Seeking Salient Facial Regions for Cross-Database Micro-Expression Recognition

Xingxun Jiang, Yuan Zong#, Wenming Zheng#, Jiateng Liu, Mengting Wei
Key Laboratory of Child Development and Learning Science of Ministry of Education,
School of Biological Science and Medical Engineering, Southeast University, Nanjing 210096, China
{jiangxingxun, xhzongyuan, wenming_zheng, jiateng_liu, weimengting}@seu.edu.cn

***Abstract*—Cross-Database Micro-Expression Recognition (CDMER) aims to develop the Micro-Expression Recognition (MER) methods that satisfy different conditions (equipment, subjects, and scenes) in practical application, i.e., the MER method of strong domain adaption ability. CDMER faces two obstacles: 1) the severe feature distribution gap between the training and test databases and 2) the feature representation bottleneck for micro-expression (ME) such local and subtle facial expressions. To solve these obstacles, this paper proposes a novel Transfer Group Sparse Regression method, namely TGSR, which seeks and selects those salient facial regions to 1) promote a more precise measurement of the difference between source and target databases by the operation in the feature level to alleviate their difference better, and to 2) improve the extracted hand-crafted feature to be more effective and explicable for better MER. We use two public micro-expression databases, i.e., CASME II and SMIC, to evaluate the proposed TGSR. Experimental results show that TGSR learns the discriminative feature and outperforms most state-of-the-art subspace-learning-based domain adaption methods for CDMER.**

I. Introduction

Micro-Expression (ME) is a low amplitude and short duration facial expression which may reflect subjects' genuine emotions[1], [2].It plays an indispensable role in criminal investigations[3], clinical diagnosis[4], and human-computer interaction[5], [6], [7], [8].

Due to its enormous potential value, many efforts have been made to design an automatic Micro-Expression Recognition (MER) system in the last few decades to give ME full play[9], [10], [11], [12], [13], [14], [15], [16], [17]. In the age of traditional pattern recognition, feature extraction was the most important step for automatic MER. Thus researchers focused on this and proposed many well-designed hand-crafted feature, such as, LBP-TOP[18], LBP-SIP[10], FDM[15], and FHOFO[19], [20], [21], [22], [23], to better collaborate with classifiers. In the age of deep learning, researchers integrated the feature extraction and classification process into an end-to-end approach, and proposed a series of creative approaches for MER. Though these methods brought promising results, they are only evaluated in one single database. This may not satisfy the practical MER required to be used in different domains, i.e., recognizing ME captured by various equipment from various subjects and scenes.

To learn a domain-robust model, researchers turned their attention to domain adaptive MER methods. A new challenging topic has thus emerged, i.e., Cross-Database Micro-Expression Recognition (CDMER), which mimics the domain variation problem in practical application. The CDMER task belongs to a typical cross-database emotion recognition task, which evaluates the model's domain adaptive ability operated by training the model in one micro-expression database (source database) and testing in the other (target database). In fact, cross-database emotion recognition tasks have been extensively studied in natural image[24], speech[25], [26], facial expression[27], [28], and EEG[29], [30] modality. They reveal the potential problems and provide baseline methods for emerging CDMER problem. Like them, CDMER[31], [32], [33] faces two obstacles: 1) the severe feature distribution gap between the training and test databases, and 2) the feature representation bottleneck of ME such local and subtle facial expression, which requires professional knowledge.

It has been widely validated that salient region selection benefits face-related tasks. Inspired by this, this paper adopts facial region selection to seek the salient regions from the whole face to 1) promote a more precise measurement of the difference between the source and target databases by the operations in the feature level to alleviate this difference better, and to 2) improve the extracted hand-crafted feature to become more effective and explicable for better MER. We propose a novel Transfer Group Sparse Regression method (TGSR) which introduces a learnable binary sparse regression matrix shared between the source and target databases to implement the facial region selection for MER. Specially, TGSR contains three terms: a regression term to bridge the micro-expression data and labels, a Frobenius norm sparse term to make the learnable regression matrix sparse and to supervise the salient region selection, and a joint feature distribution adaption term including the learnable regression matrix to bridge the source and target features in the label space. TGSR learns the discriminative features by combining the optimization of these three terms with the learnable sparse regression matrix. We evaluate our method on CASME II[34] database and SMIC database[35]. Experimental results and corresponding visualization demonstrate that our proposed TGSR can effectively solve the two obstacles above and outperform most state-of-the-art subspace-learning-based domain adaption methods for CDMER.

# indicates the corresponding authors

II. Method

A. The Generation of Micro-Expression Features

Extracting ME feature is the first step of MER. As shown in Fig. 1, we firstly used a grid-based multi-scale spatial division scheme[36] to divide the cropped ME sequence into K spatial local sequence. Then we extracted d-dimensional hand-crafted spatio-temporal feature $\boldsymbol{x}_k$ of each spatial local sequence and obtain the hierarchical feature $\boldsymbol{x}^{\nu} = [\boldsymbol{x}_1^{\mathrm{T}}, \cdots, \boldsymbol{x}_K^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{Kd}$ of the ME sequence by concatenating these spatial local features one by one. Suppose that we have N_s source and N_t target micro-expression samples, the feature matrix of source and target database can be denoted as $\boldsymbol{X}^s = \left[{\boldsymbol{X}_1^s}^{\mathrm{T}}, \cdots, {\boldsymbol{X}_K^s}^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{R}^{Kd \times N_s}$ and $\boldsymbol{X}^t = \left[{\boldsymbol{X}_1^t}^{\mathrm{T}}, \cdots, {\boldsymbol{X}_K^t}^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{R}^{Kd \times N_t}$. Here, each column of $\boldsymbol{X}^s$ and $\boldsymbol{X}^t$ is a feature vector like $\boldsymbol{x}^{\nu}$, they respectively denote the micro-expression feature from the source and target database. $\boldsymbol{X}_i^s \in \mathbb{R}^{d \times N_s}$ and $\boldsymbol{X}_i^t \in \mathbb{R}^{d \times N_t}$ respectively denote the feature matrix of i-th spatial local sequence from the source and target databases. The label of source micro-expression database is denoted as $\boldsymbol{L}^s = [\boldsymbol{l}_1^s, \cdots, \boldsymbol{l}_{N_s}^s] \in \mathbb{R}^{C \times N_s}$, where C is the total class number and the j-th column of $\boldsymbol{L}^s$ denotes the label vector of the j-th source micro-expression sample. The label vector of j-th sample $\boldsymbol{l}_j^s = [l_{j,1}^s, \cdots, l_{j,C}^s]^{\mathrm{T}}$ is a one-hot vector in which only one element $l_{j,c}^s$ equals one and the other are zero. It indicates j-th sample from the source database belongs to c-th micro-expression category.

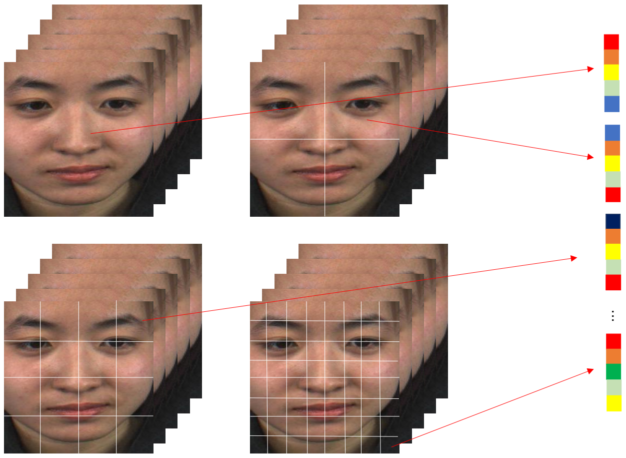

Fig. 1. The grid-based multi-scale spatial division scheme for extracting micro-expression feature.

B. Proposed Method

As mentioned above, the basic idea of our TGSR is building a linear regression model with a learnable shared regression matrix $\boldsymbol{C} = [\boldsymbol{C}_1^{\mathrm{T}}, ..., \boldsymbol{C}_K^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{Kd \times C}$ to realize the mapping from features to labels and the selection of those salient facial regions for better dealing with the CDMER problem, which can be formulated as Equ. (1),

$$\min_{\boldsymbol{C}_i} \left\| \boldsymbol{L}^s - \sum_{i=1}^{K} \boldsymbol{C}_i^{\mathrm{T}} \boldsymbol{X}_i^s \right\|_F^2 + \xi f_1(\boldsymbol{C}_i) + \lambda f_2(\boldsymbol{C}_i), \quad (1)$$

where $\boldsymbol{C}_i \in \mathbb{R}^{C \times d}$ is such a domain-invariant regression matrix to bridge the features of the i-th facial region from the source database and corresponding label information, $f_1(\boldsymbol{C}_i)$ and $f_2(\boldsymbol{C}_i)$ are the well-designed regularizations, and ξ and λ are corresponding hyper-parameters control the balance of the regression term and two regularization terms.

Regularization $f_1(\boldsymbol{C}_i)$ is called the database difference elimination one, which measures the difference between the source and target micro-expression features. Thus minimizing $f_1(\boldsymbol{C}_i)$ together with regression optimization will alleviate the significant feature distribution difference. We use the maximum mean discrepancy (MMD) to serve as this regularization term, which can be expressed as Equ. (2),

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

where $\Phi(\cdot)$ is a kernel mapping operator projecting features from the original space to an infinite one. We use $\mathbf{1}_s \in \mathbb{R}^{N_s}$ and $\mathbf{1}_t \in \mathbb{R}^{N_t}$ to respectively transform the source and target feature into a scalar value. Here $\mathbf{1}_s$ and $\mathbf{1}_t$ are the vectors filled with scalar value 1. However, the kernel mapping operator is unsolvable, so we further modify the MMD into Equ. (3) to serve as $f_1(\boldsymbol{C}_i)$,

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

By relaxing the kernel-mapped feature difference between the source and target databases into the difference in the label space, Equ. (2) becomes solvable. Meanwhile, by optimizing the feature representation using facial region selection, Equ. (3) becomes a more precise measurement of the difference between the source and target databases, which aims to alleviate this difference better. Regularization $f_2(\boldsymbol{C}_i)$ is a sparse term designed to select the salient facial regions to promote the extracted feature more effective and explicable for MER, which is defined as Equ. (4),

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

Thus, substituting Equ. (3) and Equ. (4) into Equ. (1), we can rewrite the objective function into Equ. (5),

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

C. Optimization

We can use the Alternative Direction Method (ADM)[37] and Inexact Augmented Lagrangian Multiplier (IALM)[38] to solve Equ. (5). We first rewrote Equ. (5) into Equ. (6),

$$\min_{\boldsymbol{C}_i} \left\| \tilde{\boldsymbol{L}} - \sum_{i=1}^{K} \boldsymbol{C}_i^{\mathrm{T}} \tilde{\boldsymbol{X}}_i \right\|_F^2 + \lambda \sum_{i=1}^{K} \|\boldsymbol{C}_i\|_F \,, \tag{6}$$

where $\tilde{\boldsymbol{L}} = [\boldsymbol{L}^s, \mathbf{0}]$, $\mathbf{0} \in \mathbb{R}^{c\times 1}$, $\tilde{\boldsymbol{X}}_i = \left[\boldsymbol{X}_i^s, \sqrt{\xi}(\frac{1}{N_s}\boldsymbol{X}_i^s\mathbf{1}_s - \frac{1}{N_t}\boldsymbol{X}_i^t\mathbf{1}_t)\right]$. Then we introduce a new variable, $\boldsymbol{D} = [\boldsymbol{D}_1^{\mathrm{T}}, \cdots, \boldsymbol{D}_K^{\mathrm{T}}]^{\mathrm{T}}$, which equals $\boldsymbol{C} = [\boldsymbol{C}_i^{\mathrm{T}}, \cdots, \boldsymbol{C}_K^{\mathrm{T}}]^{\mathrm{T}}$, thus the optimization of Equ. (6) can be converted into a constrained one as Equ. (7),

$$\min_{\boldsymbol{C},\boldsymbol{D}} \left\| \tilde{\boldsymbol{L}} - \sum_{i=1}^{K} \boldsymbol{D}_i^{\mathrm{T}} \tilde{\boldsymbol{X}}_i \right\|_F^2 + \lambda \sum_{i=1}^{K} \|\boldsymbol{C}_i\|_F \,, \quad \text{s.t.}\ \boldsymbol{D}_i = \boldsymbol{C}_i. \tag{7}$$

Subsequently, we can obtain the corresponding augmented Lagrange function as Equ. (8) shown,

$$\begin{aligned} \Gamma(\boldsymbol{C}_i, \boldsymbol{D}_i, \boldsymbol{P}_i, \mu) = & \left\| \tilde{\boldsymbol{L}} - \sum_{i=1}^{K} \boldsymbol{D}_i^{\mathrm{T}} \tilde{\boldsymbol{X}}_i \right\|_F^2 + \lambda \sum_{i=1}^{K} \|\boldsymbol{C}_i\|_F \\ & + \sum_{i=1}^{K} \operatorname{tr}\left[\boldsymbol{P}_i^{\mathrm{T}}(\boldsymbol{C}_i - \boldsymbol{D}_i)\right] + \frac{\mu}{2}\sum_{i=1}^{K} \|\boldsymbol{C}_i - \boldsymbol{D}_i\|_F^2 \,, \end{aligned} \tag{8}$$

where $\boldsymbol{P}_i \in \mathbb{R}^{d\times C}$ denotes the Lagrangian multiplier matrix corresponding to the i-th facial spatial local block sequence, and μ is the given trade-off coefficient.

To learn the optimal $\boldsymbol{C}_i$, we only need to minimize the Lagrange function of Equ. (8) while iteratively update $\boldsymbol{C}_i$ and $\boldsymbol{D}_i$. Specifically, we repeat the following four steps until convergence:

1) Fix $\boldsymbol{C}$, $\boldsymbol{P}$, μ and update $\boldsymbol{D}$:

In this step, the optimization problem with respect to the sub-matrix $\boldsymbol{D}_i$ of $\boldsymbol{D}$ can be written as Equ. (9), whose closed-form solution as Equ. (12) shown,

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

where $\boldsymbol{P}^{\mathrm{T}} = [\boldsymbol{P}_1^{\mathrm{T}}, \cdots, \boldsymbol{P}_K^{\mathrm{T}}]$, $\boldsymbol{P} \in \mathbb{R}^{Kd\times C}$, $\boldsymbol{P}_j \in \mathbb{R}^{d\times C}$.

2) Fix $\boldsymbol{D}$, $\boldsymbol{P}$, μ update $\boldsymbol{C}$:

In this step, the optimization problem with respect to the sub-matrix $\boldsymbol{C}_i$ of $\boldsymbol{C}$ can be written as Equ. (10),

$$\begin{aligned} \min_{\boldsymbol{C}_i} \lambda \sum_{i=1}^{K} \|\boldsymbol{C}_i\|_F + \sum_{i=1}^{K} \operatorname{tr}\left[\boldsymbol{P}_i^{\mathrm{T}}(\boldsymbol{C}_i - \boldsymbol{D}_i)\right] \\ + \frac{\mu}{2}\sum_{i=1}^{K} \|\boldsymbol{C}_i - \boldsymbol{D}_i\|_F^2 \,. \end{aligned} \tag{10}$$

We can convert Equ. (10) into Equ. (11), and obtain the optimal $\boldsymbol{C}$ using Equ. (13).

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

3) Update $\boldsymbol{P}$ and μ.

4) Check the convergence of $\|\boldsymbol{C} - \boldsymbol{D}\|_\infty < \varepsilon$.

Algorithm 1 shows more optimization details.

Algorithm 1 The Algorithm to solve the optimal regression matrix $\boldsymbol{C}$ in TGSR method.

```
Input: Data matrix $\tilde{L}$ and $\tilde{X} = [\tilde{X}_1^T, \cdots, \tilde{X}_K^T]^T$, the number of salient facial spatial local region $\kappa$, the scalar parameter $\rho$, $\mu_{max}$.
  • Initializing the regression matrix $C = [C_1^T, \cdots, C_K^T]^T$
  • Initializing the Lagrangian multiplier matrix $P = [P_1^T, \cdots, P_K^T]^T$ and the trade off coefficient $\mu$.
Repeating steps 1) to 4) until convergence.
1: Fix $C, P, \mu$ and update $D$:
     $D = \left(\mu I_{Kd} + 2\tilde{X}\tilde{X}^T\right)^{-1} \left(2\tilde{X}\tilde{L}^T + P + \mu C\right)$,   (12)
   where $I_{Kd} \in \mathbb{R}^{Kd\times Kd}$ is the identity matrix.
2: Fix $D, P, \mu$ and update $C$:
     Calculate $d_i = \left\| D_i - \frac{P_i}{\mu} \right\|_F$, and sort the value of
   $d_i$, such that $d_{i_1} \geq d_{i_2} \geq \cdots \geq d_{i_K}$, Let $\lambda = \mu d_{i_{\kappa+1}}$,
   then update $C$ according to
     $C_i = \begin{cases} \frac{d_i - \frac{\lambda}{\mu}}{d_i}(D_i - \frac{P_i}{\mu}), & \frac{\lambda}{\mu} < d_i, \\ \mathbf{0}, & \frac{\lambda}{\mu} \geq d_i. \end{cases}$   (13)
3: Update $P$ and $\mu$:
     $P = P + \mu(D - C)$, $\mu = \min(\rho\mu, \mu_{max})$
4: Check convergence:
     $\|C - D\|_\infty < \varepsilon$
Output: The solved $\hat{C}$ of regression matrix $C$.
```

D. Application for CDMER

Based on the labeled source and the unlabeled target databases, TGSR can get the solved $\hat{\boldsymbol{C}}$ of the regression matrix $\boldsymbol{C}$. Then we can predict the label vector $\boldsymbol{l}^{te}$ of the micro-expression sample with feature $\boldsymbol{x}_i^{te} \in \mathbb{R}^{Kd}$ by solving the optimization problem as Equ. (14),

$$\min_{\boldsymbol{l}^{te}} \left\| \boldsymbol{l}^{te} - \sum_{i=1}^{K} \hat{\boldsymbol{C}}_i^{\mathrm{T}} \boldsymbol{x}_i^{te} \right\|_F^2 , \quad \text{s.t.}\quad \boldsymbol{l}^{te} \geq 0.\mathbf{1}^{\mathrm{T}}\boldsymbol{l}^{te} = 1, \tag{14}$$

where $\hat{\boldsymbol{C}}_i \in \mathbb{R}^{d\times C}$ is the solved regression matrix of i-th facial spatial local region, and $\hat{\boldsymbol{C}}^{\mathrm{T}} = \left[\hat{\boldsymbol{C}}_1^{\mathrm{T}}, \cdots, \hat{\boldsymbol{C}}_K^{\mathrm{T}}\right]$, $\hat{\boldsymbol{C}}^{\mathrm{T}} \in \mathbb{R}^{C\times Kd}$, $\boldsymbol{l}^{te} \in \mathbb{R}^{C}$. Then we can use $\hat{c} = \arg\max_j \left\{ \boldsymbol{l}_j^{te} \right\}$ to assign its micro-expression label to the largest entry index of the predict label vector, i.e., the micro-expression category $\hat{c}$.

III. Experiment

A. Experiment Setup

Database. We evaluated our method on Selected CASME II and SMIC database. CASME II[34] contains 255 micro-

TABLE I
THE STATISTICS OF SELECTED CASME II AND SMIC DATABASE.

Dataset	Category		
	Positive	Negative	Surprise
Selected CASME II	32	73	25
SMIC-HS	51	70	43
SMIC-VIS	23	28	20
SMIC-NIR	23	28	20

expression samples from 26 subjects with seven category micro-expressions, i.e., *Disgust*, *Fear*, *Happiness*, *Others*, *Repression*, *Sadness*, and *Surprise*. We selected the samples of *Disgust*, *Happiness*, *Repression*, and *Surprise* to be the Selected CASME II. SMIC[35] records three modality samples of 16 subjects with three category micro-expressions, i.e., *Positive*, *Negative*, and *Surprise*. Specially, a high-speed camera at 100 frames/s captures the HS subset, which contains 164 samples; a general visual camera at 25 frames/s captures the VIS subset, which includes 71 samples; and a near-infrared camera captures the NIR subset, which contains 71 samples. To make Selected CASME II and the three subsets of the SMIC database share the same labels, we further relabelled the Selected CASME II: relabelled the samples of *Happiness* into *Positive*, the samples of *Disgust* and *Repression* into *Negative*, and maintain the samples of *Surprise* with label *Surprise*. Tab. I summarizes the basic information of Selected CASME II and SMIC.

Protocol. The cross-database protocol is designed to develop models with promising domain adaption performance operated by training model in Source database (S) and testing in the Target database (T), which is denoted as S→T. Following [36], we employed two types of unsupervised CDMER experiments: TYPE-I is implemented between every two subsets of SMIC, and TYPE-II is implemented between Selected CASME II and every subset of SMIC. We denote SMIC-HS, SMIC-VIS, and SMIC-NIR as H, V, N, and CASME II as C for short. Specially, we list the experiments of TYPE-I: H→V, V→H, H→N, N→H, V→N, N→V, and the experiments of TYPE-II: C→H, H→C, C→V, V→C, C→N, N→C.

Evaluation Metrics. We employed macro F1-score (M-F1) and accuracy (ACC) to evaluate our method. Macro F1-score is calculated by $M-F1 = \frac{1}{C}\sum_{c=1}^{C}\frac{2p_c r_c}{p_c+r_c}$, where p_c and r_c are the precision and recall of the c-th category micro-expression, and C is the category number. Macro F1-score is suitable because unbalanced sample problems widely existed in the CDMER problem.

Implementation Detail. In the experiments, we used the facial landmarks from the first frame of each ME sequence to make the bounding box for face region cropping, then transformed the cropped face of CASME II and SMIC into 308 × 257 pixels and 170 × 139 pixels, respectively. We employed the Temporal Interpolation Model (TIM)[48], [49] to normalize the ME sequence into fixed 16 frames in temporal sequence and then resized them into 112×112 pixels in spatial. We used a grid-based multi-scale spatial division scheme for a ME sequence to divide the whole face into 1×1, 2×2, 4×4, 8×8 four scales totally $K = 85$ face block sequence to extract and concatenate corresponding LBP-TOP features[18] to serve as the micro-expression representation. Here, the neighboring radius R and the number of neighboring points P for LBP-TOP are set as $(R, P) = (3, 8)$. Two hyper-parameters are involved in solving our proposed method, i.e., the salient or valid facial block number κ and the trade-off hyper-parameter ξ for weighting the contribution of the MMD term. Here, salient facial block number κ is an integer variable, and trade-off hyper-parameter ξ is a consecutive variable. Following the work in [46], [36], we used a grid searching strategy to search the optimal hyper-parameters for the best macro F1-score, and we also reported the corresponding accuracy. Specially, we searched the hyper-parameter κ from a preset parameter interval [1:1:85], and searched the hyper-parameter ξ from a preset parameter interval [0.001:0.0002:0.01 0.01:0.002:0.1 0.1:0.02:1 1:0.2:10 10:2:100 100:20:1000].

B. Results and Analysis

Tab. II and Tab. III respectively show the TYPE-I and TYPE-II experimental results. We calculated the average performance (macro F1-score and Accuracy) of these two types of experiments. In the TYPE-I task, from Exp.1 to Exp.6, SVM[39], IW-SVM[40], TCA[41], GFK[42], SA[43], STM[44], TKL[45], TSRG[46], DRLS[47] achieved 0.6003, 0.6911, 0.6238, 0.7223, 0.6917, 0.6440, 0.6964, 0.6991, 0.6552 in term of Macro F1-score, and achieved 61.62%, 69.74%, 64.01%, 72.44%, 69.44%, 65.78%, 69.92%, 70.05%, 66.60% in terms of accuracy. In the TYPE-II task, from Exp.7 to Exp.12, SVM[39], IW-SVM[40], TCA[41], GFK[42], SA[43], STM[44], TKL[45], TSRG[46], DRLS[47] achieved 0.4112, 0.4876, 0.5177, 0.5161, 0.4877, 0.3982, 0.5051, 0.5348, 0.5102 in terms of macro F1-score, and achieved 45.55%, 50.73%, 53.45%, 54.31%, 50.90%, 45.61%, 52.33%, 56.22%, 53.71% in terms of accuracy. From the bold-lighted in Tab. II and Tab. III, we can see that proposed TGSR outperform those state-of-the-art methods on beyond half tasks and achieve the best performance in 7 of the total 12 tasks. In the TYPE-I task, from Exp.1 to Exp.6, the best macro F1-score is achieved in hyper-parameters (κ,ξ) at (85, 0.0022), (46, 0.0036), (14, 4000), (85, 0.0044), (12, 44), (12, 280), respectively. In the TYPE-II task,from Exp.7 to Exp.12, the best macro F1-score is achieved in hyper-parameters (κ,ξ) at (62,0.0012), (28,0.0980), (85,0.0030), (85,0.0280), (85,0.0016), (75,0.0220), respectively.

In addition, at least three apparent characteristics we can find in the tables. Firstly, we find that domain adaption tricks can promote method performance. IW-SVM vastly outperforms SVM in average performance by learning a group of importance weights: in the TYPE-I task, it improves 0.0908 and 8.12% in terms of macro F1-score and accuracy, and in the TYPE-II task, it improves 0.0764 and 5.18% in terms of macro F1-score and accuracy. And the domain-adaption-based methods perform better than the non-domain-adaption based:

TABLE II
EXPERIMENTAL RESULTS (MARCO F1-SCORE AND ACCURACY) ON EITHER TWO SUBSETS OF SMIC, I.E., HS, VIS, AND NIR. THE MICRO-EXPRESSIONS ARE NEGATIVE, POSITIVE, AND SURPRISE. THE BEST RESULTS IN EACH EXPERIMENT ARE HIGHLIGHTED IN BOLD. MACRO F1-SCORE IS DENOTED BY M-F1, AND ACCURACY IS DENOTED BY ACC.

Method	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[39]	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[40]	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[41]	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[42]	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[43]	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[44]	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[45]	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[46]	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[47]	0.8604	85.92	0.6120	60.98	0.6599	66.20	0.5599	55.49	0.6620	69.01	0.5771	61.97
Ours	**0.9150**	**91.55**	**0.6226**	**62.20**	0.5847	60.56	**0.6272**	**61.59**	0.6984	70.42	**0.8403**	**84.51**

TABLE III
EXPERIMENTAL RESULTS (MACRO F1-SCORE AND ACCURACY) ON SELECTED CASME II AND ONE SUBSET OF SMIC (HS, VIS, AND NIR) DATABASE. THE MICRO-EXPRESSION CATEGORY INCLUDES NEGATIVE, POSITIVE, AND SURPRISE. THE BEST RESULTS IN EACH EXPERIMENT ARE HIGHLIGHTED IN BOLD. MACRO F1-SCORE IS DENOTED BY M-F1, AND ACCURACY IS DENOTED BY ACC.

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)	
	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC
SVM[39]	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[40]	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[41]	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[42]	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[43]	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[44]	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[45]	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[46]	**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[47]	0.4924	**53.05**	0.5267	59.23	0.5757	57.75	0.5942	60.00	0.4885	49.83	0.3838	42.37
Ours	0.5001	51.83	0.5061	56.92	0.5906	59.15	**0.6403**	**63.85**	**0.5697**	**57.75**	**0.4474**	**48.46**

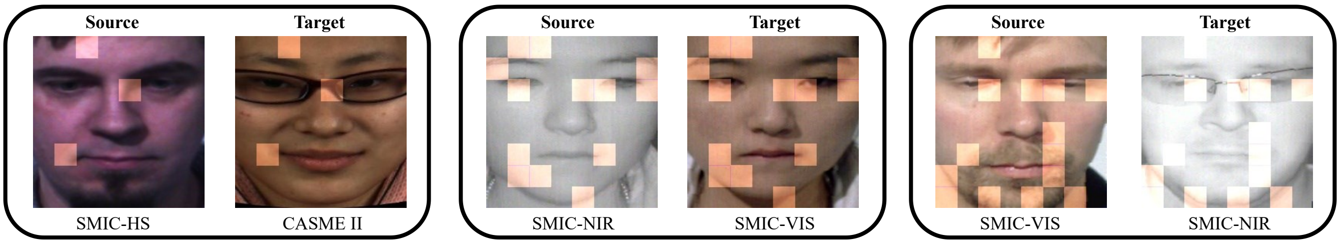

Fig. 2. The visualization of salient facial regions selected by proposed TGSR method for three cross-database micro-expression tasks. The salient facial regions are highlighted above and evidenced by the regression matrix $\boldsymbol{C}$.

all methods except SVM used domain adaption tricks, and it is clear that they are better than SVM in average performance (both macro F1-score and accuracy) for all tasks steadily.

Secondly, From Tab. II and Tab. III, we observe that results in TYPE-I are generally better than TYPE-II. The tasks themselves cause it. Experiments in TYPE-I selected two different modalities in the SMIC database, and TYPE-II selected two different databases, i.e., CASME II and one subset of the SMIC database. Thirdly, we also find that the experiment exchanging the source and target database has an obvious performance gap: the performance of SVM is better in Exp.1(H→V) than in Exp.2(V→H), better in Exp.3(H→N) than in Exp.4(N→H), better in Exp.11(C→N) than in Exp.12(N→C). Exp.1(H→V) used the high-speed camera captured image sequences as the source database and the general visual camera captured as the target database, which is exchanged in Exp.2(V→H). We guess that a high-speed camera may capture more subtle spatial-temporal facial action that contributed to the MER performance. Exp.3(H→V) used the high-speed camera captured image sequences as the source database and the near-infrared image sequence as the target database, which is exchanged as the target database in Exp.4(N→H). The high-speed camera captures subset HS and the near-infrared captures subset NIR are both from the SMIC

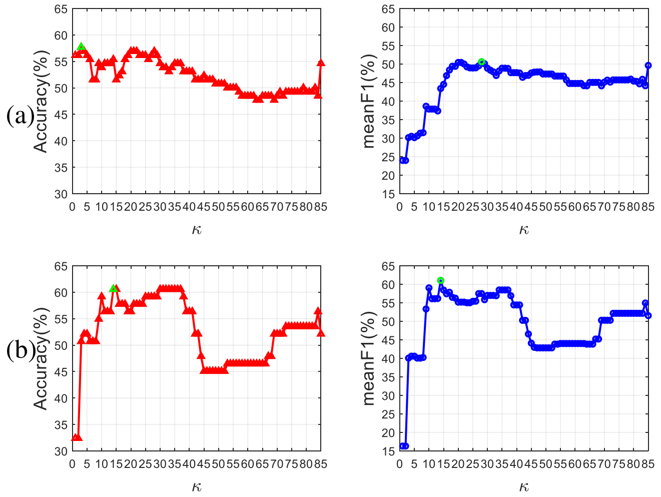

Fig. 3. The hyper-parameter discussion of selected facial local region number κ. (a) shows the experimental results of Exp.8(H→C) and (b) shows the experimental results of Exp.4(H→N).

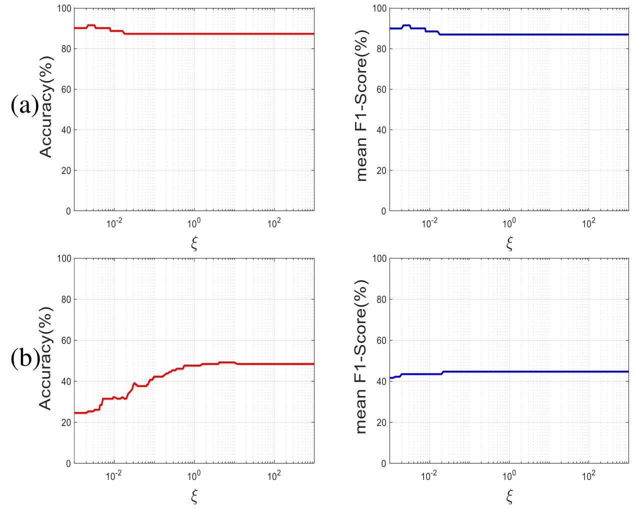

Fig. 4. The hyper-parameter discussion of trade-off coefficient ξ. (a) shows experimental results of Exp.1(H→V), and (b) shows the experimental results of Exp.12(N→C).

database, but the near-infrared subset NIR discards color information while the high-speed subset HS preserved. We find that learning the MER model from colored to uncolored shows better performance than that from uncolored to colored, which also verifies the research that human understanding of emotions is color-related in [50]. The same phenomenon exists in Exp.11(C→N) and Exp.12(N→C) which involving two different databases, i.e., CASME II and SMIC. But the gap between Exp.11(C→N) and Exp.12(N→C) is larger than Exp.3(N→H) and Exp.4(N→H), which tells us the database difference is larger than the color difference.

C. Hyper-parameter Discussion

Two hyper-parameters are involved in solving the optimal coefficient matrix $\boldsymbol{C}$ of proposed TGSR, i.e., the salient facial block number κ and the trade-off hyper-parameter ξ for weighting MMD term. Their settings affect model performance; thus, it is easy to question whether the proposed TGSR performs well in a steady hyper-parameter range. We conducted two experiments to investigate model sensitiveness to hyper-parameter κ and ξ. In the first experiment, we fixed hyper-parameter ξ and varied the selected facial block number κ from 1 to $K = 85$ to explore model sensitiveness to hyper-parameter κ and then recorded the corresponding macro F1-score and accuracy. We selected Exp.4(H→N) and Exp.8(H→C) as the typical, on behalf of TYPE-I and TYPE-II CDMER task, to discuss model sensitiveness to hyper-parameter κ and show corresponding performance curve as Fig. 3. We can apparently see that the performance of the first hyper-parameter discussion experiment peaks at a low κ and then deteriorates as the increment of κ.It indicates that salient facial regions for MER are exiguous and valuable features may be drowned. In the second experiment, we fixed hyper-parameter κ and varied hyper-parameter coefficient ξ from 10^{-3} to 10^{3}, to explore model sensitiveness to hyper-parameter ξ. We selected Exp.1(H→V) and Exp.12(N→C) as the typical, on behalf of TYPE-I and TYPE-II CDMER task, to discuss the sensitiveness to hyper-parameter ξ, and show the performance as Fig. 4. From Fig. 4, we find that MMD tends to improve model performance across a wide range of hyper-parameter ξ, and a proper selection of hyper-parameter ξ will yield better performance.

D. Visualization

To verify if the proposed TGSR learned the correct facial local region for the CDMER task, we selected Exp.5(V→N), Exp.6(N→V), and Exp.8(H→C) as the typical, to visualize the selected facial region as Fig. 2, which evidenced by that nonzero element in the regression matrix $\boldsymbol{C}$. We observe that the highlighted area focuses on eye corners, mouth corners, and those facial areas with apparent muscle movement, which is consistent with the micro-expression definition in anatomy. Thus we can believe that TGSR achieved a competitive performance by learning an explicable feature.

IV. Conclusion

This paper proposes a novel Transfer Group Sparse Regression, TGSR, to cope with the Cross-Database Micro-Expression Recognition (CDMER) problem. TGSR seeks and selects the salient facial regions by learning a shared binary sparse regression matrix between the source and target databases to 1) promote a more precise measurement for better alleviating the feature difference between the source and target databases in the label space, and to 2) improve the extracted hand-crafted feature to become more effective and explicable for better MER. Experiments and visualizations demonstrate that TGSR effectively learns well-designed features and outperforms most state-of-the-art subspace-learning-based domain adaption methods for CDMER.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (NSFC) under the Grants U2003207, 61921004, and 61902064, in part by the Fundamental Research Funds for the Central Universities under Grant 2242022k30036, and in part by the Zhishan Young Scholarship of Southeast University.

References

[1] P. Ekman and W. V. Friesen, "Nonverbal leakage and clues to deception," *Psychiatry*, vol. 32, no. 1, pp. 88–106, 1969. 1

[2] P. Ekman, *Telling lies: Clues to deceit in the marketplace, politics, and marriage (revised edition)*. WW Norton & Company, 2009. 1

[3] 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*, vol. 37, no. 4, pp. 217–230, 2013. 1

[4] 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*, 2009, pp. 1–35. 1

[5] X. Jiang, Y. Zong, W. Zheng, C. Tang, W. Xia, C. Lu, and J. Liu, "Dfew: A large-scale database for recognizing dynamic facial expressions in the wild," in *Proc. ACM MM*, 2020, pp. 2881–2889. 1

[6] S. Li, W. Zheng, Y. Zong, C. Lu, C. Tang, X. Jiang, J. Liu, and W. Xia, "Bi-modality fusion for emotion recognition in the wild," in *Proc. ICMI*, 2019, pp. 589–594. 1

[7] 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 Trans. Ind. Informat.*, vol. 14, no. 9, pp. 3822–3830, 2017. 1

[8] C. Yang, J. Luo, Y. Pan, Z. Liu, and C.-Y. Su, "Personalized variable gain control with tremor attenuation for robot teleoperation," *IEEE Trans. Syst., Man, Cybern.*, vol. 48, no. 10, pp. 1759–1770, 2017. 1

[9] T. Pfister, X. Li, G. Zhao, and M. Pietikäinen, "Recognising spontaneous facial micro-expressions," in *Proc. ICCV*. IEEE, 2011, pp. 1449–1456. 1

[10] 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 *Proc. ACCV*. Springer, 2014, pp. 525–537. 1

[11] 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 Trans. Image Process.*, vol. 24, no. 12, pp. 6034–6047, 2015. 1

[12] 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 Trans. Affect. Comput.*, vol. 7, no. 4, pp. 299–310, 2015. 1

[13] D. H. Kim, W. J. Baddar, and Y. M. Ro, "Micro-expression recognition with expression-state constrained spatio-temporal feature representations," in *Proc. ACM MM*, 2016, pp. 382–386. 1

[14] 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*, vol. 99, no. 6, pp. 1694–1697, 2016. 1

[15] F. Xu, J. Zhang, and J. Z. Wang, "Microexpression identification and categorization using a facial dynamics map," *IEEE Trans. Affect. Comput.*, vol. 8, no. 2, pp. 254–267, 2017. 1

[16] S. Happy and A. Routray, "Fuzzy histogram of optical flow orientations for micro-expression recognition," *IEEE Trans. Affect. Comput.*, vol. 10, no. 3, pp. 394–406, 2017. 1

[17] Y. Zong, X. Huang, W. Zheng, Z. Cui, and G. Zhao, "Learning from hierarchical spatiotemporal descriptors for micro-expression recognition," *IEEE Trans. Multimedia*, vol. 20, no. 11, pp. 3160–3172, 2018. 1

[18] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 915–928, 2007. 1, 4

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Proc. NeurIPS*, vol. 25, pp. 1097–1105, 2012. 1

[20] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997. 1

[21] M. Wei, W. Zheng, Y. Zong, X. Jiang, C. Lu, and J. Liu, "A novel micro-expression recognition approach using attention-based magnification-adaptive networks," in *Proc. ICASSP*. IEEE, 2022, pp. 2420–2424. 1

[22] J. Liu, W. Zheng, and Y. Zong, "Sma-stn: Segmented movement-attending spatiotemporal network formicro-expression recognition," *arXiv preprint arXiv:2010.09342*, 2020. 1

[23] W. Xia, W. Zheng, Y. Zong, and X. Jiang, "Motion attention deep transfer network for cross-database micro-expression recognition," in *Proc. ICPR*. Springer, 2021, pp. 679–693. 1

[24] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, 2009. 1

[25] B. Schuller, B. Vlasenko, F. Eyben, M. Wöllmer, A. Stuhlsatz, A. Wendemuth, and G. Rigoll, "Cross-corpus acoustic emotion recognition: Variances and strategies," *IEEE Trans. Affect. Comput.*, vol. 1, no. 2, pp. 119–131, 2010. 1

[26] J. Liu, W. Zheng, Y. Zong, C. Lu, and C. Tang, "Cross-corpus speech emotion recognition based on deep domain-adaptive convolutional neural network," *IEICE Transactions on Information and Systems*, vol. 103, no. 2, pp. 459–463, 2020. 1

[27] S. Zhao, X. Zhao, G. Ding, and K. Keutzer, "Emotiongan: Unsupervised domain adaptation for learning discrete probability distributions of image emotions," in *Proc. ACM MM*, 2018, pp. 1319–1327. 1

[28] 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 *Proc. AAAI*, vol. 33, no. 01, 2019, pp. 2620–2627. 1

[29] W.-L. Zheng and B.-L. Lu, "Personalizing eeg-based affective models with transfer learning," in *Proc. IJCAI*, 2016, pp. 2732–2738. 1

[30] 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 Trans. Affect. Comput.*, 2018. 1

[31] Y. Zong, W. Zheng, Z. Cui, G. Zhao, and B. Hu, "Toward bridging microexpressions from different domains," *IEEE Transactions on Cybernetics*, vol. 50, no. 12, pp. 5047–5060, 2019. 1

[32] L. Li, X. Zhou, Y. Zong, W. Zheng, X. Chen, J. Shi, and P. Song, "Unsupervised cross-database micro-expression recognition using target-adapted least-squares regression," *IEICE Transactions on Information and Systems*, vol. 102, no. 7, pp. 1417–1421, 2019. 1

[33] X. Chen, X. Zhou, C. Lu, Y. Zong, W. Zheng, and C. Tang, "Target-adapted subspace learning for cross-corpus speech emotion recognition," *IEICE Transactions on Information and Systems*, vol. 102, no. 12, pp. 2632–2636, 2019. 1

[34] 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*, vol. 9, no. 1, p. e86041, 2014. 1, 3

[35] X. Li, T. Pfister, X. Huang, G. Zhao, and M. Pietikäinen, "A spontaneous micro-expression database: Inducement, collection and baseline," in *Proc. FG*. IEEE, 2013, pp. 1–6. 1, 4

[36] 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 Trans. Knowl. Data Eng.*, 2020. 2, 4

[37] Z. T. Qin and D. Goldfarb, "Structured sparsity via alternating direction methods." *Journal of Machine Learning Research*, vol. 13, no. 5, 2012. 3

[38] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *arXiv preprint arXiv:1009.5055*, 2010. 3

[39] C.-C. Chang and C.-J. Lin, "Libsvm: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1–27, 2011. 4, 5

[40] A. Hassan, R. Damper, and M. Niranjan, "On acoustic emotion recognition: compensating for covariate shift," *IEEE Trans. Speech Audio Process.*, vol. 21, no. 7, pp. 1458–1468, 2013. 4, 5

[41] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 199–210, 2010. 4, 5

[42] B. Gong, Y. Shi, F. Sha, and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in *Proc. CVPR*. IEEE, 2012, pp. 2066–2073. 4, 5

[43] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars, "Unsupervised visual domain adaptation using subspace alignment," in *Proc. ICCV*, 2013, pp. 2960–2967. 4, 5

[44] W.-S. Chu, F. De la Torre, and J. F. Cohn, "Selective transfer machine for personalized facial action unit detection," in *Proc. CVPR*, 2013, pp. 3515–3522. 4, 5

[45] M. Long, J. Wang, J. Sun, and S. Y. Philip, “Domain invariant transfer kernel learning,” *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 6, pp. 1519–1532, 2014. 4, 5

[46] Y. Zong, X. Huang, W. Zheng, Z. Cui, and G. Zhao, “Learning a target sample re-generator for cross-database micro-expression recognition,” in *Proc. ACM MM*, 2017, pp. 872–880. 4, 5

[47] Y. Zong, W. Zheng, X. Huang, J. Shi, Z. Cui, and G. Zhao, “Domain regeneration for cross-database micro-expression recognition,” *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2484–2498, 2018. 4, 5

[48] Z. Zhou, G. Zhao, and M. Pietikäinen, “Towards a practical lipreading system,” in *Proc. CVPR*. IEEE, 2011, pp. 137–144. 4

[49] Z. Zhou, X. Hong, G. Zhao, and M. Pietikäinen, “A compact representation of visual speech data using latent variables,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 1, pp. 1–1, 2013. 4

[50] 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*, vol. 115, no. 14, pp. 3581–3586, 2018. 6