Pattern Recognition 65 (2017) 71–81

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Feature and label relation modeling for multiple-facial action unit classification and intensity estimation

Shangfei Wang[a,*], Jiajia Yang[a], Zhen Gao[a], Qiang Ji[b]

[a] Key Lab of Computing and Communication Software of Anhui Province School of Computer Science and Technology, University of Science and Technology of China, Hefei, 230027 Anhui, PR China

[b] Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, 12180 NY, USA

ARTICLE INFO

Keywords:
AU recognition
AU intensity estimation
AU relation modeling
Multi-task feature learning
Bayesian network

ABSTRACT

In this paper, we propose multiple facial Action Unit (AU) recognition and intensity estimation by modeling their relations in both feature and label spaces. First, a multi-task feature learning method is adopted to learn the shared features among the group of facial action units, and recognize or estimate their intensity simultaneously. Second, a Bayesian network is used to model the co-existent and mutual-exclusive semantic relations among action units. Finally, through probabilistic inference, the learned Bayesian network combines the results of the multi-task learning with the AU relations it captures to perform multiple AU recognition and AU intensity estimation. Experiments on the extended Cohn-Kanade database, the MMI database, the McMaster database and the DISFA database demonstrate the effectiveness of our method for both AU classification and AU intensity estimation.

1. Introduction

Recent years have seen an increasing attention and considerable progress on facial Action Unit (AU) analysis due to its wide applications in human-computer interaction [36]. The main stream of current AU analysis either recognizes each AU individually or recognizes the fixed AU combinations. They thus either ignore the dependences among multiple AUs, or cannot handle thousands of possible combinations. Only recently, several work exploits AU dependencies to facilitate AU analyses from target labels or image features. However, little work leverages the relations embedded in both AU labels and image features for AU analyses. Since several AUs can be present at the same image, the dependencies inherent in target labels and in the shared features among multiple AUs carry crucial top-down and bottom up evidence respectively for improving AU analysis.

Therefore, in this paper, we tackle the problem of AU recognition and AU intensity estimation by exploiting the relations of AUs from both shared features and target labels. First, a multi-task learning (MTL) algorithm is adopted to learn the shared features among AUs and recognize multiple AUs or estimate multiple AU intensity simultaneously. Second, a Bayesian network (BN) is used to model AUs' relations from labels by structure and parameter learning. Finally, the outputs of multi-task learning algorithm are used as the inputs of the learned BN to obtain improved multiple AU recognition and intensity estimation. Experimental results on the extended Cohn-Kanade (CK+) database and the MMI database demonstrate that MTL outperforms single task learning, and the relationship model from AU labels further improves the performance of AU classification and the cross-database experiment shows the generalization ability of our relationship model. The results on the McMaster databases **and the DISFA database** for AU intensity estimation also indicate that learning the shared feature in each AU group by MTL improve the performance by single task learning and our BN model modeling the AU relationship further improve the AU prediction result by MTL.

The paper is organized as follows: Section 2 briefly reviews the related works on AU analysis. Section 3 describes the details of our proposed AU classification and AU intensity estimation approach considering the relations among AUs from both features and labels. Section 4 provides the experiments and analyses on **four** benchmark databases. Section 5 summarizes our work briefly.

2. Related work

Usually, several AUs can be present at the same image or an image sequence. Thus, AU recognition can be formulated as a multi-label problem. Due to the large number of possible label sets, multi-label recognition is rather challenging. Successfully exploiting the dependencies inherent in multiple labels is the key to facilitate the learning

* Corresponding author.
E-mail addresses: sfwang@ustc.edu.cn (S. Wang), yang25@mail.ustc.edu.cn (J. Yang), gzgqllxh@mail.ustc.edu.cn (Z. Gao), qji@ecse.rpi.edu (Q. Ji).

http://dx.doi.org/10.1016/j.patcog.2016.12.007
Received 6 September 2015; Received in revised form 7 October 2016; Accepted 12 December 2016
Available online 15 December 2016
0031-3203/

process. Present AU recognition research can be divided into three groups.

The first group recognizes each AU individually [30,31] directly from images or sequences. They are referred to as image-based AU recognition methods. For example, Valstar and Pantic [30] detected and tracked 20 facial points, and then used a combination of gentleBoost, support vector machines, and hidden Markov models as a classifier to detect 22 AUs [30]. Maaten et.al [31] adopted Active Appearance Model (AAM) features and linear chain conditional random field for AU recognition. These works treat each AU recognition individually as one-vs.-all scheme, do not consider the AUs' relations existing in features or labels. However, multiple AUs can appear together, and there exist dependencies among them. Exploiting such dependencies may help AU recognition and modeling.

The second group recognizes AU combinations. Littlewort et al. [14] adopted a linear SVM with Gabor features to analyze the AU combinations of 1+2, 2+4, 1+4, and 1+2+4. Lucey et al. [17] detected a few combinations of AUs (i.e. 1, 1+2, 4, 5) using SVM and Nearest Neighbor with AAM features. Although the co-existent relations among AUs in an AU combination has been exploited by the used features and classifier in these works, the combinations are manually determined and fixed. Each combination is regarded as a new AU. Thus, it is only feasible for a few combinations, and hard to detect thousands of possible combinations. In addition, such AU combinations only capture coexistent AUs. They cannot capture AUs that are mutually exclusive of each other.

The third group explicitly exploits the co-existent and mutual exclusive relations among AUs from target labels or image features. They are referred to as model-based AU recognition methods. Tong et al. [29,28] used Gabor features and Adaboost to recognize each AU first, then they modeled the relations among AU labels by Dynamic Bayesian Network(DBN). Their method, however, learns the AU relationships from training data and such learned relations may not generalize well to a different database. To mitigate this problem, Li et al. [12] proposed to use a knowledge-driven model that satisfies specific constraints from AU relationships, then convert model parameter samples into pseudo-data and finally learned the parameters from the pseudo-data. Their method generalizes better across databases than the data-based models. Other than using DBN, Wang et al. [33] proposed a three-layer Restricted Boltzmann Machine (RBM) to capture global relations among all AUs, and to integrate the AU measurements with the high-level AU semantical relationships for AU recognition. More recently, Song et al. [26] modeled AU sparsity and co-occurrence using a Bayesian compressed sensing model. These work successfully model AU label relations, but ignore inherent AU relations in image features, which are crucial for AU analysis. Zhang et al. [37] utilized multi-task multiple kernel learning to detect multiple AUs in the same group simultaneously. Yuce et al. [35] applied the multi-label discriminant Laplacian embedding method for multiple AU recognition. These work successfully model AU relations from image features, or AU dependencies among AU labels. Therefore, current model-based AU recognition methods rarely exploit the dependencies inherent in both AU labels and image features to facilitate AU recognition [39,5].

Due to the difficulties of collecting data with AU intensity values and the limited available database, only a little research pays close attention to the intensity of facial actions. Moreover, most of them measure the intensity of each facial actions independently, such as [18,2,6,23,9,3]. In this paper, we refer to these methods as image-driven intensity estimation methods. Only recently, several works consider AU relations for AU intensity estimation. Li et al. [13] proposed using DBN to model AU relationships for measuring their intensities. Sandbach et al. [22] adopted Markov random field structures to model AU combination priors to estimate the intensity of AUs in the upper face region. Kaltwang et al. [10] proposed a generative latent tree model to estimate multiple AU intensity. They are referred to as model-based AU intensity estimation methods. Similar to AU recognition methods, few model-based AU intensity estimation methods exploits the dependencies inherent in both AU labels and image features.

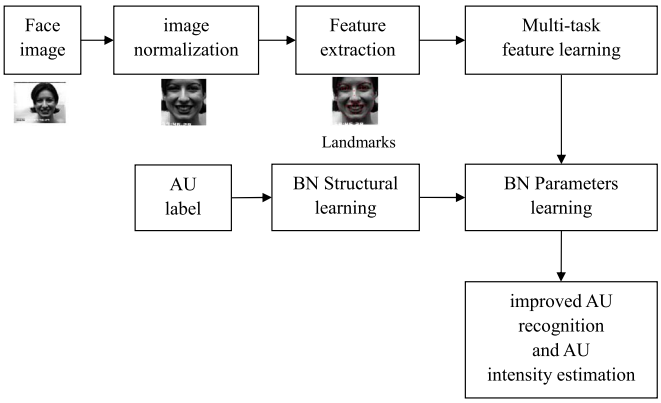

Fig. 1. Framework of multiple AU analysis.

To the best of our knowledge, this paper is the first work to recognize AUs and estimate AU intensity by exploring their relations at both feature and label levels [34]. By learning the shared features with MTL and modeling the dependencies among AU labels with BN, the proposed approach can exploit both top-down and bottom up relations among AUs to improve multiple AU classification and intensity estimation.

3. Multiple AU analysis approach

Fig. 1 shows the framework of our multiple AU analysis approach. First, facial images are normalized and features are extracted. Second, multi-task feature learning is performed to recognize multiple AUs or estimate intensities of multiple AUs. Third, AU relations in labels are modeled by BN. Finally, we use the trained BN to refine the output of multi-task feature learning to improve AU classification or AU intensity estimation.

3.1. Feature extraction

Both geometric features and appearance features are extracted from the images. First, the face images are normalized to a fixed size of 330×300 according to the location of eyes. For the databases who provide feature points, we use these feature points as geometric features directly. For the databases without feature points, we detect the feature points using [32]. For appearance features, the Gabor features are extracted from the regions of the forehead, between the eyebrows, between the eyes, outer corner of eyes, and the lower jaw, as shown in Fig. 2. These appearance features present the transient features caused by the movement of muscles [38].

3.2. Multiple AU recognition and intensity estimation by multi-task learning

Different from classical single-task learning, MTL trains multiple tasks jointly. Thus, one task in multi-task learning could benefit from learning other tasks.

Suppose there are m AUs: $\Lambda = \{\lambda_i\}_{i=1}^{m}$, to be analyzed. We treat each of them as a single task. Let $L(TD_i; \mathbf{w}_i)$ stand for the loss function of the i-th AU (i.e. λ_i) learning task on its training dataset TD_i, and $\mathbf{w}_i$ are the corresponding model parameters. For the AU intensity estimation task, we use the mean square errors as the loss function. For the AU classification task, we binarize the predicted value into AU states.

In standard learning paradigms, we analyze AUs independently and no information is shared among them, which is,

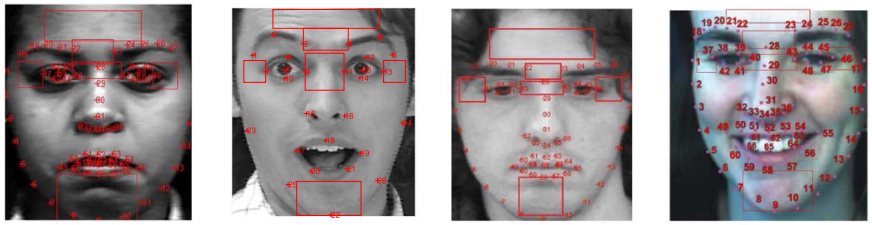

Fig. 2. Landmark points and appearance feature blocks on CK+, MMI, DISFA and McMaster databases.

$$\mathbf{W}^{\star} = \operatorname{argmin} \sum_i L(TD_i; \mathbf{w}_i) + \gamma \parallel \mathbf{W} \parallel_F^2 \tag{1}$$

where $\parallel \mathbf{W} \parallel_F^2$ is the squared Frobenius form: $\parallel \mathbf{W} \parallel_F^2 = \sum_i \parallel \mathbf{w}_i \parallel_2^2$.

In MTL, we aim at finding a joint feature subspace S where all AUs are well represented. Suppose the original feature $\mathbf{x}_i$ could be transferred to subspace S by $\mathbf{s}_i = \mathbf{S}^T\mathbf{x}_i$, where $\mathbf{S} \in \mathbb{R}^{D\times D}$ is an orthogonal matrix. Thus, the decision function turns out to be three equivalent terms,

$$\boldsymbol{\beta}_i^T \mathbf{s}_i = \boldsymbol{\beta}_i^T \mathbf{S}^T \mathbf{x}_i = \mathbf{w}_i^T \mathbf{x}_i \tag{2}$$

So, $\mathbf{w}_i = \mathbf{S}\boldsymbol{\beta}_i$. $\mathbf{S}$ and $\boldsymbol{\beta}_i$ can be simultaneously solved by

$$\mathbf{B}^{\star}, \mathbf{S}^{\star} = \operatorname{argmin} \sum_i L(TD_i; \boldsymbol{\beta}_i^T \mathbf{S}^T) + \gamma \parallel \mathbf{B} \parallel_{2,1}^2 \tag{3}$$

where $\parallel \mathbf{B} \parallel_{2,1}^2 = \sum_{d=1}^{D} \sqrt{\sum_i \boldsymbol{\beta}_{di}^2}$. This term calculates the 2-norm values of every dimension across all the AU recognition tasks, which captures the relationships in the feature-level space among AUs. Solving Eq. (3) can obtain the representation of the shared subspace S simultaneously, which can achieve our goal. Besides, to reduce the complexity of the optimization algorithm, Argyriou et. al [1] proved that a closed solution could be achieved by transferring Eq. (3) into the following problem:

$$\mathbf{W}^{\star} = \operatorname{argmin} \sum_i L(TD_i; \mathbf{w}_i) + \gamma \parallel \mathbf{W} \parallel_{\star}^2, \tag{4}$$

where $\parallel \mathbf{W} \parallel_{\star}$ is the trace norm of the parameter matrix.

To exploit the commonality among tasks, MTL is performed for a group of tasks, that have something in common. Suppose m AU analysis can be divided into P groups, and $q_{pi} \in \{0, 1\}$ to indicate whether the i-th task is assigned to group p. Let $\mathbf{Q}$ be the group assignment matrix composed by q_{pi} and $\mathbf{Q}_p \in \mathbb{R}^{n\times n}$ be a diagonal matrix with diagonal elements equal to q_{pi}. $\mathbf{W}_p$ is the parameter matrix for the p-th group, which is the parameter of the model. Multiple AU recognition task within a group are learned jointly, and the learning procedure for each group is independent. Kang et al. proposed an automatic grouping method to find the optimal $\mathbf{Q}$ in [11]. In our approach, before seeking the optimized model parameter $\mathbf{W}$, we defined $\mathbf{Q}$ manually according to the locations of AUs, as shown in Table 4. For each group, we get the optimized parameter $\mathbf{W}_p$ by solving the problem as follows:

$$\mathbf{W}^{\star} = \operatorname{argmin} \sum_i L(TD_i; \mathbf{w}_i) + \gamma \sum_p \parallel \mathbf{W}_p \parallel_{\star}^2 \tag{5}$$

$$\textbf{where} \parallel \mathbf{W}_p \parallel_{\star}^2 = Trace[\mathbf{W}\mathbf{Q}_p(\mathbf{W}\mathbf{Q}_p)^T]^{\frac{1}{2}}.$$

3.3. AU relationship modeling from labels by BN

As a probabilistic graphical model, BN can effectively capture the dependencies among variables in data. For this work, we use BN to capture the dependencies among AU labels.

3.3.1. BN structure and parameters learning

A BN is a directed acyclic graph (DAG) $G = (\Lambda, E)$, where $\Lambda = \{\lambda_i\}_{i=1}^m$ represents a collection of m nodes and E denotes a collection of arcs.

Given the data of multiple AU labels $TD = \{\lambda_i^j\}$, where $i = 1, 2, \ldots, m$ is an index to the number of nodes, and $j = 1, 2, \ldots, n$ is index to the number samples. The structure and parameter learning is to find a structure G that maximizes a score function. In this work, we employ the Bayesian Information Criterion (BIC)[24] score function which is defined as Eq. (6)

$$Q^{BIC}(G, \theta: G^*, \theta^*) = E_{G^*,\theta^*}[logP(TD|G, \theta)] - \frac{Dim(G)}{2}\log n \tag{6}$$

where the first term is the log-likelihood function of structure G with respect to data TD, representing how well G fits the data. The second term is a penalty relating to the complexity of the network, where $Dim(G)$ is the number of independent parameters and n is the number of samples. The BN structure learning algorithm proposed by Campos and Ji [4] are adopted to learn the dependencies among multiple AUs here.

For the AU classification task, after the BN structure is learned from the groundtruth labels, we link each node to the corresponding node for measurements as shown in Figs. 3 and 4. The parameters can be learned from the groundtruth labels and their measurements of the training data.

For the AU intensity estimation task, a continuous BN consists of continuous variables should be learned from groundtruth AU intensities. However, the structure learning for a continuous BN is more complex than that for a discrete BN. Therefore, in this paper, we

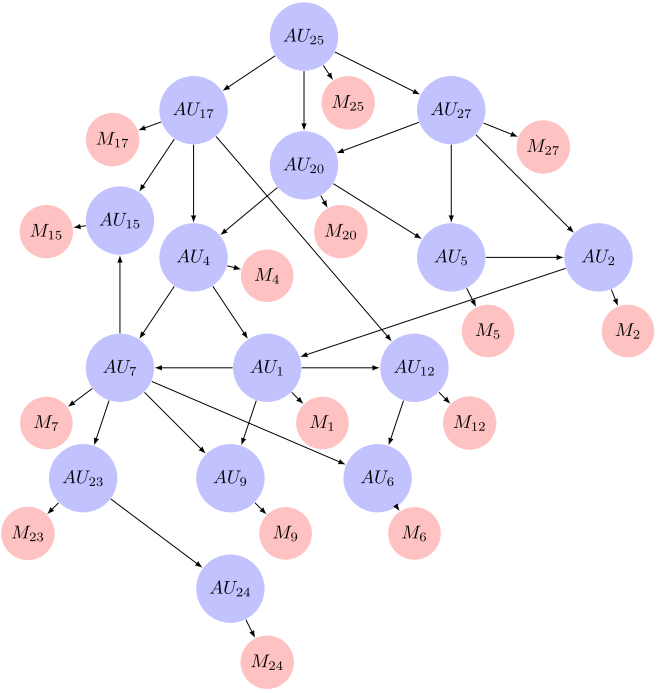

Fig. 3. The BN model learned from the CK+ database. "M" represents the AU measurement, "AU" represents the ground-truth AU label.

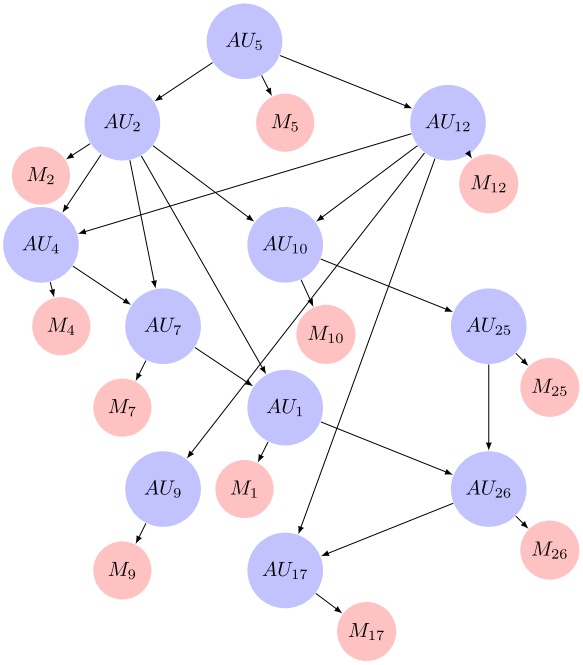

Fig. 4. The BN model learned from the MMI database. "M" represents the AU measurement, "AU" represents the ground-truth AU label.

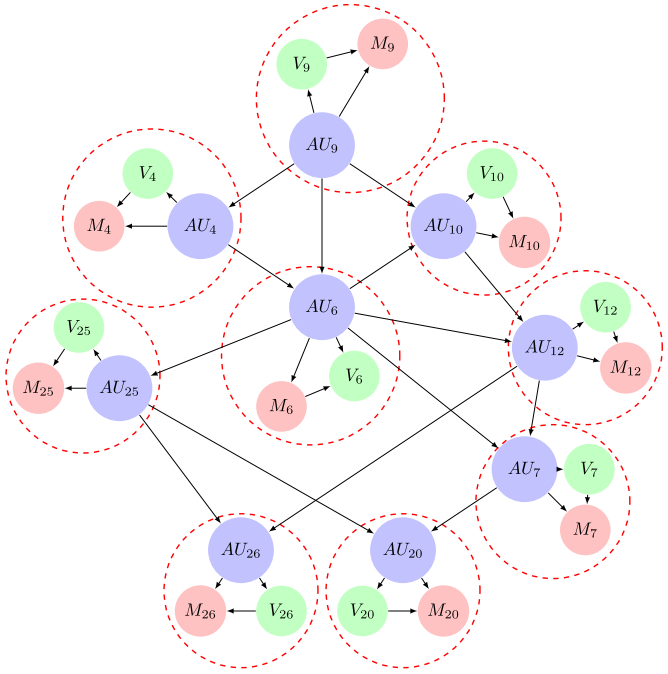

Fig. 5. The BN model learned from the McMaster database. "M" represents the measurement of AU intensity, "AU" represents the ground-truth AU intensity. "V" represents the discrete value of AUs discretized based on the ground-truth intensity. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

propose a method to simplify the BN structure learning with continuous nodes. We first discretize each continuous AU intensities into the label with two classes based on the mean value of the ground-truth value. We use the discrete labels to learn the BN structure to model the relationships among AUs. After learned the BN structure based on binarized AUs, we link each node with two corresponding gaussian nodes representing the ground-truth label and measurement as shown in the red circle in Figs. 5 and 6.

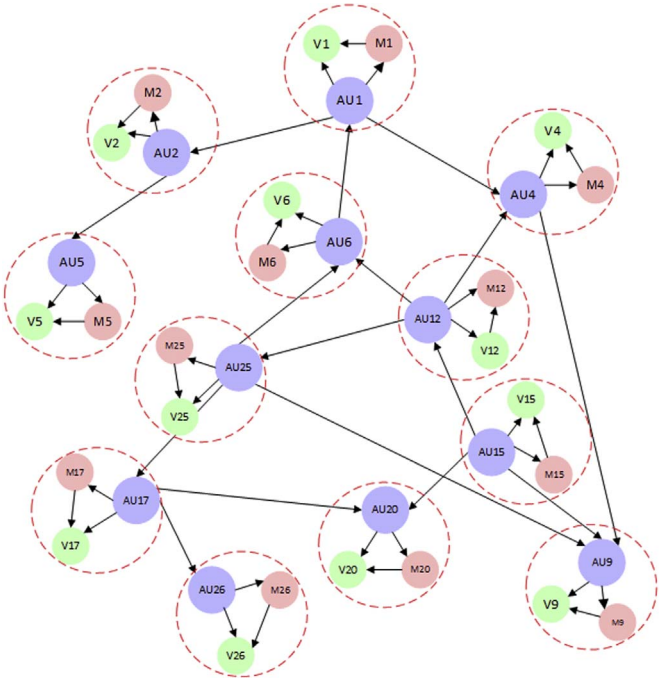

Fig. 6. The BN model learned from the DISFA database. "M" represents the measurement of AU intensity, "AU" represents the ground-truth AU intensity. "V" represents the discrete value of AUs discretized based on the ground-truth intensity. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

During the training, the parameters of our BN network are estimated using the maximum-likelihood estimation (MLE), as shown in Eq. (7):

$$\theta_{MLE} = argmax_{\theta\in\Theta}(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} lnP(\lambda_j^i|pa(\lambda_j^i);\theta)) \tag{7}$$

where θ is the parameter (i.e. the conditional probability of each node), $\lambda^1, \lambda^2, \ldots, \lambda^n$ are n independent samples, $\lambda_j{}^i$ is the value of the *jth* node of the *ith* sample, $Pa(\lambda_j^i)$ is the value of the *jth* node's parents of the *ith* sample.

3.3.2. BN inference

A complete BN model is obtained after the parameter and structure learning. Given the AU measurements obtained from MTL, the true AU category or AU intensity of the input sample is estimated through BN inference. During the BN inference, the posterior probability can be estimated by combining the likelihood from measurement with the prior model.

For the AU classification, let λ_i and M_i, $i \in \{1,\ldots,m\}$, denote the AU label variable and the corresponding measurement obtained from MTL respectively. Then, most probable explanation (MPE) [21] inference is used to estimate the joint probability of multiple AUs.

$$\mathbf{Y}^{\star} = \underset{\lambda_1,\lambda_2,\ldots\lambda_m}{\operatorname{argmax}} P(\lambda_1, \lambda_2, \ldots\lambda_m \Big| M_1,\ldots,M_m) = \underset{\lambda_1,\lambda_2,\ldots\lambda_m}{\operatorname{argmax}} (\prod_{i=1}^{m} P(M_i \Big| \lambda_i) \prod_{i=1}^{m} P(\lambda_i \Big| pa(\lambda_i))) \tag{8}$$

The first part of the equation is the likelihood of λ_j given the measurements and the second part is the product of the conditional probabilities of each category node λ_j given its parents $pa(\lambda_j)$, which are BN model parameters that have been learned. In AU classification, the inferred states of AUs are the states ($\mathbf{Y}^{\star} = (\lambda_1,\ldots,\lambda_m)$) with the highest probability given $M_1,\ldots,M_m$.

For AU intensity prediction, we use L_i, λ_i and M_i, $i \in \{1, 2,\ldots,m\}$,

denote the discrete labels, the corresponding AU intensity values and the corresponding measurements gained from MTL respectively. Then,

$$\mathbf{Y}^{\star} = \underset{L_1,\ldots,L_m,\lambda_1,\lambda_2,\ldots\lambda_m}{\operatorname{argmax}} P(L_1, L_2,\ldots,L_m, \lambda_1, \lambda_2,\ldots,\lambda_m | M_1, M_2,\ldots,M_m)$$
$$= \underset{L_1,\ldots,L_m,\lambda_1,\lambda_2,\ldots\lambda_m}{\operatorname{argmax}} \left(\prod_{i=1}^{m} P(M_i | L_i, \lambda_i) \prod_{i=1}^{m} P(\lambda_i | L_i) \prod_{i=1}^{m} P(L_i | pa(L_i))\right) \quad (9)$$

The condition probability in the equation is learned in the training phase. The inferred results are $\mathbf{Y}^{\star} = (\lambda_1,\ldots,\lambda_m, L_1,\ldots,L_m)$ with the highest probability given $M_1,\ldots,M_m$. In practice, we use the junction tree algorithm [8] to estimate the posterior probability effectively.

4. Experiments and results

4.1. Experimental conditions

To validate our approach, we conduct AU recognition experiments on the CK+ database [15] and the MMI database [20], and AU intensity estimation on the McMaster database [16] and the DISFA database [19].

The CK+ database consists of 593 posed expression image sequences, starting from the neutral frame and ending at the peak frame, from 123 subjects. 593 apex images signed with AU labels are selected, and the AUs, whose positive sample number is larger than 50, are used in the experiment, which are: AU1, AU2, AU4, AU5, AU6, AU7, AU9, AU12, AU15, AU17, AU20, AU23, AU24, AU25, and AU27. The sample distribution over AU states on the CK+ database is shown in Table 1. A 10-fold cross validation is adopted.

The MMI database consists of over 2900 videos and high-resolution images of 75 subjects. It contains recordings of the full temporal pattern of facial expressions, from neutral, through a series of onset, apex, and offset phases, and finally back again to a neutral face. In this experiment, 623 samples are selected, and 11 AUs are used in the experiment: AU1, AU2, AU4, AU5, AU7, AU9, AU10, AU12, AU17, AU25, and AU26. The sample distribution over AU states on the MMI database is shown in Table 1. A 10-fold cross validation is adopted.

The McMaster database contains face videos of patients suffering from shoulder pain. Totally, 200 sequences of 25 subjects are recorded (48,398 frames). AU intensities are provided for each frame. AUs 4,6,7,9,10,12,20,25,26 and 27 are labeled on 6 levels (0–5) and AU 43 are labeled 2 levels(present or not). The sample distribution over AU intensity levels are listed in Table 2. For AU 27, its data are very unbalanced and AU 43 are labeled on only 2 levels, therefore we use the rest 9 AUs for our experiment. We adopt the leave-one-subject cross validation for our AU intensity estimation experiment.

Table 1
Sample distribution over AU states on the CK+ and MMI databases.

	CK+ database		MMI database	
Label	0	1	0	1
AU1	418	175	428	195
AU2	476	117	393	230
AU4	399	194	398	225
AU5	491	102	388	235
AU6	470	123	–	–
AU7	472	121	448	175
AU9	518	75	518	105
AU10	–	–	523	100
AU12	462	131	503	120
AU15	498	95	–	–
AU17	390	203	478	145
AU20	514	79	–	–
AU23	533	60	–	–
AU24	535	58	–	–
AU25	269	324	173	450
AU26	–	–	328	295
AU27	512	81	–	–

Table 2
Sample distribution over AU intensity levels on the McMaster Database.

Intensity	0	1	2	3	4	5
AU4	47,324	202	509	225	74	64
AU6	42,841	1776	1663	1327	681	110
AU7	45,034	1360	991	608	305	100
AU9	47,975	93	151	68	76	35
AU10	47,873	171	208	63	61	22
AU12	41,511	2145	1799	2158	736	49
AU20	47,692	286	282	118	0	20
AU25	45,992	766	803	611	138	88
AU26	46,306	430	918	265	478	1
AU27	48,380	6	3	3	6	0
AU43	45,964	2434	–	–	–	–

The DISFA database [19] contains spontaneous facial expressions of 27 young adults while watching emotion eliciting videos. Each facial image has been annotated with six scale AU intensity for 12 AUs (i.e. AU1, AU2, AU4, AU5, AU6, AU9, AU12, AU15, AU17, AU20, AU25, and AU26) by an expert FACS rater. The samples distribution of database is shown in Table 3. In our experiment, the 19,850 samples whose sum of intensity is larger than 10 are selected. The leave-one-subject cross validation is adopted.

We divided the AUs into several groups according to their locations, as shown in Table 4. For the AUs in the same group, we learn the shared features using the MTL described in Section 3.2.

To demonstrate the effectiveness of our methods, three experiments are conducted: AU recognition/AU intensity prediction considering each AU individually (single task), AU recognition/AU intensity prediction using MTL (MTL), and our approach consider AU relations from both features and labels (MTL+BN).

For AU classification, we calculate metrics from two aspects: example-based view and label-based view. For example-based view, we calculate accuracy and F1-score on each AU as well as the average of them. For label-based measures, we adopt macro F1-score and micro F1-score [27]. For AU intensity prediction, we use three metrics: the Pearson correlation coefficient, the intraclass correlation coefficient (ICC) [25] and the mean squared error(MSE).

4.2. Experimental results and analyses for AU classification

4.2.1. Analyses from example-based view

Table 5 provides the AU recognition results on the CK+ database from example-based view. From Table 5, we can find the follows:

First, comparing the AU recognition as single task and AU recognition as multi-task, the accuracies of 9 AUs and the F1-scores of 13 AUs increase under the help of shared feature learning, demonstrating the effectiveness of our proposed multi-task AU recog-

Table 3
Sample distribution over AU intensity levels on the DISFA database.

Intensity	0	1	2	3	4	5
AU1	14,942	344	286	2432	1291	555
AU2	16,388	144	83	2124	752	359
AU4	12,518	478	744	2335	2768	1007
AU5	19,168	213	192	189	62	26
AU6	9738	2734	3281	3367	589	141
AU9	16,120	377	772	2277	283	21
AU12	9312	816	796	6301	2453	172
AU15	16,815	1825	385	780	45	0
AU17	16,281	1160	758	1540	100	11
AU20	18,788	58	323	681	0	0
AU25	4827	860	1120	7720	4464	859
AU26	11,493	3923	1701	2327	234	172

Table 4
AU group definitions.

Groups (Facial regions)	CK+	MMI	McMaster	DISFA
Eye	AU1, AU2, AU4, AU5, AU7	AU1, AU2, AU4, AU5, AU7	AU4, AU7	AU1, AU2, AU4, AU5
Mouth and chin	AU12, AU15, AU17, AU20, AU23, AU24, AU25, AU27	AU10, AU12, AU17, AU25, AU26	AU10, AU12, AU20, AU25, AU26	AU12, AU15, AU17, AU20, AU25, AU26
Cheek and nose	AU6, AU9	AU9	AU6, AU9	AU6, AU9

nition with the facial-region grouping strategy. Specifically, there exist a significant improvement for AU9, whose F1-score increases by 0.5592, and accuracy increases by 7.59%. Furthermore, the performance of its group partener – AU6 is also improved, with an increase of 0.1005 in F1-score and 1.53% in accuracy. Such improvement indicates that each AU in this group benefits from the shared feature space, especially for the AUs that take a small fraction of the samples, such as AU9, whose proportion of the total sample is 12.64%. Similar phenomena can be found in AU15(16.02%), AU20(13.32%), AU23 (10.11%) and AU24(9.78%), whose F1-scores increase by 0.6693, 0.4741, 0.4368 and 0.2133 respectively. In addition, the average accuracies and F1-scores increase by 2.15% and 0.1813, further confirming the effectiveness of MTL.

Second, comparing the result of our method using both MTL and BN with the one using MTL, there are 12 AUs with improvement in accuracy and 8 AUs with increase in F1-scores. The most significant improvement occurs in the F1-score of AU24, whose increase is 0.1117. From Fig. 3, we can find that the only link to AU24 is AU23. Thus, the improvement may be cause by the dependency between them. Furthermore, the average accuracies and F1-scores increase by 0.15% and 0.0283, demonstrating that the BN model can improve the results from MTL.

Fig. 3 shows the learned BN structure, and Table 6 lists the dependencies between each AU pair: $P(\lambda_j|\lambda_i)$, measuring the probability of label λ_j happens, given label λ_i happens. Comparing the learned BN with the dependency table, we find that the label pairs whose conditional probabilities are top ranked or bottom ranked are linked in the BN in most cases. For example, the link from AU1 to AU9 shows the mutual exclusive relationship, since $P(AU9|AU1)$ is zero. While the link from AU2 to AU1 represents the strong co-occurrence relationship between AU1 and AU2, since $P(AU1|AU2)$ is 1.0. Therefore, the learned BN can systematically capture the relations among AU labels. Thus, the learned BN can calibrate the AU recognition results from MTL.

Third, our proposed AU recognition method considering AU relations in both feature and label space performs best among the three AU recognition experiments, with the highest average accuracy and F1-scores.

Table 7 provides the AU recognition results on the MMI database from example-based view. Similar to the analysis on the CK+ database, we can find the following from Table 7:

First, comparing the AU recognition as single task and AU recognition as multi-task, the accuracies and F1-scores of 9 AUs increase by sharing feature spaces, which shows the effectiveness of our proposed multi-task AU recognition with the facial-region grouping strategy. Specifically, there exist a significant improvement for AU1, whose accuracy increases by 17.49% and F1-score increases by 0.3485. Furthermore, the performance of its group partner — AU7 are also improved, with a increase of 0.0398 in F1-score and 1.60% in accuracy. In addition, the average accuracies and F1-scores increase by 1.86% and 0.0411, which further confirming the effectiveness of MTL.

Second, comparing the result of our method with the one using MTL, there are 3 AUs with improvement in accuracy and F1-score, and all of the improvement are less than 1%/0.01. Furthermore, the average accuracies and F1-score of our proposed method slightly decrease by 0.05% and 0.0006, which indicates a minor improvement with the BN structure. However, the accuracy and F1-score of AU1 increase in both single task and MTL results, which shows that the structure learned from MMI could improve the recognition performance on AU1.

Fig. 4 shows the learned BN structure, and Table 8 lists all the dependencies. Comparing the learned BN with the dependency table, we find that the label pairs whose conditional probabilities are top ranked or bottom ranked are linked in the BN in most cases. For example, the link from AU5 to AU12 shows the mutual exclusive relationship, since $P(AU12|AU5)$ is 0.04. While the link from AU25 to AU26 represents the strong co-occurrence relationship between AU25 and AU26, since $P(AU25|AU26)$ is 1.0. Therefore, the learned BN can systematically capture the relations among AU labels. Thus, the learned BN can improve the recognition results from both single task and MTL, especially for AU1.

From Table 5 and 7, we can find that multi-task AU recognition is better than single task AU recognition for both databases; the learned

Table 5
AU recognition on CK+ database from example-based view.

AU	Accuracy (%)/F1-score of single task	Accuracy (%)/F1-score of single task +BN	Accuracy (%)/F1-score of MTL	Accuracy (%)/F1-score of MTL+BN	Accu.(%) in [14]
1	84.82/0.6875	84.49/0.6783	87.18/0.7610	84.65/0.6873	97.5
2	92.75/0.7795	92.75/0.7795	93.76/0.8213	92.75/0.7882	87.1
4	80.27/0.6139	79.60/0.6033	73.86/ 0.6301	78.58/0.6319	97.4
5	89.88/0.6341	90.73/0.6784	90.73/0.6893	91.40/0.7213	87.0
6	87.18/0.6162	87.18/0.6162	88.53/0.7167	88.87/0.7105	80.2
7	83.98/0.3949	83.98/0.4172	82.80/ 0.5565	84.32 /0.5550	89.1
9	89.71/0.3297	89.71/0.3297	97.30/0.8889	96.96/0.8732	100.0
12	90.22/0.7456	90.22/0.7478	90.22/0.7852	93.42/0.8482	92.4
15	84.15/0.0208	80.78/0.5512	91.06/0.6901	91.74/0.7030	91.0
17	87.86/0.8182	87.18/0.8100	85.67/0.7826	89.54/0.8480	89.0
20	87.18/0.0952	88.03/0.2022	90.05/0.5693	92.75/0.6718	91.1
23	89.88/0.0000	89.88/0.0000	91.74/0.4368	91.91/0.4286	81.3
24	90.22/0.0000	90.22/0.0000	90.05/ 0.2133	90.89/0.3250	N/A
25	88.53/0.8828	88.53/0.8828	82.97/0.8308	92.07/0.9162	90.7
27	96.12/0.8456	95.95/0.8400	97.13/0.9314	97.30/0.9000	N/A
Ave.	88.18/0.4976	87.95/0.5424	90.33/0.6789	90.48/0.7072	90.29

Table 6
AU dependencies on CK+ database.

λ_i \ λ_j	AU1	AU2	AU4	AU5	AU6	AU7	AU9	AU12	AU15	AU17	AU20	AU23	AU24	AU25	AU27
AU1	1.00	1.00	0.35	0.85	0.07	0.14	0.00	0.07	0.39	0.24	0.57	0.13	0.05	0.39	0.89
AU2	0.67	1.00	0.10	0.79	0.01	0.01	0.00	0.04	0.14	0.08	0.24	0.05	0.05	0.31	0.89
AU4	0.39	0.16	1.00	0.25	0.28	0.81	0.67	0.07	0.40	0.63	0.66	0.75	0.62	0.18	0.02
AU5	0.50	0.69	0.13	1.00	0.02	0.05	0.01	0.04	0.03	0.05	0.28	0.12	0.00	0.29	0.75
AU6	0.05	0.01	0.18	0.03	1.00	0.38	0.32	0.63	0.01	0.11	0.20	0.15	0.09	0.26	0.00
AU7	0.10	0.01	0.51	0.06	0.37	1.00	0.64	0.08	0.07	0.35	0.34	0.58	0.55	0.12	0.00
AU9	0.00	0.00	0.26	0.01	0.20	0.40	1.00	0.03	0.04	0.26	0.03	0.17	0.21	0.04	0.00
AU12	0.05	0.04	0.05	0.05	0.67	0.09	0.05	1.00	0.00	0.01	0.13	0.00	0.03	0.29	0.04
AU15	0.21	0.11	0.20	0.03	0.01	0.06	0.05	0.00	1.00	0.45	0.01	0.15	0.16	0.01	0.01
AU17	0.28	0.15	0.65	0.11	0.19	0.60	0.69	0.02	0.97	1.00	0.08	0.80	0.78	0.03	0.00
AU20	0.26	0.16	0.27	0.22	0.13	0.22	0.03	0.08	0.01	0.03	1.00	0.02	0.00	0.23	0.01
AU23	0.05	0.03	0.23	0.07	0.07	0.29	0.13	0.00	0.09	0.24	0.01	1.00	0.59	0.00	0.01
AU24	0.02	0.03	0.19	0.00	0.04	0.26	0.16	0.02	0.09	0.22	0.00	0.57	1.00	0.00	0.00
AU25	0.71	0.86	0.30	0.91	0.69	0.32	0.17	0.71	0.02	0.04	0.96	0.02	0.00	1.00	1.00
AU27	0.41	0.62	0.01	0.60	0.00	0.00	0.00	0.02	0.01	0.00	0.01	0.02	0.00	0.25	1.00

Table 7
AU recognition on MMI database from example-based view.

AU	Accuracy (%)/F1-score of single task	Accuracy (%)/F1-score of MTL	Accuracy (%)/F1-score of MTL + BN	F1- score in[7]
1	80.42/0.6188	**97.91/0.9673**	**98.56/0.9771**	0.850
2	96.15/0.9462	**96.79/0.9552**	96.79/0.9552	0.822
4	97.27/0.9623	**97.59/0.9667**	96.63/0.9538	0.828
5	95.18/0.9336	94.22/0.9204	94.22/0.9204	0.825
7	92.46/0.8498	**94.06/0.8896**	**94.22/0.8929**	0.810
9	96.47/0.8842	**96.63/0.8889**	96.63/0.8889	0.959
10	95.35/0.8398	**96.15/0.8776**	**96.31/0.8808**	0.877
12	98.56/0.9610	97.91/0.9482	97.91/0.9469	0.958
17	95.35/0.8897	**95.83/0.9097**	95.83/0.9097	0.828
25	92.78/0.8673	**93.10/0.8701**	92.78/0.8640	0.795
26	88.12/0.8737	**88.28/0.8847**	88.12/0.8814	0.885
Ave.	93.46/0.8751	**95.32/0.9162**	95.27/0.9156	0.858

BN structures improve the recognition performance of single task for both databases, but it only help multi-task AU recognition on the CK+ database, not the MMI database. It may indicate that the learned relations among AUs can help AU recognition especially when the measurements are poor.

4.2.2. Analyses from label-based view

Table 9 shows the label-based results on the CK+ and the MMI databases. It is clear that all the F1 measures of MTL is higher than those of single task, which indicates that MTL outperforms the classic method. Besides, the F1-measures of our method on CK+ is the highest among all the methods, which also demonstrates the effectiveness of our method. Specifically, The macro/micro F1-score of MTL+BN on MMI database is only 0.0028/0.0020 lower comparing with MTL, which could be treated as an equivalent result.

Table 9
Label-based AU recognition results on CK+ and MMI databases.

Database	Evaluation metric	Single Task	MTL	MTL+BN
CK+	Macro F1	0.8788	0.9071	0.9148
	Micro F1	0.8805	0.9133	0.9172
MMI	Macro F1	0.9279	0.9612	0.9584
	Micro F1	0.9313	0.9652	0.9632

4.2.3. Cross database experiments

To further validate the generalization ability of our proposed approach, we perform cross-database experiments. Since the feature points provided by the CK+ database and MMI database are different, we can not validate the generalization ability of multi-task learning due to different features for different databases. Therefore, we only verify the generalization ability of the learned BN structures.

Table 10 and 11 shows the cross-database experiment results on CK + and MMI database. It is clear that for single task, the average F1-score increases when refined by BN structure learned from other database for both databases. For MTL, the performance after improved by BN structure learned from other database is not always improved. Specifically, the learned BN from the MMI database can benefit AU recognition on the CK+ database in term of F1-score, but not vice versa. Considering that MTL outperforms single task, the results indicate that the structure learning is more effective when the

Table 8
AU dependencies on MMI database.

λ_i \ λ_j	AU1	AU2	AU4	AU5	AU7	AU9	AU10	AU12	AU17	AU25	AU26
AU1	1.00	0.87	0.23	0.67	0.18	0.00	0.05	0.10	0.15	0.77	0.62
AU2	0.74	1.00	0.13	0.80	0.07	0.02	0.02	0.09	0.15	0.85	0.65
AU4	0.20	0.13	1.00	0.24	0.60	0.20	0.27	0.00	0.40	0.47	0.24
AU5	0.55	0.79	0.23	1.00	0.17	0.13	0.11	0.04	0.13	0.83	0.60
AU7	0.20	0.09	0.77	0.23	1.00	0.23	0.23	0.00	0.34	0.49	0.34
AU9	0.00	0.05	0.43	0.29	0.38	1.00	0.43	0.00	0.38	0.76	0.52
AU10	0.10	0.05	0.60	0.25	0.40	0.45	1.00	0.05	0.30	0.85	0.45
AU12	0.17	0.17	0.00	0.08	0.00	0.00	0.04	1.00	0.00	0.83	0.46
AU17	0.21	0.24	0.62	0.21	0.41	0.28	0.21	0.00	1.00	0.45	0.21
AU25	0.33	0.43	0.23	0.43	0.19	0.18	0.19	0.22	0.14	1.00	0.66
AU26	0.41	0.51	0.19	0.47	0.20	0.19	0.15	0.19	0.10	1.00	1.00

Table 10
Cross database AU recognition on the structure learnt from CK+ database.

AU in MMI	Accuracy (%)/F1-score of single task	Accuracy (%)/F1-score of single task + BN	Accuracy (%)/F1-score of MTL	Accuracy (%)/F1-score of MTL + BN
1	80.42/0.6188	**88.92/0.8313**	97.91/0.9673	97.91/0.9673
2	96.15/0.9462	96.15/0.9462	96.79/0.9552	96.15/0.9467
4	97.27/0.9623	97.27/0.9623	97.59/0.9667	97.59/0.9667
5	95.18/0.9336	95.18/0.9336	94.22/0.9204	94.22/0.9204
7	92.46/0.8498	91.97/0.8418	94.06/0.8896	94.06/0.8896
9	96.47/0.8842	96.47/0.8842	96.63/0.8889	96.63/0.8889
12	98.56/0.9610	98.56/0.9610	97.91/0.9482	97.91/0.9482
17	95.35/0.8897	95.35/0.8897	95.83/0.9097	95.83/0.9097
25	92.78/0.8647	**92.78/0.8673**	93.10/0.8701	93.10/0.8701
Ave.	93.85/0.8789	**94.74/0.9019**	96.00/0.9240	95.93/0.9231

Table 11
Cross database AU recognition on the structure learnt from MMI database.

AU in CK+	Accuracy (%)/F1-score of single task	Accuracy (%)/F1-score of single task + BN	Accuracy (%)/F1-score of MTL	Accuracy (%)/F1-score of MTL + BN
1	84.82/0.6875	**84.99/0.6920**	87.18/0.6783	87.02/ **0.7616**
2	92.75/0.7795	92.75/0.7795	93.76/0.8213	93.42/0.8186
4	80.27/0.6139	**80.78/0.6250**	73.86/0.6301	**75.38**/0.6075
5	89.88/0.6341	**91.23/0.7111**	90.73/0.6893	**91.06/0.7337**
7	83.98/0.3949	82.46/0.3882	82.8 /0.5565	81.28/0.5110
9	89.71/0.3297	89.71/0.3297	97.30/0.8889	97.30/0.8889
12	90.22/0.7456	90.22/0.7456	90.22/0.7852	**90.56/0.7879**
17	87.86/0.8182	**88.03/0.8203**	85.67/0.7826	**85.83**/0.7824
25	88.53/0.8828	88.53/0.8828	88.53/0.8308	82.97/0.8308
Ave.	87.56/0.6540	**87.63/0.6638**	87.78/0.7403	87.20/ **0.7469**

measurements are poor.

It should also be noted that the refinement in AU1 contribute the most of the improvement in recognition, in both single task and MTL. Specifically, the F1-score is increased by 0.2125 for single task on the MMI database, and 0.0833 for multi-task on the CK+ database. The strong co-occurrence relationship might be the reason. It is clear that AU2 and AU1 are always linked in the structure learnt from both databases. The dependency tables (Table 6, 8) also correspond to this relationship. Both of the structure and dependency lead to the stability of these BN models learnt from them.

4.2.4. Comparison with related work for AU classification

Although many studies have been done on AU classification, and achieved good performance, only a few work exploits the dependencies among AUs. One representative work is Tong et al. [29], who propose to use a dynamic Bayesian network (DBN) to model the relationships among different AUs and their temporal evolutions for AU classification and conduct the experiments on the CK database. Therefore, it is not exact fair to compare the performance on the CK database with that on the CK+ database. Furthermore, Tong et al. [29] adopted the true skill score as the evaluation metric. True skill score, also called as Hansen Kuiper Discriminant, is the difference between the positive rate and the false positive rate. We adopted F1-score as the evaluation metric in our paper. Tong et al. [29] used DBN to exploit dynamic patterns in AUs, which is not the focus of our paper. Therefore, we do not compare our work with theirs directly. However, we conduct similar experiment on CK+ database using the learned BN model to improve the AU recognition from single task. The results are listed in the second column of Table 5. Comparing the fourth and the second column in Table 5, we can find that our proposed method considering AU relations from both labels and features performs better than Tong et al. [29]s′ work.

There exist lots of AU classification studies considering each AU individually, which evaluate on the CK+ database. We report the comparison of the proposed method with the current work in Table 5. Comparing the experimental results of our method and those of [14], we find that the accuracies of 8 AUs increase. Furthermore, the average accuracy of our method is larger than that of [14], demonstrating the superiority of our method.

We also compared the results from the most recent research on the MMI database [7] with ours. Our average F1-score is about 5% higher than that of [7]. Specifically, the F1-scores of 8 AUs increase, which shows the advantage of our approach.

4.3. Experimental results and analyses for AU intensity estimation

4.3.1. Analyses of AU intensity estimation on the McMaster database *and the DISFA* database

Table 12 provides the AU intensity prediction results on the McMaster database. From Table 12, we can find the following:

First, compared with single task method, AU intensity prediction using MTL increases the average correlation by 0.009 and decreases the average MSE by 0.001, which demonstrates the effectiveness of our method. Using MTL decreases the MSE for almost every AU except for AU4. The results show that our proposed multi-task AU intensity estimation with the facial-region grouping strategy can benefit from feature-level AU relations to improve the prediction performance. Specifically, the correlation for AU4 is significantly improved by 0.083. Furthermore, the performance of its group partner — AU6 are also improved, for the correlation increased by 0.004 and MSE decreased by 0.006.

Second, comparing the result of our method with the one of MTL, all the AUs except for AU12 have improvement in correlation and the average correlation is increased by 0.033. Most AUs improve in ICC and the average ICC is increased by 0.032. It appears that BN can improve multi-task. Specifically, we can find that AU7 increased in correlation by 0.08 compared with MTL as shown in Table 12. Table 12 also shows that the results refined by BN become worse in the mean square error. From Table 2, we can find the samples in different intensity level in the McMaster database have uneven distribution. Therefore, MSE sometimes is not effective for evaluating the perfor-

Table 12
AU intensity prediction on McMaster database .

AU	CORR/ICC/MSE of single task	CORR/ICC/MSE of MTL	CORR/ICC /MSE of MTL + BN	CORR/MSE of LT-all[10]
4	0.118/0.269/0.147	**0.201**/ **0.301**/0.159	**0.202**/**0.309**/0.217	0.03/0.51
6	0.364/0.524/0.659	**0.368**/0.523/0.653	**0.386**/**0.541**/0.753	0.60/1.06
7	0.392/0.514/0.301	**0.397**/ **0.525**/0.316	**0.478**/**0.642**/0.361	0.11/1.19
9	0.109/0.172/0.080	0.103/0.156/ **0.078**	**0.135**/**0.214**/0.250	0.10/0.27
10	0.257/0.392/0.069	0.240/0.347/ **0.068**	**0.292**/**0.435**/0.209	0.15/0.28
12	0.381/0.561/0.800	**0.392**/0.556/0.778	0.365/0.519/0.801	0.60/1.12
20	0.064/0.088/0.067	**0.073**/ **0.091**/**0.065**	**0.103**/0.080/**0.058**	0.09/0.19
25	0.427/0.563/0.237	**0.429**/0.559/0.236	**0.447**/ **0.615**/0.338	0.18/0.72
26	0.036/0.030/0.341	0.031/ **0.033**/**0.339**	**0.085**/0.024/**0.280**	0.01/0.50
Ave.	0.239/0.346/0.300	**0.248**/0.343/**0.299**	**0.277**/ **0.375**/0.363	0.208/0.649

Table 13
AU intensity prediction on DISFA database .

AU	CORR/ICC/MSE of single task	CORR/ICC/MSE of MTL	CORR/ICC/MSE of MTL+BN	CORR/ICC/MSE of LT-all[10]
1	0.264/0.231/3.306	**0.283/ 0.249**/4.280	0.243/ **0.249**/0.817	0.41/0.32/0.44
2	0.513/0.428/1.506	0.399/0.346/2.933	**0.689**/0.284/1.673	0.44/0.37/0.39
4	0.415/0.374/3.246	**0.651/ 0.639/ 2.350**	**0.709/ 0.821**/1.836	0.5/0.41/0.96
5	0.239/0.167/0.332	**0.296/ 0.294**/0.429	**0.445**/0.104/1.686	0.29/0.18/0.07
6	0.462/0.451/1.838	**0.475/ 0.464/ 1.801**	**0.600**/0.229/1.563	0.55/0.46/0.41
9	0.406/0.399/2.161	0.385/0.381/2.344	**0.404/0.441**/1.883	0.32/0.23/0.31
12	0.638/0.629/2.015	0.631/0.622/2.056	0.460/ **0.861**/0.731	0.76/0.73/0.4
15	0.279/0.279/1.225	0.273/0.273/ **1.220**	0.127/ **0.457**/0.773	0.11/0.07/0.17
17	0.449/0.436/1.140	**0.449/ 0.438/ 1.121**	0.309/ **0.689**/3.066	0.31/0.23/0.33
20	0.135/0.128/1.262	0.122/0.114/1.272	0.040/0.003/1.614	0.16/0.09/0.16
25	0.680/0.679/1.466	0.677/0.677/1.472	0.281/ **0.840**/2.184	0.82/0.8/0.61
26	0.139/0.132/2.782	**0.158/ 0.152/ 2.705**	**0.262**/0.144/0.352	0.49/0.39/0.46
Ave.	0.385/0.361/1.857	**0.400/ 0.387**/1.999	0.381/ **0.427**/1.515	0.43/0.36/0.39

mance. For example, if the AU4 is all predicted as 0, the MSE is 0.145 which is the smallest but the results provide no information. However, this situation will fail in calculating the correlation indicating that the predicted results are meaningless.

Table 13 shows the AU intensity prediction results on the DISFA database. From Table 13, we can find the following:

First, comparing with single task method, AU intensity prediction with MTL increased the average correlation by 0.015 and the average ICC by 0.026. Specifically, the ICC and correlation of almost half AUs are increased by using MTL, and the rest of AUs are nearly the same. This demonstrates that the proposed method using the facial-region grouping procedure can improve the performance by modeling feature-level AU relation. It's worth noting that the correlation for AU4 is greatly improved by 0.236, and the ICC for AU4 and AU5 is improved by 0.265 and 0.127 respectively. The performance of their group partner, AU6, is also improved with 0.013 improvement in correlation and ICC.

Second, the proposed method outperforms MTL. Specifically, the ICC of half AUs is increased, and thus the average ICC increased by 0.04. Furthermore, the correlation of half AUs is increased, from 0.019 to 0.29, and the average MSE is also reduced by 0.484. This shows that the label dependencies modeled by BN can successfully improve the correlations between predicted AU intensities and ground-truth.

Fig. 5 and 6 show the learned BN structure from the McMaster database and the DISFA database respectively. Table 14, 15 list all the dependencies. From the learned BN structure and the dependency table, the label pairs whose conditional probabilities are top ranked or bottom ranked are linked in the BN in most cases. For example, we can find that $P(AU6|AU9)$ is 0.9 in Table 14 indicating that there are strong co-occurrence relationship between AU6 and AU9 and there is link from AU9 to AU6. The learned BN in Fig. 5 modeling this relationship improves the AU6 and AU9 prediction performance by 0.018 and 0.032 in correlation. And we can also find that $P(AU4|AU9)$ is 0.83 in Table 15 indicating that there are strong co-occurrence relationship between AU4 and AU9 and there is a link from AU4 to AU9. The learned BN in Fig. 6 modeling this relationship improves the AU4 and AU9 prediction performance in correlation by 0.058 and 0.019, and increased ICC by 0.182 and 0.06 respectively.

Table 14
AU dependencies on McMaster database.

λ_j									
λ_i	AU4	AU6	AU7	AU9	AU10	AU12	AU20	AU25	AU26
AU4	1.00	0.58	0.30	0.18	0.16	0.41	0.03	0.24	0.07
AU6	0.11	1.00	0.38	0.07	0.08	0.83	0.05	0.15	0.09
AU7	0.10	0.62	1.00	0.07	0.06	0.61	0.06	0.15	0.10
AU9	0.45	0.90	0.56	1.00	0.39	0.57	0.05	0.46	0.07
AU10	0.32	0.89	0.36	0.31	1.00	0.55	0.13	0.62	0.18
AU12	0.06	0.67	0.30	0.03	0.04	1.00	0.02	0.10	0.08
AU20	0.04	0.42	0.28	0.03	0.10	0.20	1.00	0.40	0.19
AU25	0.11	0.35	0.21	0.08	0.13	0.30	0.12	1.00	0.20
AU26	0.04	0.23	0.17	0.01	0.04	0.25	0.06	0.23	1.00

4.3.2. Comparison with related work for AU intensity estimation

Among the few studies on AU intensity estimation, three works [13,22] exploit the dependencies among AUs using DBN [13] or MRF [22] or generative latent tree [10]. The first two works did not report the experimental results on the McMaster database, while the last work conducted experiments on both databases. Therefore, we compare our work with Kaltwang et al.'s [10]. From Table 12, we can find that our proposed method can achieve better performance with higher correlation and lower MSE on the McMaster database, further demonstrating the advantage of our method.

For the DISFA database, from Table 13, we can find the ICCs of our method are higher than those of [10] in most cases. However, the correlation of our method is slightly lower than theirs. These shows our method can achieve comparable performance with Kaltwang et al.'s [10]. Since Kaltwang et al. adopted all the images of the DISFA database in their experiments, the number of AUs with lower intensity in their experiments is much larger than ours. Therefore, the MSE may not be an effective metric.

5. Conclusion

In this paper, we tackle the problem of AU recognition and AU intensity estimation by exploiting the relations of AUs from both shared features and target labels, which carry crucial top-down and bottom up evidence for improving AU analysis, but has not been thoroughly exploited yet. First, we formulate the AU analyses task into MTL problem with face-region grouping strategy. It is different from the traditional method where all the AUs share the same features space, or none AUs share feature space. Second, we construct the relations among AUs' labels with structural and parameter learning using BN. Finally, we further improve the result of MTL using the learned BN. The results for AU classification and AU intensity estimation show our approach treating AU analyses as a MTL problem outperforms traditional single task methods, and the co-occurrence or exclusive relationship among AUs could also be obtained by BN and that improves the recognition and estimation results. The cross database AU recognition experiments demonstrate the generalization ability of label-level dependency captured by BN.

Conflict of interest

None declared.

Table 15
AU dependencies on DISFA database.

λ_j												
λ_i	AU1	AU2	AU4	AU5	AU6	AU9	AU12	AU15	AU17	AU20	AU25	AU26
AU1	1	0.67	0.29	0.05	0.04	0.11	0.06	0.01	0.02	0.01	0.23	0.15
AU2	0.88	1	0.23	0.08	0	0.01	0.06	0	0.005	0.01	0.12	0.07
AU4	0.20	0.12	1	0.004	0.14	0.35	0.03	0.09	0.23	0.07	0.40	0.12
AU5	0.78	0.91	0.09	1	0	0	0	0.03	0	0.04	0	0.01
AU6	0.04	0	0.22	0	1	0.21	0.64	0.03	0.01	0.09	0.81	0.06
AU9	0.17	0.01	0.83	0	0.33	1	0.01	0.11	0.04	0.09	0.56	0.10
AU12	0.03	0.02	0.02	0	0.30	0.003	1	0	0	0	0.95	0.04
AU15	0.04	0	0.64	0.01	0.17	0.34	0	1	0.26	0.16	0.43	0.12
AU17	0.05	0.01	0.84	0	0.01	0.06	0	0.13	1	0.08	0.05	0.04
AU20	0.05	0.05	0.62	0.02	0.51	0.33	0	0.19	0.20	1	0.59	0.27
AU25	0.08	0.03	0.19	0	0.25	0.11	0.65	0.03	0.01	0.03	1.	0.16
AU26	0.24	0.09	0.26	0	0.09	0.09	0.14	0.04	0.02	0.07	0.77	1

Acknowledgements

This work has been supported by the National Science Foundation of China (Grant No. 61175037, 61228304, 61473270), and the project from Anhui Science and Technology Agency (1508085SMF223).

References

[1] A. Argyriou, T. Evgeniou, M. Pontil, Convex multi-task feature learning, Machine Learning, vol. 73, pp. 243–272, 2008, 10.1007/s10994-007-5040-8. [Online]. Available: http://dx.doi.org/10.1007/s10994-007-5040-8
[2] M.S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, J. Movellan, Fully automatic facial action recognition in spontaneous behavior, in: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition, 2006, FGR 2006, , 2006, pp. 223–230.
[3] S.W. Chew, Recognising facial expressions with noisy data, 2013.
[4] C.P. de Campos, Q. Ji, Efficient structure learning of bayesian networks using constraints, J. Mach. Learn. Res. 12 (3) (2011) 663–689.
[5] S. Eleftheriadis, O. Rudovic, M. Pantic, Multi-conditional latent variable model for joint facial action unit detection, in: Proceedings of the in International Conference on Computer Vision (ICCV), Santiago, Chile, December 2015.
[6] L. Jeni, J.M. Girard, J.F. Cohn, F. De La Torre et al., Continuous au intensity estimation using localized, sparse facial feature space, in: Proceedings of the 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 2013 , 2013, pp. 1–7.
[7] B. Jiang, M. Valstar, B. Martinez, M. Pantic, A dynamic appearance descriptor approach to facial actions temporal modeling, IEEE Trans. Cybern. 44 (February (2)) (2014) 161–174.
[8] D. Kahle, T. Savitsky, S. Schnelle, V. Cevher, Junction tree algorithm, STAT 631 (2008).
[9] S. Kaltwang, O. Rudovic, M. Pantic, Continuous pain intensity estimation from facial expressions, in: Proceedings of the Advances in Visual Computing, Springer, 2012, pp. 368–377.
[10] S. Kaltwang, S. Todorovic, M. Pantic, Latent trees for estimating intensity of facial action units, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 296–304.
[11] Z. Kang, K. Grauman, F. Sha, Learning with whom to share in multi-task feature learning, in: Proceedings of the 28th International Conference on Machine Learning ICML-11, 2011, pp. 521–528.
[12] Y. Li, J. Chen, Y. Zhao, Q. Ji, Data-free prior model for facial action unit recognition, IEEE Trans. Affect. Comput. 4 (2) (2013) 127–141.
[13] Y. Li, S.M. Mavadati, M.H. Mahoor, Q. Ji, A unified probabilistic framework for measuring the intensity of spontaneous facial action units, in: Proceedings of the 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition FG, 2013 2013, pp. 1–7.
[14] G. Littlewort, J. Whitehill, T. Wu, I. Fasel, M. Frank, J. Movellan, M. Bartlett, The computer expression recognition toolbox (cert), in: Proceedings of the 2011 IEEE International Conference on,Automatic Face Gesture Recognition and Workshops FG 201), march 2011, pp. 298–305.
[15] P. Lucey, J. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews, The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression, in: Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), june 2010, pp. 94–101.
[16] P. Lucey, J.F. Cohn, K.M. Prkachin, P.E. Solomon, I. Matthews, Painful data: The unbc-mcmaster shoulder pain expression archive database, in: Proceedings of the 2011 IEEE International Conference on Automatic Face & Gesture Recognition and Workshops FG 2011, pp. 57–64.
[17] S. Lucey, A.B. Ashraf, J.F. Cohn, Investigating spontaneous facial action recognition through aam representations of the face, in: Proceedings of the In Face Recognition, Delac, 2007, pp. 275–286.
[18] M.H. Mahoor, S. Cadavid, D.S. Messinger, J.F. Cohn, A framework for automated measurement of the intensity of non-posed facial action units, in: Proceedings of the IEEE Computer Society Conference on, Computer Vision and Pattern Recognition Workshops, 2009, pp. 74–80.
[19] S. Mavadati, M. Mahoor, K. Bartlett, P. Trinh, J. Cohn, Disfa: a spontaneous facial action intensity database, IEEE Trans. Affect. Comput. 4 (2) (2013) 151–160.
[20] M. Pantic, M. Valstar, R. Rademaker, L. Maat, Web-based database for facial expression analysis, in: Proceedings of the IEEE International Conference on Multimedia and Expo, ICME, July 2005, pp. 5 pp.–.
[21] J. Pearl, Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann, 1988.
[22] G. Sandbach, S. Zafeiriou, M. Pantic, Markov random field structures for facial action unit intensity estimation, in: Proceedings of the IEEE International Conference on, Computer Vision Workshops (ICCVW), 2013, pp. 738–745.
[23] A. Savran, B. Sankur, M.T. Bilge, Regression-based intensity estimation of facial action units, Image Vis. Comput. 30 (10) (2012) 774–784.
[24] G. Schwarz, Estimating the dimension of a model, Ann. Stat. 6 (2) (1978) 461–464.
[25] P.E. Shrout, J.L. Fleiss, Intraclass correlations: uses in assessing rater reliability, Psychol. Bull. 86 (2) (1979) 420.
[26] Y. Song, D. McDuff, D. Vasisht, A. Kapoor, Exploiting sparsity and co-occurrence structure for action unit recognition, in: Proceedings of the 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), vol. 1. IEEE, 2015, pp. 1–8.
[27] M.S. Sorower, A literature survey on algorithms for multi-label learning, Oregon State University, Corvallis, 2010.
[28] Y. Tong, J. Chen, Q. Ji, A unified probabilistic framework for spontaneous facial action modeling and understanding, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2) (2010) 258–273.
[29] Y. Tong, W. Liao, Q. Ji, Inferring facial action units with causal relations, in Computer Vision and Pattern Recognition, in: Proceedings of the 2006 IEEE Computer Society Conference on, vol. 2, 2006, pp. 1623–1630.
[30] M. Valstar, M. Pantic, Fully automatic recognition of the temporal phases of facial actions, IEEE Trans. Syst., Man, Cybern., Part B: Cybern. 42 (1) (2012) 28–43.
[31] L. van der Maaten, E. Hendriks, Action unit classification using active appearance models and conditional random fields, Cogn. Process. 13 (2012) 507–518. http://dx.doi.org/10.1007/s10339-011-0419-7 [Online])(. Available: http://dx.doi.org/10.1007/s10339-011-0419-7).
[32] P. Wang, Q. Ji, Multi-view face and eye detection using discriminant features, Comput. Vis. Image Underst. 105 (2) (2007) 99–111.
[33] Z. Wang, Y. Li, S. Wang, Q. Ji, Capturing global semantic relationships for facial action unit recognition, in: Proceedings of the IEEE International Conference on, Computer Vision (ICCV), Dec 2013, pp. 3304–3311.
[34] L.Y. Q.J. Yachen Zhu, Shangfei Wang, Multiple-facial action unit recognition by shared feature learning and semantic relation modeling, in: Proceedings of the 22nd International Conference on Pattern Recognition, 2014.
[35] A. Yüce, H. Gao, J.-P. Thiran, Discriminant multi-label manifold embedding for facial action unit detection, in: Proceedings of the 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), vol. 6, 2015, pp. 1–6.
[36] Z. Zeng, M. Pantic, G. Roisman, T. Huang, A survey of affect recognition methods: audio, visual, and spontaneous expressions, IEEE Trans. Pattern Anal. Mach. Intell. 31 (1) (2009) 39–58.
[37] X. Zhang, M. Mahoor, Simultaneous detection of multiple facial action units via hierarchical task structure learning, in: Proceedings of the 22nd International Conference on Pattern Recognition (ICPR), Aug 2014, pp. 1863–1868.
[38] Y. Zhang, Q. Ji, Active and dynamic information fusion for facial expression understanding from image sequences, IEEE Trans. Pattern Anal. Mach. Intell. 27 (5) (2005) 699–714.
[39] K. Zhao, W.-S. Chu, F. De la Torre Frade, J. Cohn, H. Zhang, Joint patch and multi-label learning for facial action unit detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2015.

Shangfei Wang received her BS in Electronic Engineering from Anhui University,

Hefei, Anhui, China, in 1996. She received her MS in circuits and systems, and the PhD in signal and information processing from University of Science and Technology of China (USTC), Hefei, Anhui, China, in 1999 and 2002. From 2004 to 2005, she was a postdoctoral research fellow in Kyushu University, Japan. Between 2011 and 2012, Dr. Wang was a visiting scholar at Rensselaer Polytechnic Institute in Troy, NY, USA. She is currently an Associate Professor of School of Computer Science and Technology, USTC. Dr. Wang is a senior member of the IEEE and a member of the ACM. Her research interests cover affective computing, and probabilistic graphical models.

Jiajia Yang received her BS in Software engineering from Dalian maritime university in 2015, and she is currently pursuing her MS in Computer Science in the University of Science and Technology of China, Hefei, China. Her research interesting is affective computing.

Zhen Gao received his BS in computer science from Nanjing University of Science and Technology in 2013, and received his MS in Computer Science in the University of Science and Technology of China, Hefei, China in 2016. His research interesting is affective computing.

Qiang Ji received his PhD in Electrical Engineering from the University of Washington. He is currently a Professor with the Department of Electrical, Computer, and Systems Engineering at Rensselaer Polytechnic Institute (RPI). He recently served as a program director at 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 at University of Illinois at Urbana-Champaign, the Robotics Institute at Carnegie Mellon University, the Dept. of Computer Science at University of Nevada at Reno, and the US Air Force Research Laboratory. Prof. Ji currently serves as the director of the Intelligent Systems Laboratory (ISL) at RPI. Prof. Ji's research interests are in computer vision, probabilistic graphical models, information fusion, and their applications in various fields. 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. Prof. Ji is an editor on several related IEEE and international journals and he has served as a general chair, program chair, technical area chair, and program committee member in numerous international conferences/workshops. Prof. Ji is a fellow of the IEEE and IAPR.