $\hat{\mathbf{C}}_i$ 是经过训练学习到的微表情标签关于到的第 i 个面部局部区域特征的投影矩阵。TGSr预

Algorithm 1 TGSr模型中投影矩阵 \mathbf{C} 最优参数的求解算法。

输入: 标签矩阵 \mathbf{L} 、特征矩阵 \mathbf{X} 、面部显著局部区域数 $\kappa < K$ 、权衡系数 ρ 和 μ_{max} 。

- 初始化投影矩阵 \mathbf{C}
- 初始化拉格朗日乘子矩阵 \mathbf{P} 和权衡系数 μ 。

重复步骤1)到4)至收敛

- 1: 保持矩阵 \mathbf{C} 、矩阵 \mathbf{P} 和权衡系数 μ 不变，更新投影矩阵 \mathbf{D} :

$$\mathbf{D} = (\mu \mathbf{I} + 2\mathbf{X}\mathbf{X}^T)^{-1} (2\mathbf{X}\mathbf{L}^T + \mathbf{P} + \mu \mathbf{C})$$

- 2: 保持矩阵 \mathbf{P} 、矩阵 \mathbf{D} 和权衡系数 μ 不变，更新投影矩阵 \mathbf{C}

计算 $d_i = \left\| \mathbf{D}_i - \frac{\mathbf{P}_i}{\mu} \right\|_F$ ，并对 d_i 的值进行排序。假设排序后顺序为 $d_{i_1} \geq d_{i_2} \geq \dots \geq d_{i_K}$ ，令 $\lambda = \mu d_{i_{\kappa+1}}$ ，然后更新投影矩阵 \mathbf{C} 。

$$\mathbf{C}_i = \begin{cases} \frac{d_i - \frac{\lambda}{\mu}}{d_i} (\mathbf{D}_i - \frac{\mathbf{P}_i}{\mu}), & \frac{\lambda}{\mu} < d_i \\ 0, & \frac{\lambda}{\mu} \geq d_i \end{cases} \quad (9)$$

- 3: 更新矩阵 \mathbf{P} 及其权衡参数 μ

$$\mathbf{P} = \mathbf{P} + \mu (\mathbf{Z} - \mathbf{C})$$

$$\mu = \min(\rho\mu, \mu_{max})$$

- 4: 检验收敛:

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

输出: 投影矩阵 \mathbf{C} 和权衡系数 λ

测微表情类别为标签向量 \mathbf{l}^{te} 最大值对应的微表情类别，如式11。

$$\hat{c} = \arg \max_j \{\mathbf{l}_j^{te}\} \quad (11)$$

3. 实验

3.1. 实验设置

数据库。 我们在经过挑选的CASME II数据库和SMIC数据上评价我们的方法。SMIC[14]收集了16个被试的微表情图片序列，包含三个子集(HS、VIS和NIR)。HS子集由帧率100帧/秒的高速摄像机

表 1. 经过挑选的CASME II数据库与SMIC数据库(包含HS、VIS和NIR三个子集)的样本信息。

数据库	正性	负性	惊奇
经过挑选的CASME II	32	73	25
SMIC(HS)	51	70	43
SMIC(VIS)	23	28	20
SMIC(NIR)	23	28	20

获得, VIS子集由帧率25帧/秒的普通摄像头获得, NIR子集由近红外摄像机获得。HS子集、VIS子集和NIR子集, 分别拥有164个样本、71个样本和71个样本。SMIC数据库包含了三种微表情:正性、负性和惊奇。CASME II[26]收集了来自26个被试的247个微表情样本, 其包含了5种微表情:开心、惊奇、厌恶、沮丧和其他。经过挑选的CASME II数据集是排除了CASME II数据库中的“其他”类别, 并将开心重标为正性, 厌恶和沮丧重标为负性后得到的数据库。表1总结了CASME II数据库和SMIC数据库的基本信息。

协议。 我们遵循[30]的协议, 设置了两类跨库微表情识别实验。第一类是挑选SMIC数据库三个子集HS、VIS和NIR中的两个, 如HS和VIS, 将其作为源数据库和目标数据库, 进行跨库微表情识别。另一类是选择CASME II数据库, 并挑选SMIC数据库三个子集HS、VIS和NIR中的一个, 如CASME II和HS, 作为源数据库和目标数据库, 进行跨库微表情识别。用 H 、 V 、 N 和 C 代表SMIC数据库的HS子集、SMIC数据库的VIS子集、SMIC数据库的NIR子集和CASME II数据库, 我们便可以表示从源数据库 S 到目标数据库 T 的跨库微表情实验任务 $S \rightarrow T$ 。比如, 从SMIC数据库的HS子集到CASME II数据库的跨库微表情识别任务可以表示为 $H \rightarrow C$

评价指标。 我们用Mean F1-Score和WAR来评估我们的方法。WAR即加权平均召回率, 也即模型识别的整体精度。Mean F1-Score可由 $MeanF1 - Score = \frac{1}{c} \sum_{i=1}^c \frac{2p_i r_i}{p_i + r_i}$ 计算而得, 这里, p_i 和 r_i 分别是第 i 类微表情的精度和召回率, c 是微表情的类别数。由于跨库微表情识别中存在严重的样本不平衡问题,

所以用Mean F1-Score作为评价指标是十分合适的。

实现细节。 为了方法的公平对比, 我们使用CASME II数据库和SMIC数据库收集者提供的图片序列。我们使用时间插值模型(TIM), 将所有图片序列归一化为16帧, 并将每张图片调整为 112×112 分辨率大小。我们使用了 $(1 \times 1, 2 \times 2, 4 \times 4, 8 \times 8)$ 四种空间尺度划分方式, 获得了一共 $K = 85$ 块多尺度人脸局部区域块。我们抽取每个人脸局部区域块的LBP-TOP($R = 3, P = 8$)[31]特征, 获取了微表情的层级式时空特征描述。我们用网格搜索的方式搜索参数空间, 以此获取我们模型的最佳超参。

3.2. 结果及分析

表2和表3分别展现了第一类实验和第二类实验的结果。我们计算了不同方法在两类实验上的平均表现性能(Mean F1-Score和WAR)。第一类实验中, SVM[2], IW-SVM[10], TCA[19], GFK[8], SA[6], STM[3], TKL[17], TSRG[35], DRLS[37]和我们的方法在Mean F1-Score/WAR性能指标上, 分别达到了0.6003/61.62, 0.6911/68.07, 0.6238/64.01, 0.7223/72.44, 0.6917/69.44, 0.6440/65.78, 0.6964/69.92, 0.6991/70.05, 0.6552/66.60和0.7141/71.80的效果。第二类实验中, VM[2], IW-SVM[10], TCA[19], GFK[8], SA[6], STM[3], TKL[17], TSRG[35], DRLS[37]和我们的方法在Mean F1-Score/WAR性能指标上, 分别达到了0.4112/45.55, 0.4876/50.73, 0.5177/53.45, 0.5161/54.31, 0.4877/50.90, 0.3982/45.61, 0.4925/51.93, 0.5347/56.22, 0.5102/53.71和0.5424/56.33的效果。从表2和表3的加粗部分中, 可以看到我们方法的实验性能超过了大部分目前的最佳方法。并且, 我们的方法在超过一半的实验中取得了目前最好的结果: 总共12个实验, 在7个实验上取得了最佳性能。

此外, 我们从表格中还能看到其他三个显著的特点。首先, 我们发现领域自适应方法确实可以提高模型性能, 并且领域自适应方法的表现好于非领域自适应方法。IW-SVM通过学习一组加权系数, 这使得它在各实验上的平均性能优

表 2. 挑选在SMIC数据库三个子集(HS、VIS、NIR)中任意两个作为源数据库和目标数据库，进行实验得到的结果。评价指标为Mean F1-Score/WAR。三类微表情分别是:正性、负性和惊奇。实验的最高结果用加粗的字体标出。

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)
SVM[2]	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[10]	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[19]	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[8]	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[6]	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[3]	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[17]	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[35]	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[37]	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	0.6234 / 61.59	0.6984 / 70.42	0.8403 / 84.51

表 3. 挑选CASME II数据库和SMIC数据库三个子集(HS、VIS和NIR)之一作为源数据库和目标数据库，进行实验得到的结果。评价指标为Mean F1-Score/WAR。三类微表情分别是:正性、负性和惊奇。实验的最高结果用加粗的字体标出。

Method	Exp.7(C→H)	Exp.8(H→C)	Exp.9(C→V)	Exp.10(V→C)	Exp.11(C→N)	Exp.12(N→C)
SVM[2]	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[10]	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[19]	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[8]	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[6]	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[3]	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[17]	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[35]	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[37]	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

于SVM: 在第一类实验上提升了0.0911/6.45(Mean F1-score/WAR)，在第二类实验上提升了0.0764/5.18 (Mean F1-score/WAR)。对比方法中，除SVM外，其他方法都使用了领域自适应技巧。我们可以发现，它们在平均性能上都稳定地、大幅地超过SVM。

其次，从表2和表3中，我们发现第一类实验的结果，普遍优于第二类实验结果。第一类实验选择了SMIC数据库中两个不同模态的子集，第二类实验选择了两个不同的数据库: CASME II数据库和SMIC数据库的任一子集。第一类实验的数据来源差距较小，而第二类实验的数

据来源差距较大。这使得第一类实验的性能优于第二类实验的性能。第三，我们发现源域和目标域数据库互换的任务，识别性能相差较大。以SVM的性能为代表，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)，Exp.11(C→N)的结果优于Exp.12(N→C)。Exp.1(H→V)使用高速相机获取图片序列作为源域数据，普通相机获取的图片序列作为目标域数据；而Exp.2(V→H)的源域数据和目标域数据与Exp.1(H→V)相反。我们认为，高速相机能够更好地获取面部精细的时空运动，这

有助于微表情的识别。Exp.3(N→H)使用高速相机获取的图片序列作为源域数据，近红外摄像头捕捉的图片序列作为目标域数据；Exp.4(N→H)的源域数据和目标域数据与Exp.3(N→H)相反。虽然，用高速摄像头获取的HS子集，与用近红外摄像头获取的NIR子集都来源于SMIC数据库，但是近红外摄像头获取的NIR子集较之高速摄像头捕捉的HS子集，是丢失了颜色信息的。我们发现从带有颜色信息数据学习微表情识别模型并迁移到不带有颜色信息数据的效果，是优于从不带有颜色信息的数据迁移到带有颜色信息的数据的。这印证了人类对情感的理解是与颜色相关的[1]。同样的现象也出现在Exp.11(C→N)和Exp.12(N→C)中。Exp.11(C→N)和Exp.12(N→C)，涉及了两个不同的数据库：CASME II和SMIC。我们发现Exp.11(C→N)和Exp.12(N→C)的性能差异大于Exp.3(N→H)和Exp.4(N→H)的实验差异。这告诉我们，对跨库微表情识别而言，数据库的差异是大于颜色差异的。

4. Conclusion

本文中，我们提出了一种用于跨库微表情识别的迁移组稀疏回归模型TGSR。TGSR模型能够挑选出用于跨库微表情识别的显著人脸局部区域、淘汰掉冗余信息，所以TGSR可以学习到从特征映射到标签的稀疏投影矩阵。这个矩阵有利于模型识别性能的提升和计算加速。并且，TGSR可以有效消除源域微表情数据库和目标域微表情数据库的特征差异。实验结果表明，我们的方法能超过目前大部分跨库微表情识别的最佳方法。

参考文献

- [1] C. F. Benitez-Quiroz, R. Srinivasan, and A. M. Martinez. Facial color is an efficient mechanism to visually transmit emotion. *Proceedings of the National Academy of Sciences*, 115(14):3581–3586, 2018. 6
- [2] C.-C. Chang and C.-J. Lin. Libsvm: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, 2(3):1–27, 2011. 4, 5
- [3] W.-S. Chu, F. De la Torre, and J. F. Cohn. Selective transfer machine for personalized facial action unit detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3515–3522, 2013. 4, 5
- [4] P. Ekman. *Telling lies: Clues to deceit in the marketplace, politics, and marriage (revised edition)*. WW Norton & Company, 2009. 1
- [5] P. Ekman and W. V. Friesen. Nonverbal leakage and clues to deception. *Psychiatry*, 32(1):88–106, 1969. 1
- [6] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars. Unsupervised visual domain adaptation using subspace alignment. In *Proceedings of the IEEE international conference on computer vision*, pages 2960–2967, 2013. 4, 5
- [7] M. Frank, M. Herbasz, K. Sinuk, A. Keller, and C. 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
- [8] B. Gong, Y. Shi, F. Sha, and K. Grauman. Geodesic flow kernel for unsupervised domain adaptation. In *2012 IEEE conference on computer vision and pattern recognition*, pages 2066–2073. IEEE, 2012. 4, 5
- [9] S. Happy and A. Routray. Fuzzy histogram of optical flow orientations for micro-expression recognition. *IEEE Transactions on Affective Computing*, 10(3):394–406, 2017. 1
- [10] A. Hassan, R. Damper, and M. Niranjan. On acoustic emotion recognition: compensating for covariate shift. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(7):1458–1468, 2013. 4, 5
- [11] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997. 1
- [12] D. H. Kim, W. J. Baddar, and Y. M. Ro. Micro-expression recognition with expression-state constrained spatio-temporal feature representations. In *Proceedings of the 24th ACM international conference on Multimedia*, pages 382–386, 2016. 1

- [13] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012. [1](#)
- [14] X. Li, T. Pfister, X. Huang, G. Zhao, and M. Pietikäinen. A spontaneous micro-expression database: Inducement, collection and baseline. In *2013 10th IEEE International Conference and Workshops on Automatic face and gesture recognition (fg)*, pages 1–6. IEEE, 2013. [1](#), [3](#)
- [15] Y. Li, W. Zheng, Y. Zong, Z. Cui, T. Zhang, and X. Zhou. A bi-hemisphere domain adversarial neural network model for eeg emotion recognition. *IEEE Transactions on Affective Computing*, 2018. [1](#)
- [16] Y.-J. Liu, J.-K. Zhang, W.-J. Yan, S.-J. Wang, G. Zhao, and X. Fu. A main directional mean optical flow feature for spontaneous micro-expression recognition. *IEEE Transactions on Affective Computing*, 7(4):299–310, 2015. [1](#)
- [17] M. Long, J. Wang, J. Sun, and S. Y. Philip. Domain invariant transfer kernel learning. *IEEE Transactions on Knowledge and Data Engineering*, 27(6):1519–1532, 2014. [4](#), [5](#)
- [18] P. Lu, W. Zheng, Z. Wang, Q. Li, Y. Zong, M. Xin, and L. 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](#)
- [19] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang. Domain adaptation via transfer component analysis. *IEEE Transactions on Neural Networks*, 22(2):199–210, 2010. [4](#), [5](#)
- [20] S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009. [1](#)
- [21] T. Pfister, X. Li, G. Zhao, and M. Pietikäinen. Recognising spontaneous facial micro-expressions. In *2011 international conference on computer vision*, pages 1449–1456. IEEE, 2011. [1](#)
- [22] B. Schuller, B. Vlasenko, F. Eyben, M. Wöllmer, A. Stuhlsatz, A. Wendemuth, and G. Rigoll. Cross-corpus acoustic emotion recognition: Variances and strategies. *IEEE Transactions on Affective Computing*, 1(2):119–131, 2010. [1](#)
- [23] S.-J. Wang, W.-J. Yan, X. Li, G. Zhao, C.-G. Zhou, X. Fu, M. Yang, and J. Tao. Micro-expression recognition using color spaces. *IEEE Transactions on Image Processing*, 24(12):6034–6047, 2015. [1](#)
- [24] Y. Wang, J. See, R. C.-W. Phan, and Y.-H. Oh. Lbp with six intersection points: Reducing redundant information in lbp-top for micro-expression recognition. In *Asian conference on computer vision*, pages 525–537. Springer, 2014. [1](#)
- [25] F. Xu, J. Zhang, and J. Z. Wang. Microexpression identification and categorization using a facial dynamics map. *IEEE Transactions on Affective Computing*, 8(2):254–267, 2017. [1](#)
- [26] W.-J. Yan, X. Li, S.-J. Wang, G. Zhao, Y.-J. Liu, Y.-H. Chen, and X. Fu. Casme ii: An improved spontaneous micro-expression database and the baseline evaluation. *PloS one*, 9(1):e86041, 2014. [1](#), [4](#)
- [27] W.-J. Yan, Q. Wu, J. Liang, Y.-H. Chen, and X. Fu. How fast are the leaked facial expressions: The duration of micro-expressions. *Journal of Nonverbal Behavior*, 37(4):217–230, 2013. [1](#)
- [28] C. Yang, J. Luo, Y. Pan, Z. Liu, and C.-Y. Su. Personalized variable gain control with tremor attenuation for robot teleoperation. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 48(10):1759–1770, 2017. [1](#)
- [29] C. Yang, H. Wu, Z. Li, W. He, N. Wang, and C.-Y. Su. Mind control of a robotic arm with visual fusion technology. *IEEE Transactions on Industrial Informatics*, 14(9):3822–3830, 2017. [1](#)
- [30] T. Zhang, Y. Zong, W. Zheng, C. P. Chen, X. Hong, C. Tang, Z. Cui, and G. Zhao. Cross-database micro-expression recognition: A benchmark. *IEEE Transactions on Knowledge and Data Engineering*, 2020. [2](#), [4](#)
- [31] G. Zhao and M. Pietikainen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE transactions on pattern analysis and machine intelligence*, 29(6):915–928, 2007. [1](#), [4](#)
- [32] S. Zhao, C. Lin, P. Xu, S. Zhao, Y. Guo, R. Krishna, G. Ding, and K. Keutzer. Cycleemotiongan: Emotional semantic consistency preserved cyclegan for adapting image emotions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 2620–2627, 2019. [1](#)
- [33] S. Zhao, X. Zhao, G. Ding, and K. Keutzer. Emotiongan: Unsupervised domain adaptation for learning discrete probability distributions of image emotions. In *Proceedings of the 26th ACM international conference on Multimedia*, pages 1319–1327, 2018. [1](#)
- [34] W.-L. Zheng and B.-L. Lu. Personalizing eeg-based affective models with transfer learning. In *Proceedings of the twenty-*

fifth international joint conference on artificial intelligence, pages 2732–2738, 2016. 1

- [35] Y. Zong, X. Huang, W. Zheng, Z. Cui, and G. Zhao. Learning a target sample re-generator for cross-database micro-expression recognition. In *Proceedings of the 25th ACM international conference on Multimedia*, pages 872–880, 2017. 4, 5
- [36] Y. Zong, X. Huang, W. Zheng, Z. Cui, and G. Zhao. Learning from hierarchical spatiotemporal descriptors for micro-expression recognition. *IEEE Transactions on Multimedia*, 20(11):3160–3172, 2018. 1
- [37] Y. Zong, W. Zheng, X. Huang, J. Shi, Z. Cui, and G. Zhao. Domain regeneration for cross-database micro-expression recognition. *IEEE Transactions on Image Processing*, 27(5):2484–2498, 2018. 4, 5