

Multi-label Co-regularization for Semi-supervised Facial Action Unit Recognition

Xuesong Niu¹, Hu Han^{1,2}, Shiguang Shan^{1,2}, Xilin Chen¹

¹Institute of Computing Technology, Chinese Academy of Sciences

²Peng Cheng Laboratory

{xuesong.niu}@vipl.ict.ac.cn, {hanhu, sgshan, xlchen}@ict.ac.cn

Paper :

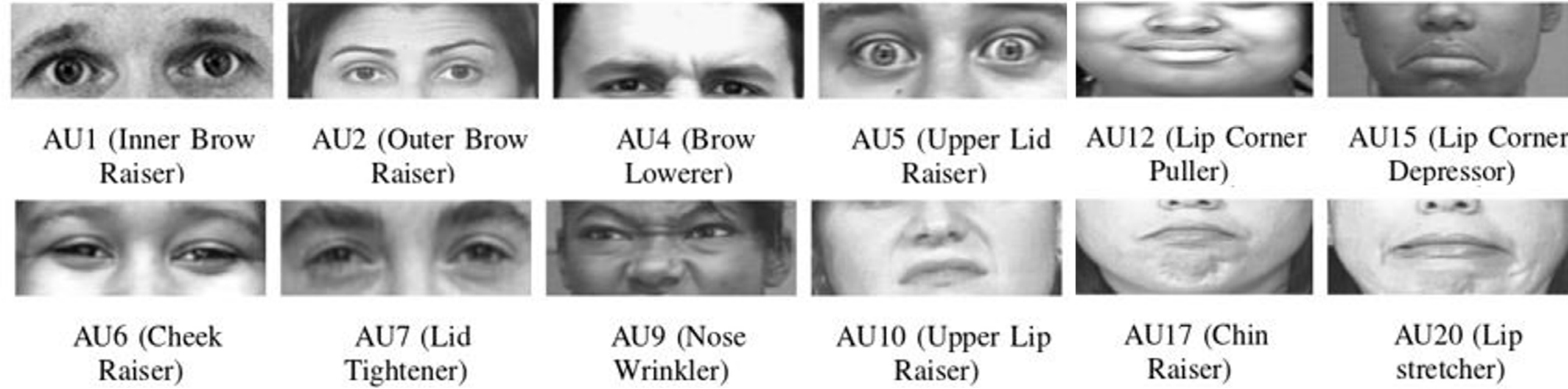


Code :



Problem

• What is facial action unit ?

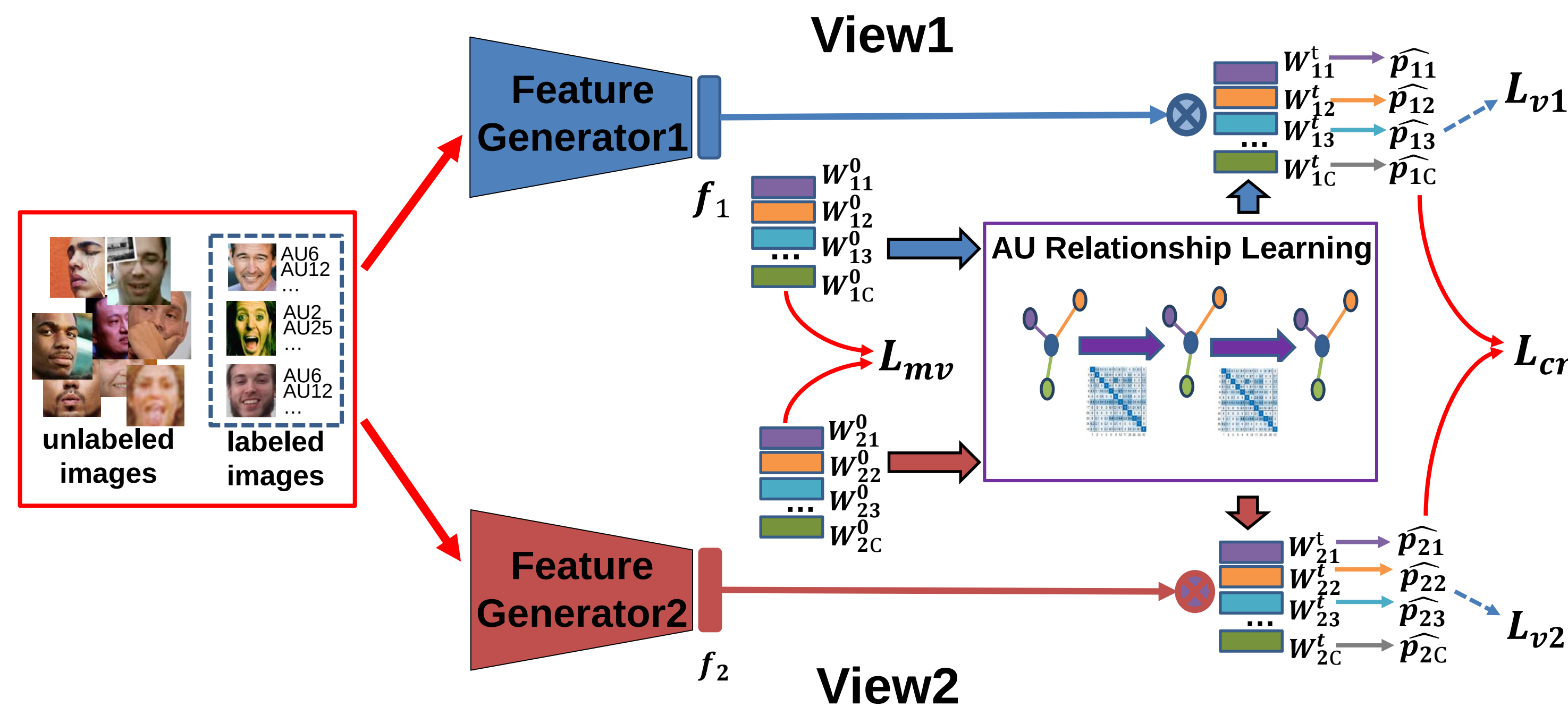


➤ Facial action units refer to a set of facial muscle movements coded by their appearance on the face, which can be used for coding nearly any anatomically possible facial expression.

➤ Since AUs are subtle, local and have significant subject-dependent variations, qualified FACS experts are required to annotate facial AUs. In addition, labeling AUs is labor-intensive and expensive, making it impractical to manually annotate a large set of face images.

Overview

• **Semi-supervised AU recognition** using multi-view co-training and AU relationship modeling to leverage massive labeled face images and a small set of labeled face images.



Method Details

• Multi-view Loss

$$L_{mv} = \frac{1}{C} \sum_{j=1}^C \frac{W_{1j}^T W_{2j}}{\|W_{1j}\| \|W_{2j}\|}$$

➤ A multi-view loss by orthogonalizing the weights of the AU classifiers of different views to encourage the two feature generators to get conditional independent features.

• Co-regularization Loss

➤ A co-regularization loss to encourage the classifiers from different views to generate similar predictions in order to make the two views to learn from each other.

$$L_{cr} = \frac{1}{C} \sum_{j=1}^C (H(\frac{\hat{p}_{1j} + \hat{p}_{2j}}{2}) - \frac{H(\hat{p}_{1j}) + H(\hat{p}_{2j})}{2})$$

$$H(p) = -(p \log p + (1 - p) \log(1 - p))$$

• AU Relationship Learning

➤ Two layer graph convolution layer to model the relationship between different AUs.

$$W_i^t = \hat{A} \text{ReLU}(\hat{A} W_i^0 H^{(0)}) H^{(1)}$$

➤ Adjacency matrix defined by the dependency matrix.

$$P_{dep} = \frac{1}{2} ([P(L_i = 1 | L_j = 1)]_{C \times C} + [P(L_i = -1 | L_j = -1)]_{C \times C})$$

$$\hat{A} = ABS((P_{dep} - 0.5) \times 2)$$

• Overall Loss Functions

➤ Binary cross-entropy loss for AU recognition

$$L_{vi} = -\frac{1}{C} \sum_{j=1}^C a_c [p_j \log \hat{p}_{ij} + (1 - p_j) \log(1 - \hat{p}_{ij})]$$

➤ Overall loss function

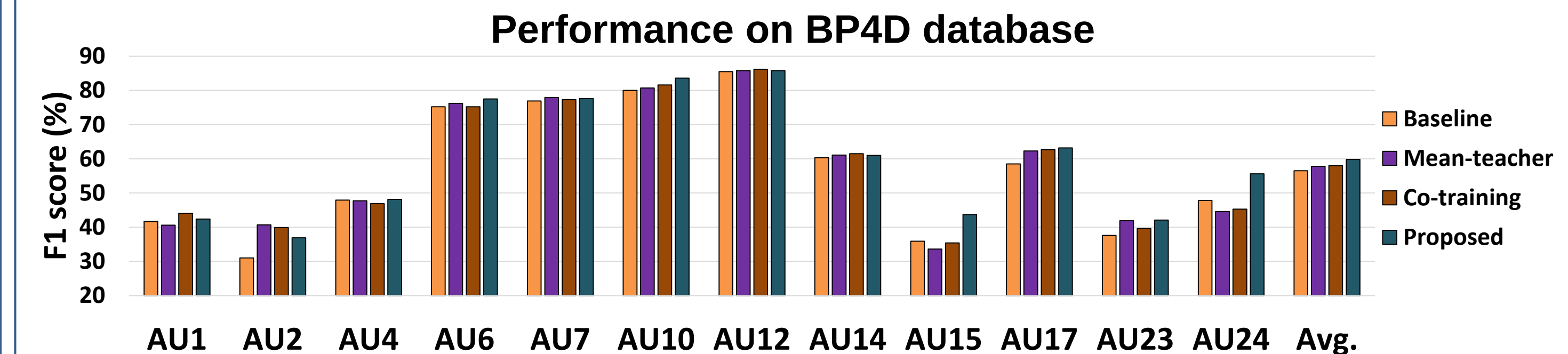
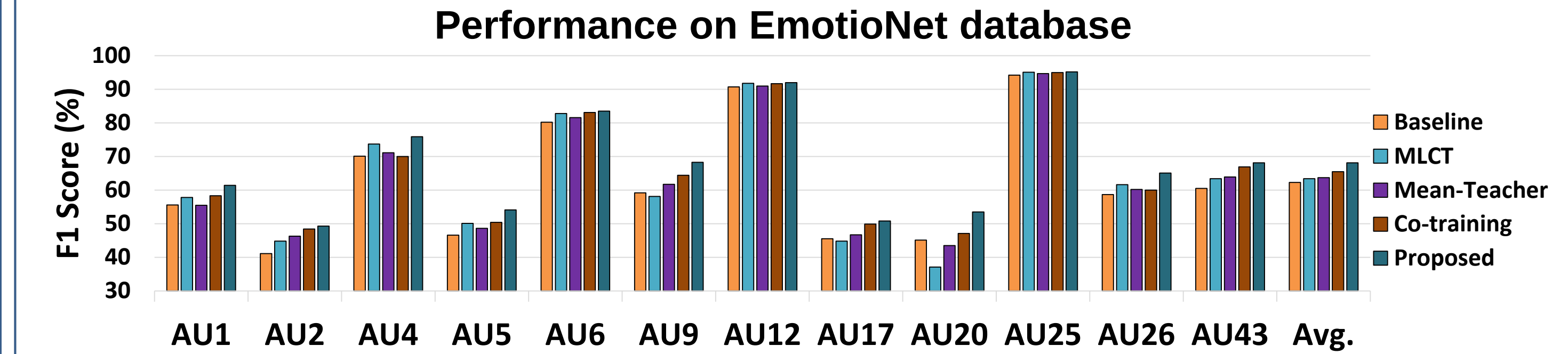
$$L = \frac{1}{2} \sum_{i=1}^2 L_{vi} + \lambda_{mv} L_{mv} + \lambda_{cr} L_{cr}$$

Experimental Results

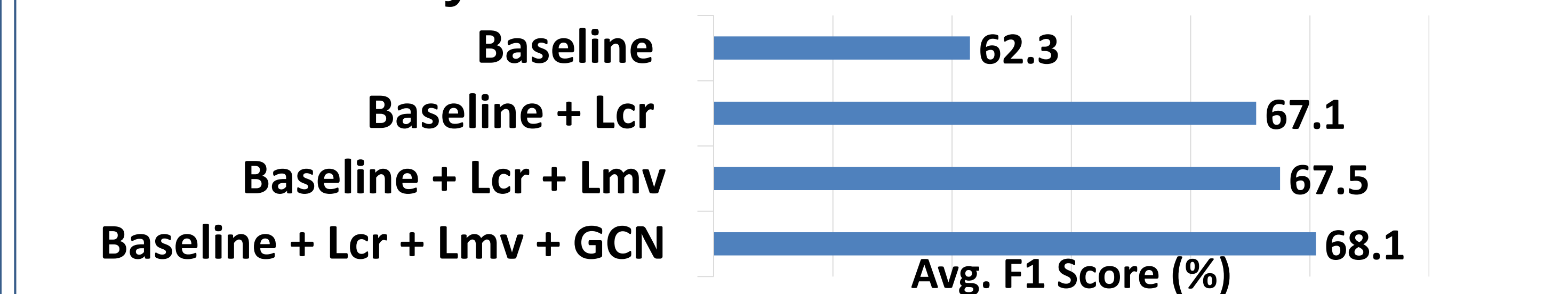
• Databases

	Type	No. labeled Imgs.	No. unlabeled Imgs.	No. AU	Protocol
EmotioNet	In-the-wild	20,722	50,000	12	Avg. of three random tests
BP4D	Spontaneous	~140,000	100,000	12	Subject-exclusive three-fold

• Results



• Ablation Study on EmotioNet



• Extension to Facial Attributes Analysis

