Super Wide Regression Network for Unsupervised Cross-database Facial Expression Recognition

Na Liu [1,3], Baofeng Zhang [1], Yuan Zong [2], Li Liu [3], Jie Chen [3], Guoying Zhao [3], Junchao Zhu [1]

[1]Key laboratory of computer vision and systems, Ministry of education, Tianjin University of Technology, China

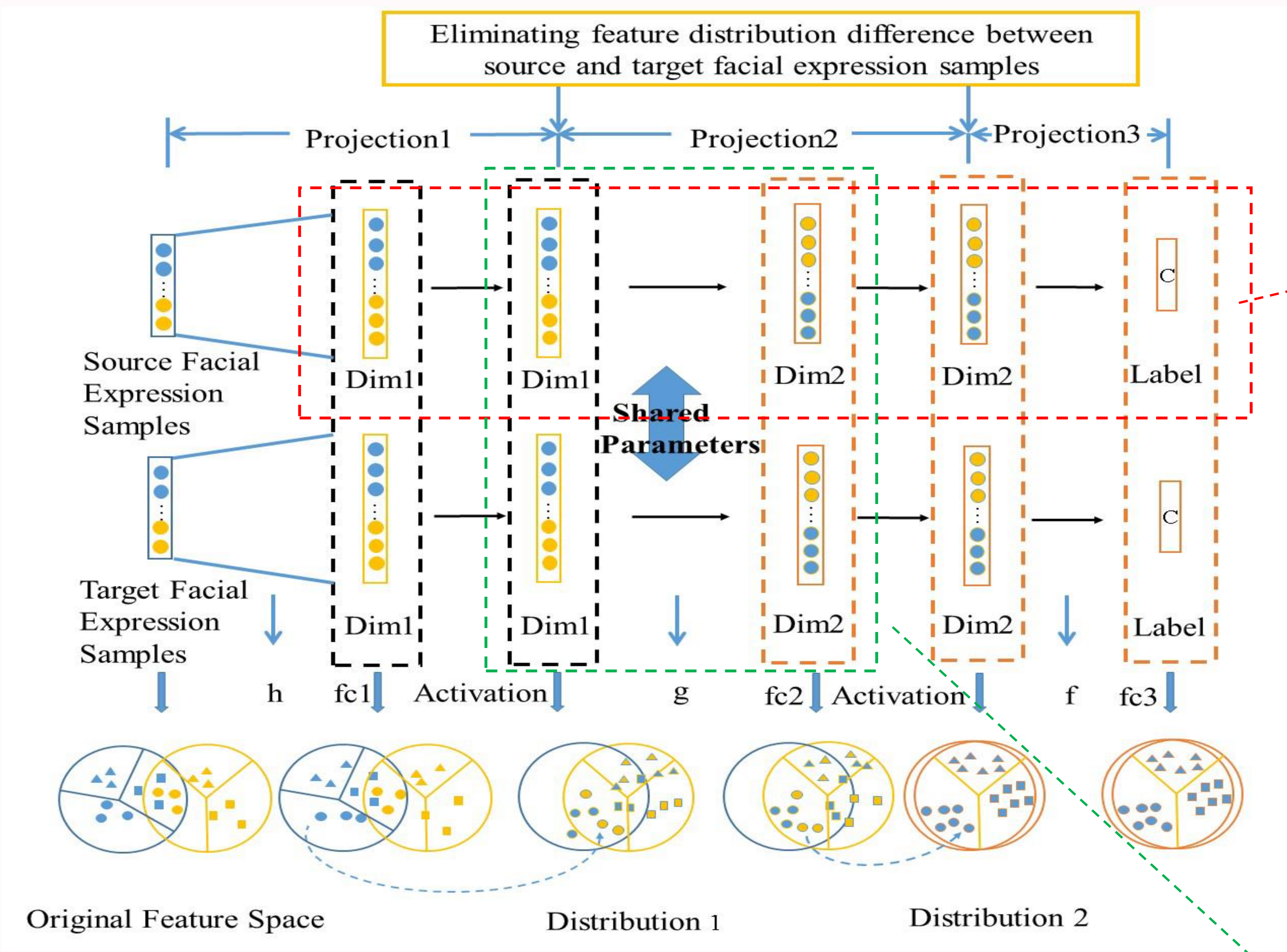
[2] Research Center for Learning Science, Southeast University, China

[3] Center for Machine Vision and Signal Analysis, University of Oulu, Finland

1.Introduction

- ***Motivation***: The problem of cross-database FER aims at learning a classifier based on a set of labeled training samples such that the learnt classifier can accurately predict the expression categories of the unlabeled testing samples.
- ***Solution***: A super wide network has been used to serve as the regression parameter instead of using projection matrix in subspace learning to build the relationship between the facial expression features and labels. Meanwhile, by using MMD criterion as regularization, we enforce the output of *SWiRN* with source and target samples as input, respectively, to have the same or similar feature distribution.
- ***Results***: *SWiRN* model achieves more promising performance than recent proposed cross-database emotion recognition

2.Method



Structure of the Super Wide Regression Network for Unsupervised Cross-Database Facial Expression Recognition

Learning a regression parameter to build the relationship between expression features and labels, which can be formulated as follows:

$$\min_{\mathbf{U}} \|\mathbf{L}^s - \mathbf{U}^T\mathbf{X}^s\|_F \quad (1)$$

we simply use the feature mean vector to measure the distribution distance between the source and target feature sets

$$\min_{\mathbf{U}} \left\| \frac{1}{n_s}\mathbf{U}^T\mathbf{X}_s\mathbf{1}_s - \frac{1}{n_t}\mathbf{U}^T\mathbf{X}_t\mathbf{1}_t \right\|^2 \quad (2)$$

By using Eq.(2) as the regularization term for Eq.(1), we will arrive at the optimization problem as follows:

$$\min_{\mathbf{U}} \|\mathbf{L}_s - \mathbf{U}^T\mathbf{X}_s\|_F^2 + \lambda \left\| \frac{1}{n_s}\mathbf{U}^T\mathbf{X}_s\mathbf{1}_s - \frac{1}{n_t}\mathbf{U}^T\mathbf{X}_t\mathbf{1}_t \right\|^2 \quad (3)$$

It should be pointed out that different from subspace learning version Eq.(3), the feature mean vector based regularization like Eq.(2) is simultaneously applied on two hidden layers and hence the final optimization problem of SWiRN becomes as follows:

$$\min_{f,g,h} \mathcal{L}(\mathbf{L}_s, f(g(h(\mathbf{X}_s))) + \lambda\left(\left\| \frac{1}{n_s} h(\mathbf{X}_s)\, \mathbf{1}_s - \frac{1}{n_t} h(\mathbf{X}_t)\, \mathbf{1}_t \right\|^2 + \left\| \frac{1}{n_s} g(\mathbf{X}_s)\, \mathbf{1}_s - \frac{1}{n_t} g(\mathbf{X}_t)\, \mathbf{1}_t \right\|^2 \right)$$

3.Experiments

fear surprise sadness angry disgust happiness

Facial expression from CK+

Results of the cross-corpus SER experiments in terms of UAR and WAR, where the common emotion states (6 classes) are Angry, Disgust, Fear, Happy, Sad and Surprise.

#	Source Database	Source Database	SVM		KMM+SVM		KLIEP+SVM		STM+SVM		SWiRN	
			UAR	WAR	UAR	WAR	UAR	WAR	UAR	WAR	UAR	WAR
1	CK+	Oulu-CASIA VIS	24.03	24.03	25.00	25.00	26.74	26.74	25.49	25.49	**28.33**	**28.33**
2	Oulu-CASIA VIS	CK+	29.52	47.57	32.68	47.57	58.67	62.78	43.07	39.48	**60.94**	**64.08**
3	CK+	eNTERFACE	16.90	16.94	21.26	21.29	**27.06**	**23.95**	18.76	18.80	20.35	20.33
4	eNTERFACE	CK+	11.45	18.54	23.39	28.48	27.60	23.95	26.11	22.65	**30.68**	**32.20**
5	eNTERFACE	Oulu-CASIA VIS	17.99	17.99	18.40	18.40	20.35	20.35	17.22	17.22	**26.95**	**26.95**
6	Oulu-CASIA VIS	eNTERFACE	17.67	17.72	18.99	19.04	18.37	18.41	18.06	18.10	**20.86**	**20.84**

Sample Number: CK+(309) eNTERFACE(1287) Oulu-CASIA VIS(1440)

4.Conclusion

- Small number of samples in source database will results in insufficient training of our model
- More unlabeled target samples may affect the discriminant ability of SWiRN since the label information given by source samples is so limited compared with a large number of unlabeled target samples.

5.Acknowledgements

This research was supported by the Natural Science Foundation of China under Grants 61172185 and 61602345, the Application Foundation and Advanced Technology Research Project of Tianjin, the Academy of Finland, Tekes Fidipro program and Infotech Oulu.