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

Xingxun Jiang, Yuan Zong*, Wenming Zheng*, Jiateng Liu, Mengting Wei

Southeast University, Nanjing, China

{jiangxingxun, xhzongyuan, wenming_zheng, Jiateng_Liu, weimengting}@seu.edu.cn

1. Introduction

■ We investigate Cross-Database Micro-Expression Recognition (CDMER), in which the labeled training and unlabeled testing samples belong to different micro-expression databases.

■ Cross-Database Micro-Expression Recognition (CDMER) faces two obstacles: 1) The severe feature distribution gap between the training and test databases, 2) The feature representation bottleneck of micro-expression such local and subtle facial expression, which requires professional knowledge.

■ To cope with this challenging problem, we propose a new Transfer Group Sparse Regression (TGSR) method based on facial region selection.

2. Proposed Method – Basic Idea

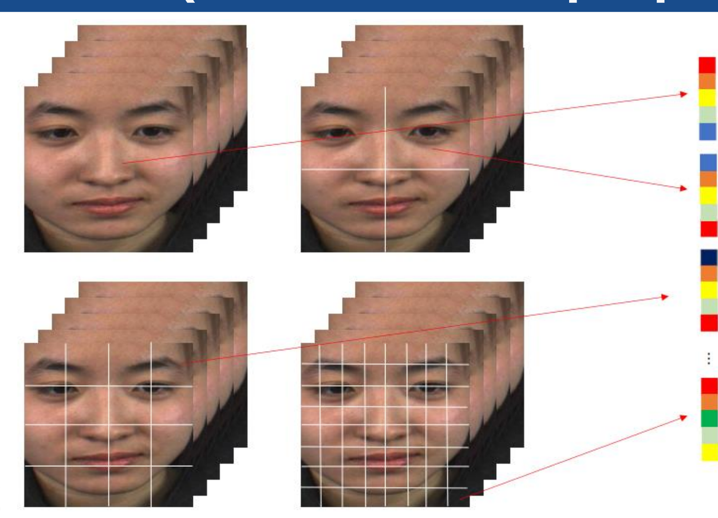
This paper adopts the facial region selection method to seek the salient regions from the whole face to

■ 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,

■ improve the extracted hand-crafted feature to become more effective and explicable for better CDMER.

3. Proposed Method – Detail (Transfer Group Sparse Regression Method, TGSR)

Feature extraction



1. A face is split into K local sequence (block);
2. Each spatial block: d dim feature $\boldsymbol{x}_k$;
3. Each micro-expression: $\boldsymbol{x}^v = [\boldsymbol{x}_1^{\mathrm{T}}, \cdots, \boldsymbol{x}_K^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{Kd}$;
4. Source database (N_s samples)
   - i-th facial block's feature is denoted by $\boldsymbol{X}_i^s \in \mathbb{R}^{d \times N_s}$;
   - Database's feature is denoted by $\boldsymbol{X}^s = \left[\boldsymbol{X}_1^{s\,\mathrm{T}}, \cdots, \boldsymbol{X}_K^{s\,\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{R}^{Kd \times N_s}$;
   - j-th sample's label is denoted by one-hot vector $\boldsymbol{l}_j^s = [l_{j,1}^s, \cdots, l_{j,C}^s]^{\mathrm{T}}$ which only $l_{j,c}^s$ equals to 1, indicating that sample belong to this category;
   - The label is denoted by $\boldsymbol{L}^s = [\boldsymbol{l}_1^s, \cdots, \boldsymbol{l}_{N_s}^s] \in \mathbb{R}^{C \times N_s}$, C is class number.
5. Target database (N_t samples)
   - i-th facial block's feature is denoted by $\boldsymbol{X}_i^t \in \mathbb{R}^{d \times N_t}$;
   - Database's feature is denoted by $\boldsymbol{X}^t = \left[\boldsymbol{X}_1^{t\,\mathrm{T}}, \cdots, \boldsymbol{X}_K^{t\,\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{R}^{Kd \times N_t}$.

Objective Function

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

The problem can be rewritten into

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

where $\tilde{\boldsymbol{L}} = [\boldsymbol{L}^s, \boldsymbol{0}], \tilde{\boldsymbol{X}}_i = \left[\boldsymbol{X}_i^s, \sqrt{\xi}\left(\frac{1}{N_s}\boldsymbol{X}_i^s \boldsymbol{1}_s - \frac{1}{N_t}\boldsymbol{X}_i^t \boldsymbol{1}_t\right)\right]$.

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. (2) can be converted in to Equ. (3)

$$\Gamma(\boldsymbol{C}_i, \boldsymbol{D}_i, \boldsymbol{P}_i, \mu) = \left\| \tilde{\boldsymbol{L}} - \sum_{i=1}^{K} \boldsymbol{D}_i^{\mathrm{T}} \tilde{\boldsymbol{X}}_i \right\|_F^2 + \lambda \sum_{i=1}^{K} \| \boldsymbol{C}_i \|_F + \sum_{i=1}^{K} \mathrm{tr}\left[\boldsymbol{P}_i^{\mathrm{T}}(\boldsymbol{C}_i - \boldsymbol{D}_i)\right] + \frac{\mu}{2} \sum_{i=1}^{K} \| \boldsymbol{C}_i - \boldsymbol{D}_i \|_F^2, \qquad (3)$$

1 2
3 4

Solved Solution

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

Input: $\tilde{\boldsymbol{L}}, \tilde{\boldsymbol{X}} = [\tilde{\boldsymbol{X}}_1^{\mathrm{T}}, \cdots, \tilde{\boldsymbol{X}}_K^{\mathrm{T}}]^{\mathrm{T}}$, salient region number κ , scalar parameter ρ and $\mu_{\max}$.

●Initializing the regression matrix $\boldsymbol{C} = [\boldsymbol{C}_1^{\mathrm{T}}, \cdots, \boldsymbol{C}_K^{\mathrm{T}}]^{\mathrm{T}}$, the Lagrangian multiplier matrix $\boldsymbol{P} = [\boldsymbol{P}_1^{\mathrm{T}}, \cdots, \boldsymbol{P}_K^{\mathrm{T}}]^{\mathrm{T}}$, and trade-off coefficient μ.

Repeating step 1) to 4) until convergence.

(1) Fix $\boldsymbol{C}, \boldsymbol{P}, \mu$ and update $\boldsymbol{D}$: $\boldsymbol{D} = (\mu \boldsymbol{I}_{Kd} + 2\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^{\mathrm{T}})^{-1}(2\tilde{\boldsymbol{X}}\tilde{\boldsymbol{L}}^{\mathrm{T}} + \boldsymbol{P} + \mu\boldsymbol{C})$

(2) Fix $\boldsymbol{D}, \boldsymbol{P}, \mu$ and update $\boldsymbol{C}$: Calculate $d_i = \left\|\boldsymbol{D}_i - \frac{\boldsymbol{P}_i}{\mu}\right\|_F$, 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

$$\boldsymbol{C}_i = \begin{cases} \dfrac{d_i - \dfrac{\lambda}{\mu}}{d_i}\left(\boldsymbol{D}_i - \dfrac{\boldsymbol{P}_i}{\mu}\right), & \dfrac{\lambda}{\mu} < d_i \\ \boldsymbol{0}, & \dfrac{\lambda}{\mu} \geq d_i \end{cases}$$

(3) Update $\boldsymbol{P}$ and μ : $\boldsymbol{P} = \boldsymbol{P} + \mu(\boldsymbol{D} - \boldsymbol{C}), \mu = \min(\rho\mu, \mu_{\max})$.

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

Output: The solved $\hat{\boldsymbol{C}}$ of regression matrix $\boldsymbol{C}$.

Application

We first extract the feature $\boldsymbol{x}_i^{te} \in \mathbb{R}^{Kd}$ of the micro-express to be predicted, then we can predict its label vector $\boldsymbol{l}^{te}$ by solving the optimization problem as Equ.(8)

$$\min_{\boldsymbol{l}^{te}} \left\| \boldsymbol{l}^{te} - \sum_{i=1}^{K} \hat{\boldsymbol{C}}_i^{\mathrm{T}} \boldsymbol{x}_i^{te} \right\|_F^2, \qquad (8)$$
$$\text{s.t. } \boldsymbol{l}^{te} \geq 0\,;\ \boldsymbol{1}^{\mathrm{T}}\boldsymbol{l}^{te} = 1,$$

where $\hat{\boldsymbol{C}}_i \in \mathbb{R}^{d \times C}$ is the solved regression matrix of i-th facial spatial local region, and $\hat{\boldsymbol{C}}^{\mathrm{T}} = [\hat{\boldsymbol{C}}_1^{\mathrm{T}}, \cdots, \hat{\boldsymbol{C}}_K^{\mathrm{T}}], \hat{\boldsymbol{C}}^{\mathrm{T}} \in \mathbb{R}^{C \times Kd}, \boldsymbol{l}^{te} \in \mathbb{R}^{C}$. Then we can use Equ.(9) to assign its micro-expression label to the largest entry index of the predict label vector, i.e., micro-expression category $\hat{c}$.

$$\hat{c} = \arg\max_j \{\boldsymbol{l}_j^{te}\} \qquad (9)$$

4. Experiments and Results

■ **Protocol:** 1. TYPE-I: between every two subsets of SMIC; 2. TYPE-II: between Selected CASME II and every subset of SMIC;

■ **Evaluation metrics:** Accuracy and Macro F1-score

■ **Hyper-parameter Discussion**:
1) Salient facial region κ,
2) MMD weighting coefficient ξ.

■ **Visualization:** Salient facial regions.

Fig. 1 : The visualization of salient facial regions.

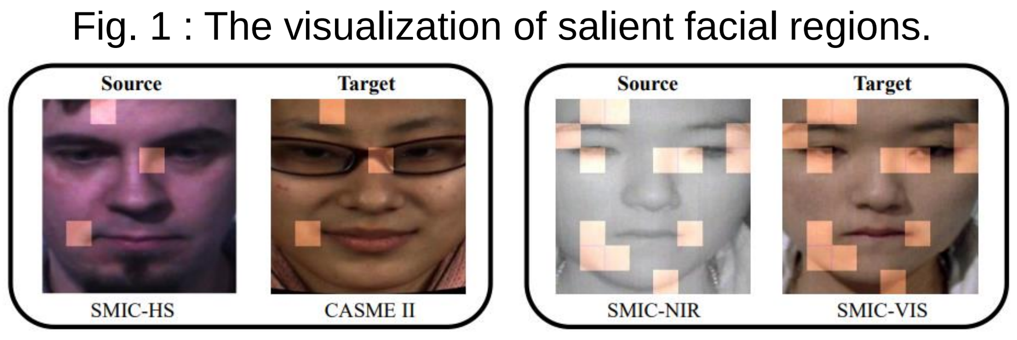


Table. I: The experimental results of TYPE-I CDMER task.

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. II: The experimental results of TYPE-II CDMER task .

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 discussion of salient facial region number.

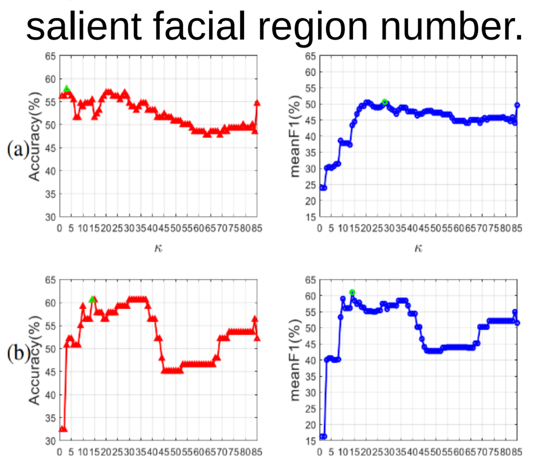
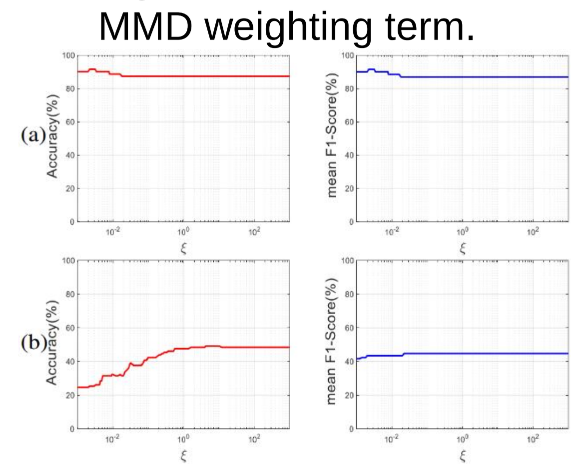


Fig. 3 : The discussion of MMD weighting term.



Conclusion and Discussion

We propose TGSR to seek the salient facial regions by learning a shared binary sparse regression matrix between the source and target databases to alleviate the severe feature difference between the source and target databases meanwhile improve extracted features. Experiments and visualizations demonstrate that TGSR outperforms most state-of-the-art domain adaption methods for CDMER.