Weakly Supervised Facial Action Unit Recognition With Domain Knowledge

Shangfei Wang, *Senior Member, IEEE*, Guozhu Peng, Shiyu Chen, and Qiang Ji, *Fellow, IEEE*

***Abstract*—Current facial action unit (AU) recognition typically includes supervised training, where the fully AU annotated training images are required. Due to the nuances of facial appearance and individual differences, AU annotation is a time-consuming, expensive, and error-prone process. Facial expression is relatively simple to label, since facial expressions describe facial behavior globally and the number of expressions appearing on a face is much less than that of AUs. Furthermore, there exist strong dependencies between AUs and expressions, referred to as domain knowledge. Such domain knowledge is inherent in facial anatomy and facial behavior. Therefore, in this paper, we propose a novel weakly supervised AU recognition method to jointly learn multiple AU classifiers with expression annotations but without any AU annotations by leveraging domain knowledge. Specifically, we first summarize the expression-dependent AU ranking from the domain knowledge of conditional probabilities of AUs given expressions. Then, we formulate the weakly supervised AU recognition as a multilabel ranking problem and propose an efficient learning algorithm to solve it. Furthermore, we extend the proposed weakly supervised AU recognition method to a semi-supervised learning scenario when partial AU labeled samples are available. Experimental results on three benchmark databases demonstrate that the proposed method can successfully exploit domain knowledge for multiple AU recognition and, thus, outperforms both state-of-the-art weakly supervised AU recognition method and the semi-supervised AU recognition method.**

***Index Terms*—Domain knowledge, facial action unit (AU) recognition, weakly supervised learning.**

I. Introduction

AUTOMATICALLY facial action unit (AU) recognition has attracted increasing attention and achieved great progresses in recent years due to its wide application prospects in many fields, such as human–computer interactions. The main stream of current facial AU recognition includes supervised learning and, thus, requires fully AU annotated images for training. Facial AUs are very hard to recognize, since facial AUs describe the local and subtle changes on a face, and multiple facial AUs may appear on a face simultaneously. Therefore, the ground truth AUs must be labeled by qualified facial action coding system (FACS) experts.

Even with fully AU annotated training images, automatic facial AU recognition is still very challenge due to the richness, ambiguity, and the dynamic nature of facial actions. Recently, several works focus on AU dependencies to improve the AU classifier's performance through either generative approaches or discriminative approaches. For generative approaches, the structure and parameters of probabilistic graphic models, such as Bayesian Network (BN) [1], [2] and hierarchical restricted Boltzman machine (HRBM) [3], are used to exploit AU dependencies from AU labels. For discriminative approaches, the dependencies among AUs are embodied by introducing the constraints in the objective function of AU classifiers. For example, Zhu *et al.* [4] as well as Zhang and Mahoor [5] exploited related AU recognition tasks as multitask learning. Zhao *et al.* [6] selected a sparse subset of facial patches based on the group sparsity and local AU relations. Chu *et al.* [7] weighted training samples according to their similarity to unlabeled test data (STM). Eleftheriadis *et al.* [8] utilized AU relations as the regularization of latent space learning (MC-LVM). All of these works successfully exploit the AU dependencies to improve the performance of AU classifiers. However, all of them require complete AU annotations to learn AU classifiers.

Only very recently, a few works start focusing on AU recognition under partial AU annotations. Wang *et al.* [9] proposed using expression labels to complement the missing AU labels through a BN. Wu *et al.* [10] and Li *et al.* [11] adopted label consistency and smoothness as constraints to facilitate AU classifier learning from partial AU labels (MLML). Song *et al.* [12] proposed a Bayesian graphical model that encodes sparsity and co-occurrence structure of facial AUs via compressed sensing and group-wise sparsity inducing priors (BGCS). Their proposed methods can handle partially observed labels by marginalizing over the unobserved values as a part of the inference procedure. All of these works still require AU annotations to train AU classifiers, although AU annotations can partially be missed.

In general, AU annotation is more expensive and harder than expression annotation, since expression categories depict facial behaviors globally, while facial AUs describe the local variations on a face. Furthermore, the number of AUs for an

Manuscript received September 2, 2017; revised May 28, 2018 and July 17, 2018; accepted August 29, 2018. Date of publication September 26, 2018; date of current version October 15, 2018. This work was supported in part by the National Science Foundation of China under Grant 61473270, Grant 917418129, and Grant 61727809, and in part by the Project from Anhui Science and Technology Agency under Grant 1804a09020038. This paper was recommended by Associate Editor B. W. Schuller. *(Corresponding author: Shangfei Wang.)*

S. Wang, G. Peng, and S. Chen are with the Key Laboratory of Computing and Communication Software of Anhui Province, School of Computer Science and Technology, University of Science and Technology of China, Hefei 230027, China (e-mail: sfwang@ustc.edu.cn; gzpeng@mail.ustc.edu.cn; sy1001@mail.ustc.edu.cn).

Q. Ji is with the Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180 USA (e-mail: qji@ecse.rpi.edu).

Digital Object Identifier 10.1109/TCYB.2018.2868194

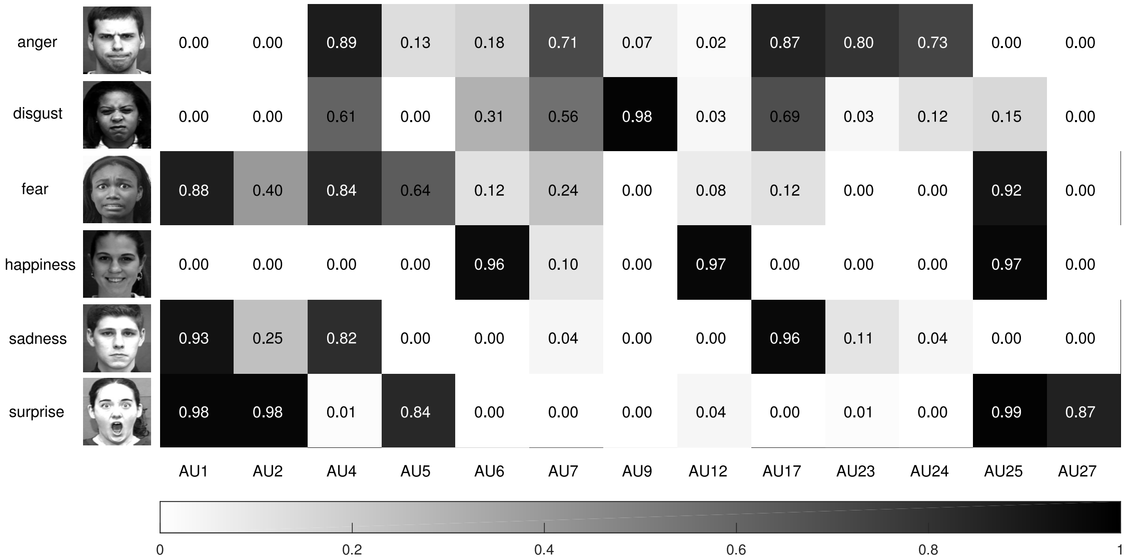

Fig. 1. Probabilities of the occurrence of AUs under six basic expressions on the CK+ database.

image is usually larger than that of expressions. Therefore, the effort for training human experts to score the AUs manually is expensive and time-consuming, while expressions are much easier to annotate, without the requirement of human experts. Large-scale expression-labeled Web images and scarce AU-labeled Web images provide further evidence that AU annotation is more expensive and harder than expression annotation.

Fortunately, expression categories and facial AUs are closely related, and most people express emotions using the same facial muscles. For instance, Du *et al.* [13] found out that in expressions of sadness, fear, and anger, people always lower their eyebrows. FACS lists emotion-related facial actions [14]. Such expression-dependent AU relations are inherent in facial anatomy and facial behavior. We refer to these relations as domain knowledge. The ground-truth AU labels of benchmark databases further confirm the strong dependencies between AUs and expressions. For example, Fig. 1 shows the probabilities of the occurrence of AUs under six basic expressions, which exist on the Extended Cohn–Kanade (CK+) database [15]. These probabilities are consistent with domain knowledge that is observed in behavior research. For example, on the CK+ database, more than 70% fear faces consist of AU1, AU4, and AU25, and less than 20% fear faces include AU6, AU9, AU12, AU17, AU23, AU24, and AU27. This is consistent with Du *et al.*'s work [13], as shown in Table I. Such inherent expression-AU dependencies can facilitate the training process of AU classifiers from facial images with expression annotations but without AU annotations.

Therefore, the goal of this paper is to learn AU classifiers from the domain knowledge of expression-AU dependencies using images with no AU annotation but with complete expression labels. To the best of our knowledge, only one work learns AU classifiers without AU annotations. Ruiz *et al.* [16] proposed hidden-task learning (HTL) to learn AU classifiers (i.e., hidden-tasks) from facial images without any annotations and extra large-scale facial images labeled with universal facial expressions (i.e., visible-tasks) through exploiting prior knowledge about the relation between expressions and AUs. Their proposed method learned both AU classifiers from images and expression classifiers from AUs. The exact AU probabilities for each expression are employed to learn expression classifiers first, then AU classifiers can be learned through embedding the output of AU classifiers as the input of expression classifiers. In addition, they extended HTL to semi-hidden task learning (SHTL) when partial AU annotated samples are provided. Unlike Ruiz *et al.*'s work, which requires both facial images without any annotations and extra large-scale facial images with basic expression annotations, we learn AU classifiers from facial images with expression labels directly, and do not collect further large-scale expression-annotated facial images. Furthermore, Ruiz *et al.*'s work learns both AU classifiers from images and expression classifiers from AUs, and the error caused by expression classifiers may propagate to the AU classifiers. Therefore, we tend to learn AU classifiers directly. Ruiz *et al.*'s work requires exact probabilities for single AUs given the expression. However, in the general case, the prior probabilities might be represented by inequalities rather than exact probabilities and some single AU probabilities are not even available. Thus, the expression-dependent AU ranking would be a better representation of domain knowledge than exact expression-dependent AU probabilities. In this paper, we first summarize expression-dependent AU probabilities from facial anatomy and behavior research as the domain knowledge. Then, we exploit the expression-dependent ranking order among AUs according to the summarized domain knowledge. After that, we formulate the weakly supervised AU recognition as a multilabel ranking problem and train AU classifiers through minimizing the rank loss. We also extend the weakly supervised AU recognition method to the semi-supervised AU recognition method when partial AU labeled data are available. We conduct within database experiments and cross-database experiments and make a comparison to state-of-the-art methods on three benchmark databases, that

TABLE I
PROBABILITIES ON AUs GIVEN EACH OF THE SIX BASIC EXPRESSIONS [13]

Expression \ AU	1	2	4	5	6	7	9	10	11	12	15	17	20	23	24	25	26	27
anger			≥ 0.7			≥ 0.7		0.26				0.52		≥ 0.7	≥ 0.7			
disgust			0.31				≥ 0.7	≥ 0.7							0.26			
fear	≥ 0.7	0.57	≥ 0.7	0.63									≥ 0.7			≥ 0.7	0.33	
happiness					0.51					≥ 0.7						≥ 0.7		
sadness	0.6		≥ 0.7		0.5				0.26		≥ 0.7	0.67						
surprise	≥ 0.7	≥ 0.7		0.66												≥ 0.7	≥ 0.7	

are, the CK+ database, the MMI database, and the UNBC-McMaster Shoulder Pain Expression Archive (McMaster) database. Experimental results demonstrate the superiority of the proposed method in automatically recognizing local facial AUs which are present in both basic and nonbasic expressions.

II. Problem Statement

Traditional supervised learning needs fully annotated target label samples during training. However, in many applications, the target labels are difficult to obtain, and the auxiliary labels may be easy to collect. Therefore, our goal is to learn a weakly supervised classifier by considering the relationship between the target labels and the auxiliary labels.

Let $S = \{(x_n, z_n), n = 1, \ldots, N\}$ denote training samples where $x_n \in R^D$ represents the D-dimensional feature vector, $z_n \in \{1, 2, \ldots, K\}$ represents the auxiliary label which is easy to obtain, K is the number of label classes, and N is the number of training samples. $Y_n = \{y_n^1, y_n^2, \ldots, y_n^Q\} \in \{0, 1\}^Q$ indicates the target labels which are unknown and Q is the number of labels. Δ represents the relationship between the target labels and the auxiliary labels. The objective of the proposed method is to learn a classifier $f(x)$ according to

$$\min \frac{1}{N}\sum_{n=1}^{N} L_{\text{unlab}}(f(x_n), z_n, \Delta) \tag{1}$$

where $L_{\text{unlab}}(f(x_n), z_n, \Delta)$ indicates the loss function that maps features to target labels through the known auxiliary labels and relationship between the target labels and the auxiliary labels.

We extend the weakly supervised method to the semi-supervised method when partial data are annotated by target labels. Similar to weakly labeled training samples S, we denote $T = \{(x_m, Y_m), m = 1, \ldots, M\}$ as fully labeled training samples where $x_m \in R^D$ is the D-dimensional feature vector, $Y_m = \{y_m^1, y_m^2, \ldots, y_m^Q\} \in \{0, 1\}^Q$ indicates the target labels, Q is the number of labels, and M is the number of training samples. Given the training data S and T, the objective of the semi-supervised method is to learn a classifier $f(x)$ according to

$$\begin{aligned} \min\ & (1-\alpha)\frac{1}{N}\sum_{n=1}^{N} L_{\text{unlab}}(f(x_n), z_n, \Delta) \\ & + \alpha\frac{1}{M}\sum_{m=1}^{M} L_{\text{lab}}(f(x_m), Y_m) \end{aligned} \tag{2}$$

where $L_{\text{unlab}}(f(x_n), z_n, \Delta)$ indicates the loss function over the weakly labeled training samples, $L_{\text{lab}}(f(x_m), Y_m)$ indicates the loss function that maps the features to target labels, and α is the coefficient. $\alpha \in [0,\ 1]$ controls the tradeoff between the minimization of the weakly labeled and fully labeled losses. Specifically, when $\alpha = 0$ the optimization problem is the same as the weakly supervised learning problem as mentioned in (1). When $\alpha = 1$, the optimization problem actually becomes a traditional supervised learning problem. In fact, any loss function can be used in (2).

In this paper, we evaluate the proposed method on multiple AU recognition problem. The expression labels are auxiliary labels and the multiple AU labels are target labels.

III. Methodology

In this section, we first summarize the relationship between AUs and expressions, i.e., domain knowledge, from behavior research, and then infer the expression-dependent AU ranking. After that, we describe the proposed weakly supervised AU recognition method, and further extend it to the semi-supervised AU recognition method.

A. Domain Knowledge

In this section, we summarize the domain knowledge for both basic expressions and nonbasic expressions. For basic expressions, the relations between AUs and six basic expressions, i.e., happiness, sadness, anger, surprise, disgust, and fear are reviewed. For nonbasic expressions, the relations between AUs and pain expressions are presented.

For the relations between AUs and six basic expressions, Du *et al.* [13] described the prototypical AUs observed in each compound expression and basic expression. The probabilities on AUs given each of the six basic expressions, i.e., anger, disgust, fear, happiness, sadness, and surprise, are shown in Table I. The blanks indicate that these probabilities on AUs are less than 20%. For example, given happiness, the probabilities of the occurrence of AU12 and AU25 are more than 70%, the probability of the occurrence of AU6 is 51%, and the probabilities of the occurrence of other AUs are less than 20%. Therefore, AU12 and AU25 have higher rankings than AU6, and AU6 has a higher ranking than other AUs. From Table I, we can obtain the expression-dependent pairwise AU ranking according to these probabilities.

Furthermore, Friesen and Ekman [14] defined a coding system named EMFACS which is given by a set of AUs. Each AU codes the fundamental actions of individuals or groups of

TABLE II
MOST FREQUENT AU COMBINATIONS GIVEN EACH OF THE SIX BASIC EXPRESSIONS [14]

Expression	AUs
anger	4+5,4+7,4+5+7,17+24,23
disgust	9,10
fear	1+2+4,20
happiness	12,6+12,7+12,6+7+12
sadness	1,1+4,15,6+15,11+15,11+17
surprise	1+2+5,1+2+26,1+2+5+26

TABLE III
PROBABILITIES ON AUS GIVEN PAIN AND NEUTRAL EXPRESSION [17]

4	6	7	9	10	12	20	25	26	43
3.2%	4.5%	5.4%	2.2%	1.6%	5.1%	0.7%	2.1%	2.9%	2.8%

muscles typically seen while producing the facial expressions of emotion. The most frequent AU combinations given to each of the six basic expressions are shown in Table II. We can find that the information obtained from Table II is mostly included in Table I. For example, AU4+AU5+AU7, AU17+AU24, and AU23 are the most frequent AU combinations given anger from Table II. Meanwhile AU4, AU5, AU7, AU17, AU23, and AU24 have higher rankings than others in Table I. This further indicates the reliability of Table I. Therefore, we only adopt the expression-dependent AU ranking inferred from Table I in the following sections.

For the relations between AUs and pain expression, Prkachin's study [17] provided the percentage of AUs coded in the entire dataset, including facial images during both painful and pain-free periods, as shown in Table III. We assume that no AUs appear on faces during pain-free period, and thus infer pairwise AU rankings under pain expression from Table III as follows: AU6, AU7, and AU12 have higher rankings than AU4, AU25, AU26, and AU43. Meanwhile, AU4, AU25, AU26, and AU43 have higher rankings than AU9, AU10, and AU20.

In summary, Tables I and III list the conditional probabilities of the occurrence of AU given expression. The AUs with higher probabilities of occurrence have higher rankings than those with lower probabilities of occurrence. We adopt the pairwise AU ranking from the domain knowledge as constraints. Specifically, we consider the pairwise ranking orders of AUs whose probabilities have exact orders. Take the anger expression as an example, the probabilities of AU4, AU7, AU23, and AU24 are all larger than 0.7, the probability of AU10 is 0.26, the probability of AU17 is 0.52, and the probabilities of AU1, AU2, AU5, and AU6 are all less than 0.2. Therefore, we consider the orders of AU1 and AU4; AU1 and AU7; AU1 and AU10; AU1 and AU17; AU1 and AU23; AU1 and AU24; AU2 and AU4; AU2 and AU7; AU2 and AU10; AU2 and AU17; AU2 and AU23; AU2 and AU24; AU4 and AU5; AU4 and AU6; AU4 and AU10; AU4 and AU17; AU5 and AU7; AU5 and AU10; AU5 and AU17; AU5 and AU23; AU5 and AU24; AU6 and AU7; AU6 and AU10; AU6 and AU17; AU6 and AU23; AU6 and AU24; AU7 and AU10; AU7 and AU17; AU10 and AU17; AU10 and AU23; AU10 and AU24; AU17 and AU23; and AU17 and AU24. We do not consider the orders among AU4, AU7, AU23, and AU24 and the orders among AU1, AU2, AU5, and AU6, since we cannot infer the rankings from their possibilities.

B. Proposed Method

For AU recognition, feature vector x_n in one sample (x_n, z_n) from S are image features, and auxiliary label z_n is the expression label. In this case, $K = 6$ when we consider six basic expressions, and $K = 1$ when we consider nonbasic expressions (i.e., pain expression). Target labels Y_n are AU labels which are unknown.

Label ranking is one of the most common approaches to solve multilabel classification problems. Given the expression and the expression-dependent AU rankings from domain knowledge, we introduce our method subject to these rankings. Rank loss imposes a penalty on a classifier when a pair of labels is incorrectly ranked. We adopt rank loss as our loss function [18], [19].

If the probability of each AU presence given expression is known, we consider the pairwise ranking orders according to the probabilities of the occurrence of the AUs. The rank loss is as follows:

$$L_{\text{unlab}}(x, z) = \frac{1}{N}\sum_{n=1}^{N}\sum_{i,j:P(y_n^i=1|z_n)>P\left(y_n^j=1|z_n\right)} \left[\left[f^i(x_n) < f^j(x_n)\right]\right] + \frac{1}{2}\left[\left[f^i(x_n) = f^j(x_n)\right]\right] \quad (3)$$

where i and j represent the indexes of the AU labels, $f^i(x_n)$ and $f^j(x_n)$ are the output value for ith and jth AU labels, and $P(y_n^i = 1|z_n)$ indicates the probability of the occurrence of the ith AU given the expression z_n. $[[\cdot]]$ is the indicator function that has a value of 1 when the conditions inside the brackets are met; otherwise, it is 0. The first term and the second term represent the loss when the predicted value $f^i(x_n)$ is smaller than $f^j(x_n)$ and the predicted value $f^i(x_n)$ is equal to $f^j(x_n)$, respectively, when the probability of the occurrence of the ith AU label is larger than that of the jth AU label.

We rewrite the rank loss by the expectation of the 0-1 function over the sample space

$$L_{\text{unlab}}(x, z) = \frac{1}{N}\sum_{n=1}^{N}\sum_{i,j:P(y_n^i=1|z_n)>P\left(y_n^j=1|z_n\right)} \left[l_{0-1}\left(f^i(x_n) - f^j(x_n)\right)\right] \quad (4)$$

and the 0-1 function can be represented as

$$l_{0-1}\left(f^i(x_n) - f^j(x_n)\right) = \begin{cases} 1 & f^i(x_n) - f^j(x_n) < 0 \\ \frac{1}{2} & f^i(x_n) - f^j(x_n) = 0 \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

Direct optimization of this expectation including the 0-1 loss is intractable. For better model learning, the surrogate loss function $l'(x)$ is generally used instead of the 0-1 loss. In this paper, we use the sigmoid loss as the surrogate loss. By applying the surrogate loss to rank loss, we obtain

$$L_{\text{unlab}}(x,z) = \frac{1}{N}\sum_{n=1}^{N} \sum_{i,j:P(y_n^i=1|z_n)>P\left(y_n^j=1|z_n\right)} \left[l_s\left(f^i(x_n)-f^j(x_n)\right)\right] \tag{6}$$

and the function l_s can be represented as

$$l_s\left(f^i(x_n)-f^j(x_n)\right) = \frac{1}{1+e^{\left(f^i(x_n)-f^j(x_n)\right)}}. \tag{7}$$

The proposed method utilizes the domain knowledge and the known expression label to learn weakly supervised classifiers to recognize AUs without AU labels. We use linear function $f(x) = wx$ as our score function. Now, our goal is to minimize the ranking loss L_{unlab} and obtain the weight w. The gradient descent approach is used to solve the problem

$$w^{(t+1)} = w^{(t)} - \eta^{(t)} \frac{\partial L_{\text{unlab}}(x,z)}{\partial w} \tag{8}$$

where t and η indicate the number of iterations and the learning rate.

The gradient of rank loss function to the weight can be computed as

$$\begin{aligned}\frac{\partial L_{\text{unlab}}(x,z)}{\partial w} &= \frac{1}{N}\sum_{n=1}^{N}\frac{\partial L_{\text{unlab}}(x_n,z_n)}{\partial f(x_n)} * \frac{\partial f(x_n)}{\partial w}\\ &= \frac{1}{N}\sum_{n=1}^{N}\frac{\partial L_{\text{unlab}}(x_n,z_n)}{\partial f(x_n)} * x_n^T \end{aligned} \tag{9}$$

where the specific gradient of rank loss function to the score function of each component can be computed as

$$\begin{aligned}\frac{\partial L_{\text{unlab}}(x_n,z_n)}{\partial f^i(x_n)} = &\sum_{j:P(y_n^i=1|z_n)>P\left(y_n^j=1|z_n\right)} \frac{\partial l_s(x)}{\partial x}\Big|_{x=f^i(x_n)-f^j(x_n)}\\ &- \sum_{j:P\left(y_n^j=1|z_n\right)>P(y_n^i=1|z_n)} \frac{\partial l_s(x)}{\partial x}\Big|_{x=f^j(x_n)-f^i(x_n)}.\end{aligned} \tag{10}$$

C. Extension to Semi-Supervised Learning

We extend the proposed method to the semi-supervised AU recognition method which can solve the problems with partial AU labels annotated data. In a semi-supervised scenario, in addition to training samples S, T contains M fully AU-labeled training samples. Target labels Y_m are multiple AU labels.

Here, we also use rank loss as the loss function. Since the AU labels are known, we consider the ranking order between relevant AU labels and irrelevant AU labels. The rank loss is as follows:

$$L_{\text{lab}}(x,Y) = \frac{1}{M}\sum_{m=1}^{M}\sum_{i,j:y_m^i=1,y_m^j=0}\left[l_s\left(f^i(x_m)-f^j(x_m)\right)\right]. \tag{11}$$

```
Algorithm 1 Training Algorithm for the Proposed Method
Input:
    weakly labelled training samples $(x_n, z_n)$,
    fully labelled training samples $(x_m, Y_m)$,
    coefficient $\alpha$, learning rate $\eta$,
    domain knowledge
Output:
    optimized parameter $w$
Randomly initialize $w$;
repeat
    for each weakly labelled training sample $(x_n, z_n)$ do
        for each label $i$ do
            Calculate $\frac{\partial L_{unlab}(x_n,z_n)}{\partial f_i(x_n)}$ as Eq. (10);
        end for
    end for
    Calculate $\frac{\partial L_{unlab}(x,z)}{\partial w}$ as Eq. (9);
    for each fully labelled training sample $(x_m, Y_m)$ do
        for each label $i$ do
            Calculate $\frac{\partial L_{lab}(x_m,Y_m)}{\partial f_i(x_n)}$ as Eq. (14);
        end for
    end for
    Calculate $\frac{\partial L_{lab}(x,Y)}{\partial w}$ as Eq. (13);
    $w \leftarrow w - \eta((1-\alpha)\frac{\partial L_{unlab}(x,z)}{\partial w} + \alpha\frac{\partial L_{lab}(x,Y)}{\partial w})$;
until Converges
Return $w$.
```

The optimization problem can be defined as

$$\min_{w}(1-\alpha)L_{\text{unlab}}(x,z) + \alpha L_{\text{lab}}(x,Y). \tag{12}$$

In (12), the first term $L_{\text{unlab}}(x,z)$ represents the rank loss function over the weakly labeled training samples, and the second term $L_{\text{lab}}(x,Y)$ represents the rank loss function over the fully labeled training samples. $\alpha \in [0,1]$ controls the trade-off between the minimization of the weakly labeled and fully labeled rank losses.

Similarly, the gradient descent approach is used to solve the above optimization problem. The gradient of weakly labeled rank loss function to the weight is computed as (9). Now, we compute the gradient of the fully labeled rank loss function $L_{\text{lab}}(x,Y)$ as follows:

$$\frac{\partial L_{\text{lab}}(x,Y)}{\partial w} = \frac{1}{M}\sum_{m=1}^{M}\frac{\partial L_{\text{lab}}(x_m,Y_m)}{\partial f(x_m)} * x_m^T. \tag{13}$$

The specific gradient of labeled rank loss function to the score function of each component can be computed as

$$\begin{aligned}\frac{\partial L_{\text{lab}}(x_m,Y_m)}{\partial f^i(x_m)} = &\sum_{j:y_m^i=1,y_m^j=0}\frac{\partial l_s(x)}{\partial x}\Big|_{x=f^i(x_m)-f^j(x_m)}\\ &- \sum_{j:y_m^j=1,y_m^i=0}\frac{\partial l_s(x)}{\partial x}\Big|_{x=f^j(x_m)-f^i(x_m)}.\end{aligned} \tag{14}$$

The detailed learning algorithm is shown in Algorithm 1.

D. Classification

During the testing phase, the predicted values are computed according to the image features and optimized w. Since we adopt rank loss during training, we need to obtain the label set size, i.e., the number of the relevant AU labels. As shown in Table I, there are up to five AUs appeared with more than 50% probabilities among available AUs given basic expressions. Therefore, the label size is set as 5. We assign labels to the testing samples if the corresponding predicted value is among the top five values.

IV. Experiments

A. Experimental Conditions

To evaluate the performance of the proposed method, we conduct experiments on three benchmark databases, i.e., the CK+ database [15], the MMI database [20], and the McMaster database [21].

The CK+ database [15] includes 593 sequences from 123 subjects and the image sequence incorporates the onset to peak formation of the facial expressions. Among them, 309 sequences from 106 subjects annotated by six basic expressions and AUs are used. AUs available for more than 10% of all samples are chosen. Thus, we obtain 309 sequences with 13 labels, i.e., AU1, AU2, AU4, AU5, AU6, AU7, AU9, AU12, AU17, AU23, AU24, AU25, and AU27.

The MMI database [20] consists of over 2900 videos and images from 75 subjects. Like the CK+ database, we choose the sequences annotated by six basic expressions and AUs, and the AUs available for more than 10% of all samples. Thus, we obtain 171 sequences from 27 subjects with 13 labels, i.e., AU1, AU2, AU4, AU5, AU6, AU7, AU9, AU10, AU12, AU17, AU23, AU25, and AU26.

The McMaster database [21] contains 200 video sequences of patients suffering from chronic shoulder pain while preforming a range of arm motion tests. From these 200 sequences, there is a total of 48 398 frames that have been FACS coded and AAM tracked. We choose the 8369 frames in pain from 24 subjects with ten labels, i.e., AU4, AU6, AU7, AU9, AU10, AU12, AU20, AU25, AU26, and AU43.

For features, 49 inner facial landmarks from the exaggerated frames are used. On the CK+ database and the McMaster database, the facial landmarks are provided by the database. On the MMI database, the facial landmarks are extracted with IntraFace [22]. The F1-measure is used as a performance metric, and five-fold subject independent cross validation is adopted.

For both weakly supervised and semi-supervised AU recognition, two experiments, i.e., within database experiments and cross-database experiments, are conducted. For weakly supervised AU recognition, only image features and expression labels are used. We compare the proposed weakly supervised AU recognition method with HTL [16], the only state-of-the-art method that learns an AU classifier without AU annotation. For the CK+ database, and the McMaster database, we directly compare with the experimental results listed in [16]. For the MMI database, we conduct AU recognition with the implementation of the HTL method, since [16] does not provide experimental results on the MMI database.

For semi-supervised AU recognition, image features, expression labels, and partial available AU labels are used. We randomly miss the AU labels with a certain proportion (50%), and conduct the experiments ten times. The averaged F1-measure and standard deviation of the F1-measure are used as the performance metrics. We compare the proposed semi-supervised AU recognition method with several related works as mentioned in Section I, i.e., SHTL [16], BN [9], MLML [10], and BGCS [12]. These works do not provide the results on the MMI database and the McMaster database. Although, they conduct experiments on the CK+ database, we cannot compare them directly because of the following reasons. Ruiz *et al.*'s work [16] used leave-one-subject-out cross validation in the semi-supervised AU recognition experiments. It means that the AU classifiers are trained with all the data from the large-scale expression-annotated database and the AU-annotated database. In our experiments, 50% training

TABLE IV
Within Database Experimental Results in Term of F1-Measure of Weakly Supervised AU Recognition

AU	CK+		MMI		McMaster	
	HTL[16]	Ours	HTL[16]	Ours	HTL[16]	Ours
1	.619	**.845**	.594	**.685**	-	-
2	.451	**.856**	.640	**.727**	-	-
4	**.816**	.676	.305	**.644**	.083	**.161**
5	**.748**	.747	.605	**.694**	-	-
6	.587	**.614**	**.388**	.381	.295	**.681**
7	.327	**.640**	.259	**.500**	.172	**.398**
9	.487	**.747**	.353	**.488**	.053	**.083**
10	-	-	.359	**.364**	.085	.013
12	.852	**.910**	.549	**.736**	.426	**.672**
17	.675	**.858**	.288	**.484**	-	-
20	-	-	-	-	.015	**.082**
23	-	.704	**.245**	.109	-	-
24	-	.471	-	-	-	-
25	.701	**.945**	.663	**.724**	.124	**.202**
26	-	-	**.566**	.467	.150	**.214**
27	-	.400	-	-	-	-
43	-	-	-	-	-	.390
Avg.	.626	**.724**	.447	**.539**	.156	**.290**

TABLE V
Cross-Database Experimental Results in Term of F1-Measure of Weakly Supervised AU Recognition on the CK+ Database and the MMI Database

AU	From CK+ to MMI		From MMI to CK+	
	HTL[16]	Ours	HTL[16]	Ours
1	.594	**.667**	.619	**.700**
2	.640	**.649**	.451	**.802**
4	.305	**.578**	**.816**	.650
5	.605	**.629**	**.748**	.707
6	**.388**	.366	**.587**	.506
7	.259	**.471**	.327	**.552**
9	**.353**	.271	**.487**	.455
12	.549	**.598**	**.852**	.745
17	.288	**.391**	**.675**	.662
23	**.245**	.068	-	.419
25	.663	**.759**	.701	**.870**
Avg.	.444	**.495**	.626	**.643**

TABLE VI
CONFUSION MATRIX OF EACH AU GIVEN ANGER EXPRESSION ON THE CK+ DATABASE

AU		AU1		AU2		AU4		AU5		AU6		AU7		AU9		AU12		AU17		AU23		AU24		AU25		AU27	
		0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
confusion	0	39	6	45	0	0	5	39	0	29	8	1	12	30	12	44	0	0	6	9	0	2	10	45	0	45	0
matrix	1	0	0	0	0	0	40	6	0	8	0	1	31	1	2	1	0	0	39	14	22	1	32	0	0	0	0

samples without AU labels. Wang *et al.* [9] conducted semi-supervised AU recognition experiments by missing each AU annotation by a certain proportion. It means that all training images have AU labels, for each image, the AU labels may not be complete. Such experimental conditions are different from ours, where 50% of the training samples are without AU labels. Wu *et al.* [10] used 20%, 40%, 60%, and 80% as the missing proportions and adopted the average precision and area under the ROC curve as performance metrics. Their missing proportions and performance metrics are different from ours. Song *et al.* [12] only provided the averaged F1-measure of 24 AUs under missing 50% labels, not the specific results of each AU. Since the experimental conditions of these related works are all different, we cannot compare them directly. We reconduct these experiments using the provided codes to make a fair comparison.

For both weakly supervised and semi-supervised AU recognition, we do not conduct the cross-database experiments between the databases with six basic expressions (i.e., the CK+ database and the MMI database) and the McMaster database, since the relations between AUs and six basic expressions are different from the relations between AUs and pain expressions.

Furthermore, to demonstrate the practicality of the proposed weakly supervised AU recognition method, we compare the performance of the proposed weakly supervised AU recognition to the state-of-the-art supervised AU recognition, which requires fully AU annotated training samples.

B. Experimental Results of Weakly Supervised AU Recognition

The weakly supervised AU recognition results of the within database and the cross-database experiments are shown in Tables IV and V, respectively.

From Tables IV and V, we can find that the proposed method outperforms HTL on both databases not only for within database AU recognition but also for cross-database AU recognition in most cases. Specifically, for the within database AU recognition, the averaged F1-measures of common AUs of the proposed method are 15.2%, 9.2%, and 12.2% higher than HTL on the CK+ database, the MMI database and the McMaster database, respectively. For specific AUs, the F1-measure of the proposed method is higher than HTL on 8 out of 10 AUs on the CK+ database and the improvements of AU1, AU2, AU7, AU9, and AU25 are more than 20%. The F1-measure of the proposed method is higher than HTL on 10 out of 13 AUs on the MMI database and 8 out of 9 AUs on the McMaster database.

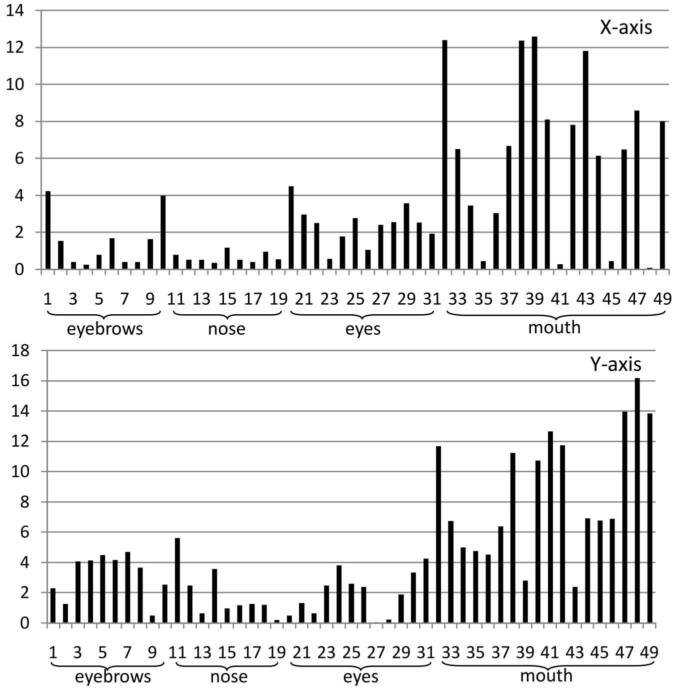

Fig. 2. Absolute values of weight coefficients for 49 points on the *x*-axis and *y*-axis for AU25 on the CK+ database.

For the cross-database AU recognition, the proposed method improves by 5.1% more than HTL in terms of the averaged F1-measure when training on the CK+ database and testing on the MMI database. The improvements of AU4 and AU7 are 27.3% and 21.2%. For the experiments of training on the MMI database and testing on the CK+ database, the proposed method has the best results on AU1, AU2, AU7, and AU25, and achieves better results than HTL in terms of averaged F1-measure.

Both HTL and the proposed method leverage the relationship between AUs and expressions to facilitate AU recognition without AU annotated training samples. However, unlike HTL, which needs exact probabilities from single AUs given expressions and learns AU classifiers indirectly by training both AU classifiers and expression classifiers using additional samples labeled with expressions, the proposed method directly learns AU classifiers from the inferred expression-dependent AU rankings. Thus, the proposed method avoids the error propagated from expression classifiers, and can capture the dependencies between expressions and AUs more efficiently. This results in better performances on both within database AU recognition and cross-database AU recognition. The experimental results demonstrate the superiority and generalization ability of the proposed method.

TABLE VII
WITHIN DATABASE EXPERIMENTAL RESULTS IN TERM OF F1-MEASURE OF SEMI-SUPERVISED AU RECOGNITION. (a) F1-MEASURE ON THE CK+ DATABASE. (b) F1-MEASURE ON THE MMI DATABASE. (c) F1-MEASURE ON THE MCMASTER DATABASE

(a)

AU	BN[9]	BGCS[12]	MLML[10]	SHTL[16]	Ours
1	.918 (.013)	.854 (.020)	.787 (.018)	.804 (.008)	.872 (.010)
2	.892 (.019)	.891 (.017)	.804 (.016)	.797 (.010)	.875 (.014)
4	.827 (.017)	.197 (.186)	.634 (.020)	.492 (.019)	.765 (.016)
5	.818 (.022)	.815 (.020)	.758 (.024)	.776 (.007)	.760 (.019)
6	.704 (.016)	.769 (.019)	.591 (.024)	.645 (.009)	.711 (.021)
7	.578 (.021)	.658 (.029)	.519 (.048)	.521 (.016)	.580 (.022)
9	.563 (.025)	.830 (.040)	.402 (.060)	.462 (.016)	.757 (.015)
12	.749 (.021)	.867 (.016)	.462 (.084)	.583 (.006)	**.913 (.019)**
17	.824 (.013)	.823 (.018)	.656 (.021)	.667 (.018)	**.847 (.010)**
23	.406 (.013)	.599 (.048)	.124 (.088)	.105 (.014)	**.683 (.046)**
24	.385 (.015)	.558 (.041)	.076 (.059)	.253 (.017)	.136 (.046)
25	.923 (.007)	.786 (.032)	.795 (.018)	.879 (.006)	**.949 (.005)**
27	.878 (.010)	.922 (.012)	.568 (.013)	.506 (.007)	**.926 (.014)**
Avg.	.728 (.006)	.736 (.010)	.568 (.013)	.596 (.003)	**.781 (.007)**

(b)

AU	BN[9]	BGCS[12]	MLML[10]	SHTL[16]	Ours
1	.634 (.032)	.490 (.075)	.673 (.026)	.658 (.018)	**.740 (.012)**
2	.685 (.036)	.565 (.053)	.700 (.031)	.734 (.016)	.715 (.022)
4	.684 (.029)	.570 (.048)	.662 (.037)	.664 (.021)	.661 (.023)
5	.660 (.035)	.602 (.038)	.676 (.032)	.622 (.019)	**.691 (.027)**
6	.342 (.030)	.421 (.045)	.457 (.085)	.420 (.011)	**.492 (.046)**
7	.439 (.043)	.424 (.064)	.482 (.052)	.460 (.028)	.466 (.042)
9	.328 (.046)	.504 (.061)	.406 (.085)	.393 (.032)	.491 (.024)
10	.290 (.036)	.336 (.061)	.300 (.082)	.399 (.019)	**.404 (.028)**
12	.530 (.016)	.672 (.034)	.646 (.040)	.577 (.017)	**.744 (.026)**
17	.400 (.021)	.410 (.042)	.322 (.072)	.395 (.031)	**.501 (.030)**
23	.225 (.026)	.211 (.053)	.046 (.031)	.253 (.017)	.136 (.046)
25	.694 (.042)	.396 (.031)	.857 (.020)	.724 (.009)	.759 (.017)
26	.630 (.030)	.370 (.083)	.655 (.033)	.615 (.027)	.635 (.028)
Avg.	.503 (.011)	.459 (.017)	.529 (.015)	.532 (.006)	**.572 (.008)**

(c)

AU	BN[9]	BGCS[12]	MLML[10]	SHTL[16]	Ours
4	.291 (.115)	.301 (.177)	.261 (.133)	.342 (.003)	.312 (.062)
6	.733 (.034)	.722 (.079)	.748 (.089)	.688 (.003)	**.786 (.014)**
7	.395 (.050)	.417 (.161)	.512 (.156)	.396 (.007)	**.545 (.031)**
9	.109 (.052)	.101 (.098)	.015 (.032)	.128 (.006)	.053 (.035)
10	.322 (.146)	.135 (.174)	.198 (.185)	.186 (.006)	.060 (.082)
12	.748 (.034)	.753 (.098)	.725 (.071)	.655 (.003)	.752 (.011)
20	.034 (.037)	.097 (.061)	.010 (.017)	.079 (.007)	.023 (.021)
25	.357 (.063)	.318 (.112)	.367 (.155)	.366 (.012)	.291 (.042)
26	.005 (.007)	.158 (.130)	.081 (.059)	.156 (.006)	.103 (.022)
43	.655 (.064)	.320 (.234)	.479 (.172)	-	.625 (.037)
Avg.	.333 (.031)	.332 (.014)	.340 (.046)	.333 (.006)	**.355 (.011)**

To further demonstrate the effectiveness of the proposed method in capturing the expression-dependent AU ranking, we analyze the confusion matrix of each AU given expression. Take the anger expression on the CK+ database as an example, as shown in Table VI, AU1, AU2, AU5, AU6, AU9, AU12, AU25, and AU27 are always classified as 0, and AU4, AU7,

TABLE VIII

CROSS-DATABASE EXPERIMENTAL RESULTS IN TERM OF F1-MEASURE OF SEMI-SUPERVISED AU RECOGNITION ON THE CK+ DATABASE AND THE MMI DATABASE. (a) TRAINING ON THE CK+ DATABASE AND TESTING ON THE MMI DATABASE. (b) TRAINING ON THE MMI DATABASE AND TESTING ON THE CK+ DATABASE

(a)

AU	BN[9]	BGCS[12]	MLML[10]	SHTL[16]	Ours
1	.705 (.048)	.361 (.087)	.661 (.010)	.563 (.035)	.669 (.015)
2	.710 (.066)	.506 (.130)	.665 (.013)	.641 (.022)	.625 (.020)
4	.442 (.048)	.542 (.052)	.576 (.027)	.509 (.075)	.561 (.027)
5	.543 (.055)	.522 (.118)	.622 (.021)	.595 (.015)	.578 (.033)
6	.444 (.035)	.313 (.033)	.312 (.041)	.391 (.005)	.404 (.018)
7	.397 (.044)	.419 (.039)	.517 (.026)	.373 (.053)	**.528 (.037)**
9	.363 (.067)	.304 (.088)	.301 (.083)	.311 (.051)	.353 (.018)
12	.462 (.088)	.347 (.065)	.600 (.062)	.550 (.007)	**.630 (.009)**
17	.311 (.163)	.384 (.034)	.329 (.039)	.363 (.032)	**.395 (.025)**
23	.194 (.041)	.228 (.061)	.006 (.020)	.216 (.025)	.091 (.050)
25	.533 (.042)	.636 (.092)	.770 (.014)	.711 (.024)	.743 (.018)
Avg.	.464 (.024)	.415 (.024)	.487 (.129)	.475 (.012)	**.507 (.009)**

(b)

AU	BN[9]	BGCS[12]	MLML[10]	SHTL[16]	Ours
1	.669 (.034)	.564 (.098)	.653 (.072)	.755 (.015)	.705 (.073)
2	.597 (.024)	.546 (.113)	.614 (.045)	.786 (.011)	.647 (.059)
4	.596 (.050)	.461 (.088)	.562 (.022)	.427 (.009)	.547 (.031)
5	.617 (.041)	.566 (.087)	.544 (.033)	.780 (.004)	.621 (.056)
6	.388 (.042)	.444 (.050)	.201 (.156)	.510 (.004)	.474 (.054)
7	.477 (.032)	.371 (.084)	.478 (.033)	.450 (.009)	.392 (.035)
9	.349 (.042)	.260 (.078)	.297 (.046)	.356 (.017)	**.379 (.051)**
12	.590 (.017)	.367 (.049)	.432 (.141)	.436 (.003)	.459 (.081)
17	.402 (.041)	.391 (.102)	.381 (.154)	.301 (.012)	**.569 (.027)**
23	.219 (.022)	.207 (.082)	.045 (.055)	.138 (.005)	**.230 (.049)**
25	.677 (.056)	.533 (.089)	.736 (.022)	.719 (.014)	.710 (.027)
Avg.	.507 (.010)	.428 (.024)	.449 (.032)	.514 (.004)	**.521 (.017)**

AU17, AU23, and AU24 are always classified as 1. These are consistent with the information in Table I. AU4, AU7, AU17, AU23, and AU24 have higher rankings than other AUs. This indicates that such expression-dependent AU rankings are successfully captured by the proposed method and result in better performance.

Since each AU is related to different local features, it may be beneficial to adopt different features for different AUs for better recognition performance. A common feature selection method is to employ the supervised feature selecting method, such as linear discriminant analysis. However, in this paper, we propose a weakly supervised AU recognition method, which does not require AU labels during training. Therefore, we use all 49 inner facial landmarks as features for all target AUs, and cannot explicitly select features for each AU. Through weakly supervised learning, the weight coefficients of the linear classifier can trade off the feature importance for each AU during training. Take AU25 on the CK+ database as an example, we analyze the learned weight coefficients of facial landmarks. The absolute values of weight coefficients for AU25 are shown in Fig. 2. From Fig. 2, we find that the points in the mouth region have larger weight coefficients than the points around the eyes, the eyebrows, and the nose. These indicate that the weight coefficients can represent the feature importance for each AUs. Therefore, it is reasonable to use all features for all target AUs in a weakly supervised learning scenario.

C. Experimental Results of Semi-Supervised AU Recognition

Tables VII and VIII show the within database results and the cross-database results of semi-supervised AU recognition, respectively.

From Tables VII and VIII, we can find that under the semi-supervised scenario, the proposed method performs best on all databases not only for within database AU recognition but also for cross-database AU recognition. Specifically, for within database experiments, the proposed method achieves better performance than other methods in terms of the averaged F1-measure on all databases. Furthermore, the proposed method achieves the best results on several AUs, such as

TABLE IX
COMPARISON TO THE STATE-OF-THE-ART WITH AU ANNOTATIONS IN TERM OF F1-MEASURE ON THE CK+ DATABASE

AU	STM[7]	HRBM[3]	MC-LVM[8]	Ours
1	.622	**.869**	.825	.845
2	.762	.855	**.870**	.856
4	.691	.726	**.792**	.676
5	-	.720	.735	**.747**
6	**.796**	.617	.728	.614
7	**.791**	.545	.575	.640
9	-	.859	**.879**	.747
12	.772	.727	.876	**.910**
17	.743	.817	**.868**	.858
23	-	.566	.673	**.704**
24	-	.353	**.510**	.471
25	-	.926	.918	**.945**
27	-	.877	**.911**	.400
Avg.	.740	.727	**.782**	.724

TABLE X
COMPARISON TO THE STATE-OF-THE-ART WITH AU ANNOTATIONS IN TERM OF F1-MEASURE ON THE MMI DATABASE

AU	SVM-HMM[23]	FFD[24]	Ours
1	.585	**.727**	.685
2	**.730**	.727	.727
4	.615	**.693**	.644
5	.561	.485	**.694**
6	.693	**.737**	.381
7	.390	.364	**.500**
9	**.887**	.692	.488
10	**.790**	.759	.364
12	**.773**	.622	.736
17	-	**.765**	.484
23	-	**.412**	.109
25	.776	**.847**	.724
26	.583	**.818**	.467
Avg.	**.671**	.665	.539

AU12, AU17, AU23, AU25, and AU27 on the CK+ database, AU1, AU5, AU6, AU10, AU12, and AU17 on the MMI database, and AU6 and AU7 on the McMaster database. Like the within database results, the proposed method achieves better cross-database results than other methods. For the experiment of training on the CK+ database and testing on the MMI database, the proposed method has the highest F1-measure on AU7, AU12, AU17, and the averaged value. For the experiment of training on the MMI database and testing on the CK+ database, the proposed method has the highest F1-measure on AU9, AU17, AU23, and the averaged value.

MLML exploits the label consistency and smoothness to fill in the missing values; BGCS handles partially observed labels by marginalizing over the unobserved values as a part of the inference procedure; and BN complements the missing AU labels through the learned AU-expression relations from ground-truth labels. Therefore, all these methods handle missing AU labels through the relations learned from partial available ground-truth labels. While the proposed method learns AU classifiers from expression-dependent AU ranking summarized from domain knowledge when AU labels are missing. The dependencies from domain knowledge are usually more general than the dependencies existing in partial available ground-truth AU labels. Although SHTL also exploits expression-AU dependencies from domain knowledge, SHTL consists of two classifiers: 1) AU classifiers from image features and 2) expression classifiers from AUs generated by the relations between expressions and AUs. SHTL learns AU classifiers indirectly, and the error caused by expression classifiers may propagate to the AU classifiers. On the contrary, the proposed method directly learns AU classifiers from the expression-dependent AU ranking summarized from domain knowledge. Thus, the proposed method is superior to current semi-supervised AU recognition methods.

TABLE XI
COMPARISON TO THE STATE-OF-THE-ART WITH AU ANNOTATIONS IN TERM OF F1-MEASURE ON THE MCMASTER DATABASE

AU	MC-LVM[8]	Ours
4	.472	.161
6	.978	.681
7	.679	.398
9	.371	.083
10	.583	.013
12	-	.672
20	-	.082
25	-	.202
26	-	.214
43	.725	.390
Avg.	**.635**	.290

D. Comparison to the State-of-the-Art Supervised Learning With Fully AU Labeled Data

We compare our weakly supervised AU recognition method to state-of-the-art supervised AU recognition with fully AU labeled data. On the CK+ database, we compare the proposed method to STM [7], HRBM [3], and MC-LVM [8], mentioned in Section I. On the MMI database, we compare the proposed method to SVM-HMM [23] and FFD [24], which are the state-of-the-art methods on the MMI database. On the McMaster database, we compare the proposed method to MC-LVM [8]. The comparison results on the three databases are illustrated in Tables IX–XI.

As shown in Tables IX–XI, the proposed method has worse performance. It is reasonable since the proposed method learns AU classifiers without any AU annotation and the other methods are traditional supervised learning with fully AU labeled data. Furthermore, in some cases, the performance of the proposed method is comparable or even better. Specifically, on the CK+ database, the proposed method is 5.8% less than the best method (MC-LVM) in terms of the averaged F1-measure. The proposed method has close results to other methods with AU labels on AU1, AU2, AU17, and AU24. Furthermore, the proposed method has the best results on AU5, AU12, AU23, and AU25. On the MMI database, the averaged F1-measure of the proposed method is about 13% lower than that of FFD. However, the performance of AU1, AU4, and AU12 are close to the state-of-the-art methods. AU5 and AU7 even have the best performances using the proposed method. The results demonstrate the effectiveness of the proposed method which learns AU classifiers without any AU annotation. Compared to the state-of-the-art supervised methods with fully AU labeled data, the proposed method has a wider application prospect since it needs no AU labels.

V. Conclusion

In this paper, we propose a weakly supervised AU recognition method without AU labels through exploiting the domain knowledge. Specifically, the expression-dependent AU rankings are obtained from the domain knowledge first. Then we train the AU classifier by adopting the rank loss to penalize the mapping functions when the labels are incorrectly ranked. We also extend the proposed weakly supervised method to the semi-supervised method to solve the problems with partial samples annotated by AU labels. The experimental results on three benchmark databases demonstrate that the proposed method can successfully leverage the domain knowledge for building multiple AU recognition classifiers, and has better AU recognition performances for both basic expressions and nonbasic expressions compared to the state-of-the-art weakly supervised and semi-supervised methods.

References

[1] Y. Tong and Q. Ji, “Learning Bayesian networks with qualitative constraints,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Anchorage, AK, USA, Jun. 2008, pp. 1–8.

[2] Y. Li, J. Chen, Y. Zhao, and Q. Ji, “Data-free prior model for facial action unit recognition,” *IEEE Trans. Affect. Comput.*, vol. 4, no. 2, pp. 127–141, Apr./Jun. 2013.

[3] Z. Wang, Y. Li, S. Wang, and Q. Ji, “Capturing global semantic relationships for facial action unit recognition,” in *Proc. IEEE Int. Conf. Comput. Vis.*, Sydney, NSW, Australia, Dec. 2013, pp. 3304–3311.

[4] Y. Zhu, S. Wang, L. Yue, and Q. Ji, “Multiple-facial action unit recognition by shared feature learning and semantic relation modeling,” in *Proc. IEEE 22nd Int. Conf. Pattern Recognit.*, Stockholm, Sweden, Aug. 2014, pp. 1663–1668.

[5] X. Zhang and M. H. Mahoor, “Simultaneous detection of multiple facial action units via hierarchical task structure learning,” in *Proc. IEEE 22nd Int. Conf. Pattern Recognit.*, Stockholm, Sweden, Aug. 2014, pp. 1863–1868.

[6] K. Zhao, W. S. Chu, F. De la Torre, J. F. Cohn, and H. Zhang, “Joint patch and multi-label learning for facial action unit detection,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 2207–2216.

[7] W.-S. Chu, F. De La Torre, and J. F. Cohn, “Selective transfer machine for personalized facial action unit detection,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Portland, OR, USA, Jun. 2013, pp. 3515–3522.

[8] S. Eleftheriadis, O. Rudovic, and M. Pantic, “Multi-conditional latent variable model for joint facial action unit detection,” in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 3792–3800.

[9] J. Wang, S. Wang, and Q. Ji, “Facial action unit classification with hidden knowledge under incomplete annotation,” in *Proc. 5th ACM Int. Conf. Multimedia Retrieval*, Shanghai, China, Jun. 2015, pp. 75–82.

[10] B. Wu, S. Lyu, B.-G. Hu, and Q. Ji, “Multi-label learning with missing labels for image annotation and facial action unit recognition,” *Pattern Recognit.*, vol. 48, no. 7, pp. 2279–2289, 2015.

[11] Y. Li *et al.*, “Facial action unit recognition under incomplete data based on multi-label learning with missing labels,” *Pattern Recognit.*, vol. 60, pp. 890–900, Dec. 2016.

[12] Y. Song, D. McDuff, D. Vasisht, and A. Kapoor, “Exploiting sparsity and co-occurrence structure for action unit recognition,” in *Proc. 11th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit.*, vol. 1. Ljubljana, Slovenia, May 2015, pp. 1–8.

[13] S. Du, Y. Tao, and A. M. Martinez, “Compound facial expressions of emotion,” in *Proc. Nat. Acad. Sci.*, vol. 111, no. 15, pp. E1454–E1462, 2014.

[14] W. V. Friesen and P. Ekman, *EMFACS-7: Emotional Facial Action Coding System*, Univ. California at San Francisco, San Francisco, CA, USA, vol. 2, 1983, p. 1.

[15] P. Lucey *et al.*, “The extended Cohn–Kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, San Francisco, CA, USA, Jun. 2010, pp. 94–101.

[16] A. Ruiz, J. Van de Weijer, and X. Binefa, “From emotions to action units with hidden and semi-hidden-task learning,” in *Proc. IEEE Int. Conf. Comput. Vis.*, Santiago, Chile, Dec. 2015, pp. 3703–3711.

[17] K. M. Prkachin, “The consistency of facial expressions of pain: A comparison across modalities,” *Pain*, vol. 51, no. 3, pp. 297–306, 1992.

[18] P. Zhong and M. Fukushima, “A new multi-class support vector algorithm,” *Optim. Meth. Softw.*, vol. 21, no. 3, pp. 359–372, 2006.

[19] M.-L. Zhang and Z.-H. Zhou, “A review on multi-label learning algorithms,” *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 8, pp. 1819–1837, Aug. 2014.

[20] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, “Web-based database for facial expression analysis,” in *Proc. IEEE Int. Conf. Multimedia Expo*, Amsterdam, The Netherlands, Jul. 2005, p. 5.

[21] P. Lucey, J. F. Cohn, K. M. Prkachin, P. E. Solomon, and I. Matthews, “Painful data: The UNBC-McMaster shoulder pain expression archive database,” in *Proc. 9th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit.*, Santa Barbara, CA, USA, Mar. 2011, pp. 57–64.

[22] F. De la Torre *et al.*, “IntraFace,” in *Proc. 11th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit.*, vol. 1. Ljubljana, Slovenia, May 2015, pp. 1–8.

[23] M. F. Valstar and M. Pantic, “Fully automatic recognition of the temporal phases of facial actions,” *IEEE Trans. Syst., Man, Cybern. B, Cybern*, vol. 42, no. 1, pp. 28–43, Feb. 2012.

[24] S. Koelstra, M. Pantic, and I. Patras, “A dynamic texture-based approach to recognition of facial actions and their temporal models,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 11, pp. 1940–1954, Nov. 2010.

Shangfei Wang (SM’15) received the B.S. degree in electronic engineering from Anhui University, Hefei, China, in 1996 and the M.S. degree in circuits and systems and the Ph.D. degree in signal and information processing from the University of Science and Technology of China (USTC), Hefei, in 1999 and 2002, respectively.

From 2004 to 2005, she was a Postdoctoral Research Fellow with Kyushu University, Fukuoka, Japan. From 2011 to 2012, she was a Visiting Scholar with Rensselaer Polytechnic Institute, Troy, NY, USA. She is currently an Associate Professor with the School of Computer Science and Technology, USTC. She has authored or co-authored over 90 publications. Her current research interests include affective computing and probabilistic graphical models.

Dr. Wang is a member of ACM.

Guozhu Peng received the B.S. degree in mathematics from the South China University of Technology, Guangzhou, China, in 2016. He is currently pursuing the M.S. degree in computer science with the University of Science and Technology of China, Hefei, China.

His current research interest includes affective computing.

Shiyu Chen received the B.S. degree in computer science from Anhui University, Hefei, China, in 2015. She is currently pursuing the M.S. degree in computer science with the University of Science and Technology of China, Hefei.

Her current research interest includes affective computing.

Qiang Ji (F'15) received the Ph.D. degree in electrical engineering from the University of Washington, Seattle, WA, USA.

He is currently a Professor with the Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute (RPI), Troy, NY, USA. He recently served as a Program Director with the National Science Foundation (NSF), where he managed NSF's computer vision and machine learning programs. He also held teaching and research positions with the Beckman Institute, Urbana, IL, USA, the University of Illinois at Urbana–Champaign, Champaign, IL, USA, the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA, the Department of Computer Science, University of Nevada at Reno, Reno, NV, USA, and the U.S. Air Force Research Laboratory. He currently serves as the Director of the Intelligent Systems Laboratory, RPI. He has published over 160 papers in peer-reviewed journals and conferences. His research has been supported by major governmental agencies, including NSF, NIH, DARPA, ONR, ARO, and AFOSR as well as by major companies, including Honda and Boeing. His current research interests include computer vision, probabilistic graphical models, information fusion, and their applications in various fields.

Prof. Ji is an editor on several related IEEE and international journals and he has served as the general chair, the program chair, the technical area chair, and a program committee member in numerous international conferences/workshops. He is a fellow of IAPR.