Posed and Spontaneous Expression Distinction Using Latent Regression Bayesian Networks

SHANGFEI WANG and LONGFEI HAO, University of Science and Technology of China, China
QIANG JI, Rensselaer Polytechnic Institute

Facial spatial patterns can help distinguish between posed and spontaneous expressions, but this information has not been thoroughly leveraged by current studies. We present several latent regression Bayesian networks (LRBNs) to capture the patterns existing in facial landmark points and to use those points to differentiate posed from spontaneous expressions. The visible nodes of the LRBN represent facial landmark points. Through learning, the LRBN captures the probabilistic dependencies among landmark points as well as latent variables given observations, successfully modeling the spatial patterns inherent in expressions. Current methods tend to ignore gender and expression categories, although these factors can influence spatial patterns. Therefore, we propose to incorporate this as a kind of privileged information. We construct several LRBNs to capture spatial patterns from spontaneous and posed facial expressions given expression-related factors. Facial landmark points are used during testing to classify samples as either posed or spontaneous, depending on which LRBN has the largest likelihood. We conduct experiments to showcase the superiority of the proposed approach in both modeling spatial patterns and classifying expressions as either posed or spontaneous.

CCS Concepts: • **Computing methodologies** → *Biometrics*;

Additional Key Words and Phrases: Latent regression Bayesian network, posed and spontaneous expression distinction, spatial pattern, privileged information

ACM Reference format:
Shangfei Wang, Longfei Hao, and Qiang Ji. 2020. Posed and Spontaneous Expression Distinction Using Latent Regression Bayesian Networks. *ACM Trans. Multimedia Comput. Commun. Appl.* 16, 3, Article 80 (July 2020), 18 pages.
https://doi.org/10.1145/3391290

1 INTRODUCTION

Recent years have seen increasing research on expression recognition [2, 30, 43], where posed and spontaneous expression distinction is a challenging work. Spontaneous expressions disclose a person's inner feelings; posed expressions can be used to dissemble and do not always represent a person's true emotions. Automatically distinguishing between spontaneous and posed

This work has been supported by the National Science Foundation of China (Grant No. 917418129), and the project from Anhui Science and Technology Agency (1804a09020038).
Authors' addresses: S. Wang and L. Hao, University of Science and Technology of China, 443 HuangShan Rd., Hefei Shi, Anhui Sheng, China, 230027; emails: sfwang@ustc.edu.cn, hlf101@mail.ustc.edu.cn; Q. Ji, Rensselaer Polytechnic Institute, 110 8th St., Troy, New York, NY 12180-3590; email: qji@ecse.rpi.edu.

1551-6857/2020/07-ART80 $15.00
https://doi.org/10.1145/3391290

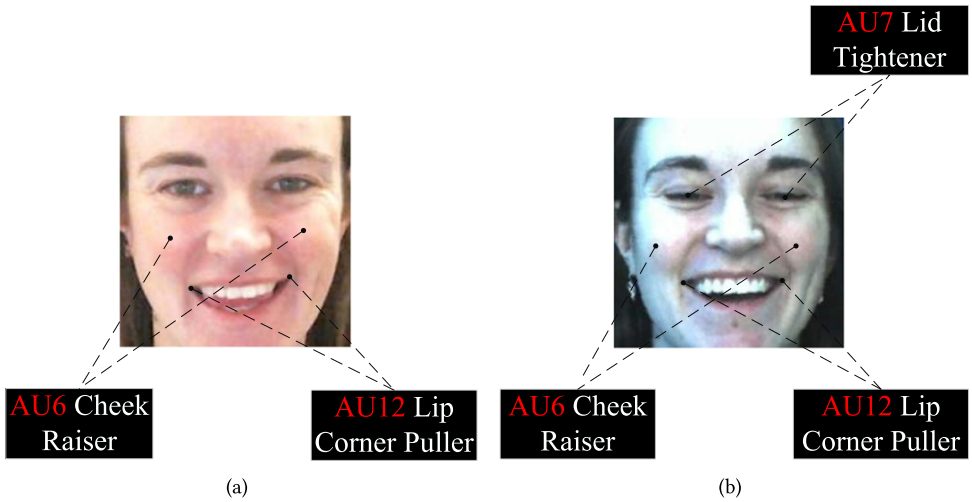


Fig. 1. Posed and spontaneous expression samples from the DISFA+ database. (a) Posed happiness. (b) Spontaneous happiness.

expressions can benefit many real-life scenarios in human–computer and human–human interaction. For example, social assistant robots can understand users more deeply and provide users timely assistance if they are able to accurately assess their feelings. Judges can feel more confident about cases if they know whether the criminal suspects are lying by distinguishing a natural expression from a manufactured one.

Due to the significant subject-dependent variations in expressions and the subtlety of the differences between posed and spontaneous expressions, classifying facial expressions is a challenging task. Current works use discriminative features or powerful classifiers. For feature extraction, either handcrafted features or learned representation through deep networks has been employed. Static and dynamic learning methods have both been used to differentiate between posed and spontaneous facial expressions.

Behavioral research indicates that the spatial patterns of a spontaneous expression are different from those of a posed expression. As shown in Figure 1(a) and Figure 1(b), in a spontaneous smile, both the orbicularis oculi (AU7) and zygomatic major (AU12) muscles contract. Only the zygomatic major contracts if the happiness expression is posed [9]. Additionally, the zygomatic major is more likely to contract symmetrically for spontaneous smiles than it is for posed ones [10].

Such inherent spatial patterns can facilitate the distinction between posed and spontaneous expressions, yet they have not been thoroughly exploited in the current research. We use a generative model, the latent regression Bayesian network (LRBN) [12, 36], to capture the embedded spatial patterns. This is a directed graphical model made up of a latent layer and a visible layer. We use the visible nodes of the LRBN to represent facial landmark points. Through learning, the parameters of the LRBN depict the probabilistic dependencies among both the latent and visible variables, faithfully representing inherent facial spatial patterns.

Current research typically ignores expression-related factors like gender or expression category, although several studies find that facial behavior patterns differ by gender [11], and different facial muscle movements formulate varying facial expressions [9, 10, 23]. Our method leverages these expression-related factors as privileged information while spatial patterns are modeled. Specifically, several LRBNs are created to model the embedded spatial patterns using posed and spontaneous expression data with different genders and expression categories. The displacement of

feature points are represented by the visible variables of the LRBN. In this way, the spatial patterns of different types of data are obtained by the probabilistic dependencies among visible nodes and among hidden nodes. During testing, images are assigned the expression label as the model with the maximum likelihood.

An earlier version of this article appears as Reference [12]. This article expands on that work by introducing these two expression-related factors during spatial pattern modeling. We then conduct experiments demonstrating that the incorporated gender and expression categories can help construct gender-specific and expression-specific spatial patterns for better distinction between posed and spontaneous expressions.

2 RELATED WORK

Current research on this task primarily focuses on feature extraction and posed versus spontaneous discriminators. Typically, spatial features and temporal features are specially designed to differentiate between posed and spontaneous facial expressions. Distance and angle are spatial features [7]; acceleration, amplitude, duration, speed, symmetry, and trajectory are temporal features [5, 33, 34]. Some studies also adopt features that are commonly used for expression recognition, including completed local binary patterns from three orthogonal planes (CLBP-TOP) [27], Gabor wavelet features [20], scale-invariant feature transform (SIFT) appearance features, and geometric facial animation parameter features [44]. Recently, learned features using deep networks have also been proposed [13]. After features have been extracted, either a static or dynamic machine learning method is employed to accomplish expression distinction. Both learning methods capture the relationships between expressions and extracted features, but dynamic classifiers also incorporate temporal dynamics. Static classifiers include linear discriminant analysis [5], support vector machines [20], Adaboost [20], gentle Boost, and relevance vector machines [34]. Dynamic classifiers include hidden Markov models [7] and dynamic Bayesian networks [33]. This research improves the ability to distinguish posed from spontaneous expressions. However, most of these feature-driven methods explore discriminative features and powerful classifiers without explicitly representing and leveraging the inherent spatial patterns.

Behavioral research shows that a posed expression is temporally and spatially different from a spontaneous one. The movement of facial muscles, i.e., the occurrence of facial action units (AUs), can be seen as certain spatial patterns. The most frequently observed AUs for spontaneous disgust, for example, are AU6 (cheek raiser), AU7 (lid tightener), and AU10 (upper lip raiser). A person deliberately depicting disgust typically activates AU4 (brow lowerer), AU7 (lid tightener), and AU17 (chin raiser). These three AUs are also frequently observed in posed sadness, but not in spontaneous sadness [23]. Such observations prove that embedded spatial patterns can aid in the distinction between spontaneous and posed expressions.

Wang et al. [38] recently proposed a model-based method that captured facial spatial patterns using multiple Bayesian networks (BNs). This method is limited by the first-order Markov assumption of Bayesian networks; only local probabilistic dependencies are captured. Wu et al. [40] proposed restricted Boltzmann machines (RBMs) to explicitly capture complex probabilistic dependencies among feature points, since RBMs use a layer of latent units to represent higher-order probabilistic dependencies [15].

While RBMs are able to model global probabilistic dependencies among visible units, the hidden units remain independent of one another other given visible units. The introduction of probabilistic dependencies among those units will improve a model's ability to explain the patterns embedded in visible units. Compared to undirected latent variable models like RBMs, the LRBN is a directed model. It is made up of a layer of visible nodes, a layer of latent nodes, and the directed connections from the latent nodes to the visible nodes. The displacement of feature points are represented by

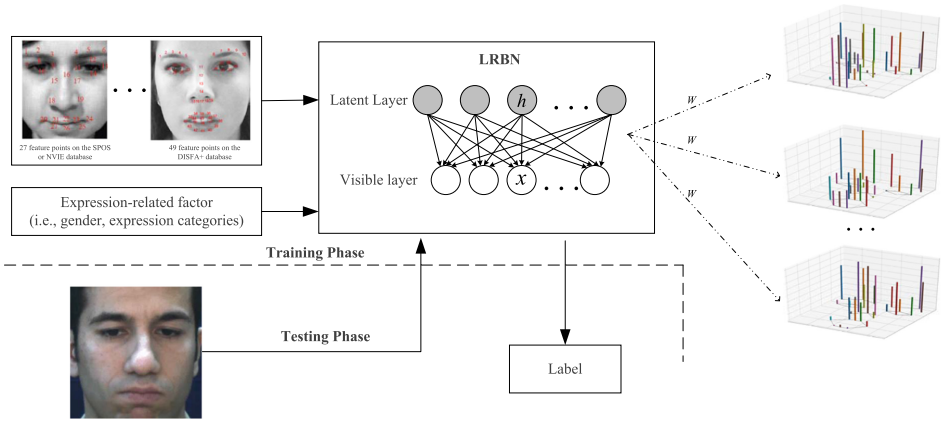


Fig. 2. Spatial patterns captured by LRBN for posed and spontaneous expression distinction.

visible variables. Spatial patterns are obtained by the probabilistic dependencies among both visible and hidden nodes. This type of model is better able to represent visible units by capturing not only the probabilistic dependencies among visible nodes, but also the probabilistic dependencies among hidden nodes given visible variables [12]. Therefore, we employ the LRBN to capture high-order and global probabilistic dependencies among the facial landmarks.

Few methods for posed and spontaneous expression distinction take advantage of expression-related factors like gender and expression categories, despite research demonstrating that males and females manifest facial expressions differently, and that different expression categories usually evoke different spatial patterns [19, 42]. Their work required those factors for training as well as testing. A sequential approach such as this is prone to propagating any errors of an expression-related factor on to the expression recognition task. Wang et al. [38] classified posed and spontaneous expressions with the help of gender and expression categories used as privileged information [35]. As privileged information, these factors are required for training but not testing. Like Wang et al., we choose to consider these factors privileged information. During training, multiple LRBNs are created to model the spatial patterns embodied in expressions given the expression-related factors. In the testing phase, the samples are assigned a label based on the LRBN with the largest likelihood.

3 PROPOSED METHOD

3.1 Brief Introduction of LRBN

The LRBN [24] is made up of a visible layer and a latent layer, as shown in Figure 2. A directed edge connects each latent variable to each visible variable. We denote visible variables as $\boldsymbol{x} = (x_1, \ldots, x_{n_d})$ and latent variables as $\boldsymbol{h} = (h_1, \ldots, h_{n_h})$, where n_d and n_h are dimensions of visible variable and latent variable respectively.

The joint probability of all variables in an LRBN, $P(\boldsymbol{x}, \boldsymbol{h})$, is shown in Equation (1),

$$P(\boldsymbol{x}, \boldsymbol{h}) = \prod_{j=1}^{n_h} P(h_j) \prod_{i=1}^{n_d} P(x_i|\boldsymbol{h}). \tag{1}$$

Equation (1) shows that the joint probability of a particular LRBN is equal to the product of the prior probabilities of a latent variable h_j, i.e., $P(h_j)$ and the conditional probabilities of a visible variable when considering all the latent variables, i.e., $P(x_i|h)$.

Equation (2) defines prior probability for a latent variable h_j, $P(h_j)$,

$$P(h_j = 1) = \text{sigm}(d_j), \tag{2}$$

where $\text{sigm}(d_j) = 1/(1 + \exp(-d_j))$ is the sigmoid function and d_j is the parameter. Essentially, it is a Bernoulli distribution.

Given latent variables, the conditional probability of a visible variable $P(x_i|\boldsymbol{h})$ is defined as linear Gaussian, as shown in Equation (3)

$$P\left(x_i|\boldsymbol{h}\right) \sim \mathcal{N}\left(\boldsymbol{w}_i^T\boldsymbol{h} + b_i, \sigma_i\right), \tag{3}$$

in which the mean of the linear Gaussian distribution is a linear combination of latent variables $\boldsymbol{h}$ values. The connecting weight between nodes h_j and x_i are represented by w_{ij}; b_i stands in as a constant, and the standard deviation is σ_i.

Incorporating Equation (2) and Equation (3) into Equation (1), we obtain Equation (4),

$$\begin{aligned} P_\Theta(\boldsymbol{x}, \boldsymbol{h}) &= \prod_j \frac{\exp(d_j h_j)}{1 + \exp(d_j)} \prod_i \mathcal{N}\left(x_i : \boldsymbol{w}_i^T\boldsymbol{h} + b_i,\ \sigma_i\right) \\ &= \frac{\exp(-E_\Theta(\boldsymbol{x}, \boldsymbol{h}))}{(2\pi)^{n_d/2} \prod_i \sigma_i \prod_j \left(1 + \exp(d_j)\right)}, \end{aligned} \tag{4}$$

where $\Theta = \{\boldsymbol{W}, \boldsymbol{\sigma}, \boldsymbol{b}, \boldsymbol{d}\}$, and

$$\begin{aligned} E_\Theta(\boldsymbol{x}, \boldsymbol{h}) = &\sum_i \frac{(x_i - b_i)^2}{2\sigma_i^2} - \sum_i \frac{x_i - b_i}{\sigma_i^2} \boldsymbol{w}_i^T\boldsymbol{h} \\ &+ \sum_i \frac{1}{2\sigma_i^2}\left(\boldsymbol{w}_i^T\boldsymbol{h}\right)^2 - \boldsymbol{d}^T\boldsymbol{h}. \end{aligned} \tag{5}$$

Compared to the Gaussian-Bernoulli restricted Boltzmann machine (GRBM) [16], which uses undirected links, the LRBN directly links visible and hidden nodes. This results in the term $\sum_i \frac{1}{2\sigma_i^2}(\boldsymbol{w}_i\boldsymbol{h})^2$ in Equation (5), which is similar to the energy function used by the GRBM. The relationships among latent variables are captured by this term. Patterns in the input data may be partially explained by probabilistic dependencies among the latent layer, taking the visible layer into account. As an additional advantage over the GRBM, an LRBN does not have a problem with intractable partition functions. Instead, the prior and conditional probabilities are multiplied to obtain the joint distribution.

3.2 Capturing Spatial Patterns through Model Learning

We construct several LRBN models using posed and spontaneous data given expression-related factors. Model inputs are the displacements of facial feature points. The LRBN can capture probabilistic dependencies among visible variables as well as the probabilistic dependencies among hidden variables through model learning. It is able to faithfully represent feature point displacements and successfully capture embedded spatial patterns.

In parameter learning, the goal is to most accurately estimate the parameters Θ when given a set of data samples $\mathcal{D} = \{\boldsymbol{x}^{(m)}\}_{m=1}^M$. M is the number of samples. This is done via marginal log

likelihood maximization as shown in Equation (6), where M represents the number of data samples,

$$\begin{aligned}\mathcal{L}(\mathcal{D};\Theta) &= \sum_m \log P_\Theta(\boldsymbol{x}^{(m)}) \\ &= \sum_m \log\left(\sum_{\boldsymbol{h}} P_\Theta(\boldsymbol{x}^{(m)},\boldsymbol{h})\right).\end{aligned} \tag{6}$$

Gradient ascent is used to maximize the above objective function. The exact gradient, taking parameter θ into account, is shown in Equation (7):

$$\nabla_\theta \mathcal{L}(\mathcal{D};\Theta) = \sum_m \sum_h P_\Theta(\boldsymbol{h}|\boldsymbol{x}^{(m)}) \frac{\partial - E_\Theta(\boldsymbol{x}^{(m)},\boldsymbol{h})}{\partial\theta}. \tag{7}$$

To obtain the gradient, we calculate the posterior probability $P_\Theta(\boldsymbol{h}|\boldsymbol{x})$ and the summation. The former is intractable for even one configuration $\boldsymbol{h}$, and the latter includes exponential terms for summation.

Variational inference algorithms are typically employed to minimize KL-divergence and approximate the true posterior distribution $P_\Theta(\boldsymbol{h}|\boldsymbol{x})$ with a factorized distribution $Q_\Phi(\boldsymbol{h}|\boldsymbol{x})$, shown in Equation (8):

$$KL(Q_\Phi(\boldsymbol{h}|\boldsymbol{x})\|P_\Theta(\boldsymbol{h}|\boldsymbol{x})). \tag{8}$$

Some approximations, such as the mean field algorithm [32], the wake-sleep algorithm [17], and inference networks [14, 18, 22, 29] result in a gap between the true and approximate posteriors, as the approximate distribution is unable to capture probabilistic dependencies. We use the true posterior probability through Gibbs sampling, which draws samples for one latent variable conditioned on all the other variables and therefore preserves probabilistic dependencies.

To deal with the exponential terms in the summation, we adopt the Markov chain Monte Carlo (MCMC) method, which is frequently utilized to estimate summation using samples as shown in Equation (9):

$$\nabla_\theta \mathcal{L}(\mathcal{D};\Theta) \approx \frac{1}{n}\sum_m \sum_n \frac{\partial - E_\Theta(\boldsymbol{x}^{(m)},\boldsymbol{h}^{(m,n)})}{\partial\theta}, \tag{9}$$

where $\boldsymbol{h}^{(m,1)},\ldots,\boldsymbol{h}^{(m,n)}$ represent n samples from $P(\boldsymbol{h}|\boldsymbol{x}^{(m)})$. We use a stochastic approximation procedure (SAP) framework to avoid multiple Gibbs chains [31]. The SAP estimates the gradient using a single sample of the latent variables.

If the learning rate γ_t satisfies conditions in Equation (10):

$$\begin{aligned}\sum_{t=1}^{\infty}\gamma_t &= \infty, \\ \sum_{t=1}^{\infty}\gamma_t^2 &< \infty.\end{aligned} \tag{10}$$

then the SAP will converge at the local optimum [41].

The gradient can then be predicted as Equation (11):

$$\nabla_\theta \mathcal{L}(\mathcal{D};\Theta) \approx \sum_m \frac{\partial - E_\Theta(\boldsymbol{x}^{(m)},\boldsymbol{h}^{(m)})}{\partial\theta}. \tag{11}$$

The derivative of w_{ij} is shown in Equation (12):

$$\frac{\partial - E_\Theta(\boldsymbol{x}^{(m)}, \boldsymbol{h}^{(m)})}{\partial w_{ij}} = \frac{h_j^{(m)}(x_i^{(m)} - \boldsymbol{w}_i^T \boldsymbol{h}^{(m)})}{\sigma_i^2}. \tag{12}$$

The gradient of other parameters can be derived similarly.

We use Gibbs sampling to draw a sample from $P(\boldsymbol{h}|\boldsymbol{x})$, so probabilistic dependencies can be preserved. We sample a single latent node while the others remain fixed,

$$h_j^t \sim P\left(h_j|\boldsymbol{x}, \boldsymbol{h}_{-j}^{t-1}\right), \tag{13}$$

where $\boldsymbol{h}_{-j}$ depicts the set of all latent variables with the exception of h_j. Algorithm 1 presents the SAP for LRBN learning.

ALGORITHM 1: Parameter Learning for an LRBN [12].

Input database $\mathcal{D} = \{\boldsymbol{x}^{(m)}\}_{m=1}^M$;
Output parameters $\Theta = \{\boldsymbol{W}, \boldsymbol{\sigma}, \boldsymbol{b}, \boldsymbol{d}\}$.

1: Randomly initialize the parameters Θ;
2: Generate Gibbs samples at time step 0;
3: **while** parameters are not converged, **do**
4: Randomly select a batch of data samples $\boldsymbol{x}$;
5: Perform Gibbs sampling to get a single sample of the latent variables for each input data, $\boldsymbol{h}^{(t)} \sim P(\boldsymbol{h}|\boldsymbol{x}, \boldsymbol{h}^{(t-1)})$, **where** $\boldsymbol{h}^{(t)} = (h_1^{(t)}, \ldots, h_{n_h}^{(t)})$, $P(\boldsymbol{h}|\boldsymbol{x}, \boldsymbol{h}^{(t-1)}) = (P(h_1|\boldsymbol{x}, \boldsymbol{h}_{-1}^{t-1}), \ldots, P(h_{n_h}|\boldsymbol{x}, \boldsymbol{h}_{-n_h}^{t-1}))$;
6: Compute the gradient using Equation (12);
7: Update the parameters, $\theta_t = \theta_{t-1} + \gamma_t \nabla_\theta \mathcal{L}(\boldsymbol{x})$.
8: **end while**

3.3 Using LRBN Inference for Posed and Spontaneous Expression Recognition

In this work, the LRBN is generative, modeling different types of facial expressions. Training results in several models $\boldsymbol{M} = \{\mathcal{M}^{(i,j)}, i \in \{1, 2\}, j \in \{1, \ldots, n\}\}$, where $i \in \{1, 2\}$ represents spontaneous and posed expressions, and n is the number of expression-related factors. For instance, if gender is privileged information, the n is 2. Taking into account the features of test image $\boldsymbol{x}$, posed and spontaneous expression distinction is conducted according to its likelihood for models $P(\boldsymbol{x}|\boldsymbol{M})$:

$$i, j = \arg\max_{i,j} P(\boldsymbol{x}|\boldsymbol{M}), \tag{14}$$

where i indicates the label of posed or spontaneous expression and j is the index of expression-related factors. Compared with our previous work, we have extended the proposed models through incorporating expression-related factors, i.e., gender and expression categories, as privileged information during spatial pattern modeling.

The inference task refers to computing the marginal likelihood $P(\boldsymbol{x})$ as shown in Equation (15):

$$P(\boldsymbol{x}) = \sum_{\boldsymbol{h}} P(\boldsymbol{x}, \boldsymbol{h}). \tag{15}$$

However, it would be intractable to directly compute $P(\boldsymbol{x})$ because of the exponential terms in the summation. We use a collection of samples drawn from the model given the input variables to implement the conservative sampling-based log-likelihood (CSL) method [3] for estimating the log-probability as shown in Equation (16):

$$\log \hat{P}(\boldsymbol{x}) = \log \text{mean}_{\boldsymbol{h} \in S} P(\boldsymbol{x}|\boldsymbol{h}), \tag{16}$$

where S indicates a set of latent variable samples h collected from $P(\boldsymbol{h}|\boldsymbol{x})$. It has been observed [3] that as the length of the Markov chain approaches infinity, the CSL estimator approaches the ground-truth log-likelihood. The expectation of the estimator is lower bound of the true log-likelihood.

4 EXPERIMENTS

4.1 Experimental Conditions

Several databases provide posed expressions as well as spontaneous expressions, including the BBC smile database [25], the UvA-Nemo smile database [6], the MAHNOB-Laughter database [26], the spontaneous vs. posed facial expression (SPOS) database [28], the USTC-NVIE (NVIE) database [37], and the DISFA+ database [21]. The first three databases consist of the smile expression only, while the last three contain multiple expression categories. As we want to analyze the performance of our models on general cases, we conduct our experiments on the SPOS, NVIE, and DISFA+ databases.

The SPOS database captures posed and spontaneous expressions for six emotions (i.e., happiness, sadness, anger, surprise, fear, and disgust) from seven subjects. There are 147 spontaneous and 84 posed expressions. Every expression sequence begins with an onset frame and concludes at the apex frame. We used all of the onset frames and apex frames for every expression sequence for our experiments.

The NVIE database contains onset and apex frames of those six expressions for both posed and spontaneous expression subsets. Following the same sample selection criteria as that in Wang et al. [39], we use 514 posed expressions and 514 spontaneous expressions. The samples are taken from 55 male subjects and 25 female subjects.

The DISFA+ database is extended from the DISFA database [21], adding posed expression videos of nine subjects. The DISFA database consists of spontaneous expression videos of happiness, surprise, disgust, sadness, and fear, collected from 27 subjects as they watched YouTube videos. The expression videos of the nine subjects who have both spontaneous expressions and posed expressions were used in our experiments. Each video starts with a neutral expression, and the apex frames are extracted by AU intensity, yielding 572 posed and 252 spontaneous expression samples.

Following the experimental conditions in Wang et al. [39] for the NVIE and the SPOS databases, we used displacement of 27 facial feature points (see Figure 2) between the apex and onset frames as the features. For the DISFA+ database, the features were defined as the displacement between apex and onset frames of 49 feature points (as shown in Figure 2) provided by the database constructors. Z-score normalization was used to normalize the features [1] so they satisfy standard Gaussian distribution and are unit free.

Subject-independent experiments were conducted to eliminate the influence of subject-dependent factors on model performance. We adopted leave-one-subject-out cross-validation for the SPOS and DISFA+ databases, as there were fewer subjects. For the NVIE database, we divided subjects into ten groups containing eight subjects each and applied leave-one-group-out cross-validation.

Three experiments were conducted: posed and spontaneous expression distinction without considering expression-related factors (i.e., the PS model), posed and spontaneous expression distinction using gender information as privileged information (i.e., the PS_gender model), and posed and spontaneous expression distinction using expression categories as privileged information (i.e., the PS_expression model). For the PS model, we built two LRBNs, one from posed samples and one from spontaneous samples. For the PS_gender model, we constructed four LRBNs: one from posed female samples, one from spontaneous female samples, one from posed male samples, and one

Table 1. Experimental Results on the SPOS, NVIE, and DISFA+ Databases

		The SPOS database		The NVIE database		The DISFA+ database	
		P	S	P	S	P	S
PS	P	49	35	501	13	572	0
	S	21	129	0	514	78	174
	Accuracy(%)	76.07		98.74		90.53	
	F1-score	0.6364		0.9872		0.9362	
PS_gender	P			507	7	572	0
	S	/		0	514	48	204
	Accuracy(%)			99.32		94.17	
	F1-score			0.9931		0.9597	
PS_expression	P			497	17	546	26
	S	/		0	514	4	248
	Accuracy(%)			98.35		96.36	
	F1-score			0.9832		0.9733	

Note: "P" represents posed expression and "S" represents spontaneous expression.

from spontaneous male samples. For the PS_expression model, we trained $N \times 2$ LRBNs from samples with expression information, where N is the number of expression categories. Since the SPOS database does not have enough samples to train posed models as well as spontaneous models for each gender and expression category, we conducted the PS_gender and PS_expression models on the NVIE and the DISFA+ databases only. During the training phase, the number of latent nodes is limited to avoid complex networks and over-fitting. The number of hidden nodes on the SPOS and NVIE databases is the same as Reference [12], which are 100 and 200. The number of hidden nodes on the DISFA+ database is 200. During the testing phase, we approximated the value of the log-likelihood by collecting 100,000 samples from a Markov chain for each test sample. We evaluated results using F1 score and accuracy.

4.2 Experimental Results and Analyses

4.2.1 Results and Analyses of Posed and Spontaneous Expression Distinction. Table 1 shows the results of our experiments on the task of posed versus spontaneous expression distinction. We can make the following observations:

First, for the PS model, the proposed LRBN achieves high accuracy and F1 scores for the recognition of posed and spontaneous expressions on all three databases, demonstrating its capability in capturing the global spatial patterns inherent in the two kinds of expressions. The performances on the NVIE database and the DISFA+ database are significantly better that the results of experiments on the SPOS database. This is reasonable, as these databases have nearly five times more samples. It further proves the importance of data size for pattern recognition. The models also perform better on the NVIE database than the DISFA+ database. While similar in size, the data distribution is much more balanced for the NVIE database, leading to better performance.

Second, the PS_gender model achieves higher accuracy and F1 scores than the PS model on both the NVIE and the DISFA+ databases. Specifically, the PS_gender model improves accuracy by 0.58% and 3.64% on the NVIE and the DISFA+ databases, respectively. For F1 score, the PS_gender model achieves 0.59% and 2.35% improvement over the PS model on the NVIE and the DISFA+ databases, respectively. This demonstrates that gender information available during training is beneficial for modeling gender-specific facial spatial patterns in both posed expressions and spontaneous expressions, which improves distinction performance during testing.

Table 2. Experimental Results on the NVIE Database

		Male	Female	Happiness	Disgust	Fear	Surprise	Anger	Sadness
PS	Accuracy(%)	98.39	99.65	100.00	98.92	100.00	99.36	96.70	97.50
	F1-score	0.9837	0.9964	1.0000	0.9891	1.0000	0.9935	0.9659	0.9744
PS_gender	Accuracy(%)	99.20	99.65	100.00	100.00	100.00	99.36	96.70	100.00
	F1-score	0.9919	0.9964	1.0000	1.0000	1.0000	0.9935	0.9659	1.0000
PS_expression	Accuracy(%)	97.99	99.29	100.00	100.00	99.26	96.15	96.15	98.13
	F1-score	0.9795	0.9929	1.0000	1.0000	0.9926	0.9600	0.9600	0.9809

Table 3. Experimental Results on the DISFA+ Database

		Male	Female	Happiness	Disgust	Fear	Surprise	Sadness
PS	Accuracy(%)	88.92	91.85	91.15	91.48	86.67	90.16	90.68
	F1-score	0.9264	0.9444	0.9422	0.9421	0.9130	0.9314	0.9371
PS_gender	Accuracy(%)	91.89	96.04	94.25	94.89	95.00	93.03	94.92
	F1-score	0.9451	0.9721	0.9617	0.9644	0.9655	0.9504	0.9647
PS_expression	Accuracy(%)	97.03	95.81	95.58	96.02	98.33	95.90	98.31
	F1-score	0.9786	0.9688	0.9686	0.9707	0.9880	0.9688	0.9877

Third, compared to the PS model, the PS_expression model achieves better performance on the DISFA+ database. Accuracy on the DISFA+ database improves by 5.83%, and F1 score improves by 3.71%. This demonstrates that expression information available during the training phase may allow the LRBN to better capture expression-specific patterns to some extent. Results on the NVIE database are comparable. Compared with the performance of the PS model on other databases, the PS model already performs well enough on the NVIE database, so that the space of improvement is little. That is may be the reason that the PS_expression model on the NVIE database performs a little bit worse than the baseline PS model.

To analyze the statistical difference, we conducted t-test for LBRN versus LBRN + gender and LBRN+expression on the DISFA+ and NVIE database. On the DISFA+ database, the p-value is 5.4259e-20 and 3.2463e-21 for gender and expression respectively. On the NVIE database, the p-value is 0.0098 for gender. This p-value is less than 0.0500. The results indicate that the performance improvements by introducing gender and expression are statistically significant.

Tables 2 and 3 show the result respectively for different genders and expression categories when comparing the PS_gender/PS_expression model against PS model. From Tables 2 and 3, we can see the following. First, PS_gender model has achieved better results than PS model on different genders. This demonstrates that gender information helps improve distinction performance. Second, compared with the PS model, the PS_expression model greatly improves the results on the DISFA+ database. This indicates that expression information may help to capture expression-specifc patterns to improve distinction performance.

To avoid complex networks and over-fitting, we also do ablation studies to discuss the number of latent nodes. Take DISFA+ database as example, we perform ablation experiment of PS model on the DISFA+ database, the results are shown in Figure 3. From Figure 3, we find that when the number of hidden nodes is 200, the PS model performs best on the DISFA+ database. Therefore, we set the number of hidden nodes as 200 for the experiments on the DISFA+ database.

4.2.2 Spatial Pattern Analysis. To investigate the learned facial spatial patterns inherent in posed expressions and spontaneous expressions, the parameter W captured by LRBN models is

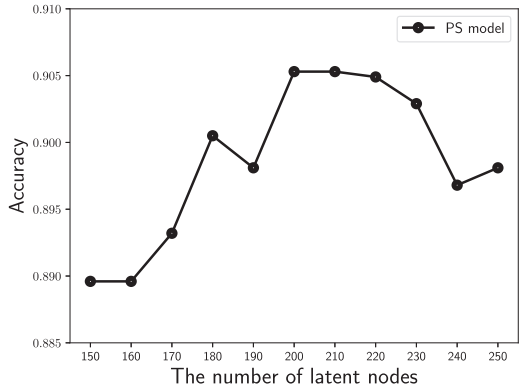


Fig. 3. The results of PS model on the DISFA+ database.

visualized. For each visible node v_i, we get W_i by average w_{ij} for all latent nodes, where i is the index of the visible node v_i and j is the index of the latent node h_j. Logically, as parameter W_i increases, the influence of v_i on captured spatial patterns also increases. Each facial landmark point corresponds to two visible nodes representing the x- and y-axes. We summarize the W_i for the x- and y-axes for each facial landmark point. Finally, the parameter W captured by the LRBN is visualized in a three-dimensional coordinate to illuminate the captured spatial patterns. Figures 4, 5, 6, and 7 show the captured spatial patterns embedded in posed expressions and spontaneous expressions for the PS model, the PS_gender model, and the PS_expression model.

From Figure 4, we find the following. First, every database shows that captured spatial patterns differ when expressions are posed rather than spontaneous. This finding is corroborated by current behavioral research. Second, on all three databases, the distribution of W is more symmetrical when an expression is posed. This may indicate that spontaneous expressions have more complex spatial patterns than posed expressions. Third, the W from the mouth region for posed expressions is larger than those for spontaneous expressions, while the W from the eye and eyebrow regions for posed expressions are smaller than those for spontaneous expressions on all three databases. This may indicate that the mouth region is more important to the display of posed expressions, while the eye and eyebrow regions are more essential for conveying spontaneous expressions. This is reasonable, since it is easier for people to control mouth movements than it is to control eye and eyebrow movements. These successfully captured spatial patterns result in good performance by the PS model described in Section 4.2.1 on posed and spontaneous expression distinction. We also find that on the NVIE and DISFA+ databases, the differences between the types of expressions are much greater than they are on the SPOS database. This may lead to the lower accuracy and F1 scores of the PS model on the SPOS database, as compared to the NVIE and DISFA+ databases. The learned Ws of the three databases are different for both types of models, proving the existence of database bias.

Figure 5 shows the differences in male and female spatial patterns. The difference in Ws between a male posed expression and a male spontaneous expression is much smaller than those for the females. This indicates that males may be better at disguising their expressions than females, since their posed and spontaneous expressions are more similar. Males usually spend more time in social situations and have more opportunity to display and practice their posed expressions. This obvious difference in the captured spatial patterns results in the good performance of the PS_gender model described in Section 4.2.1.

From Figures 6 and 7, we find unique patterns for the distributions of parameters W learned from posed and spontaneous expressions. It proves that different expressions tend to evoke

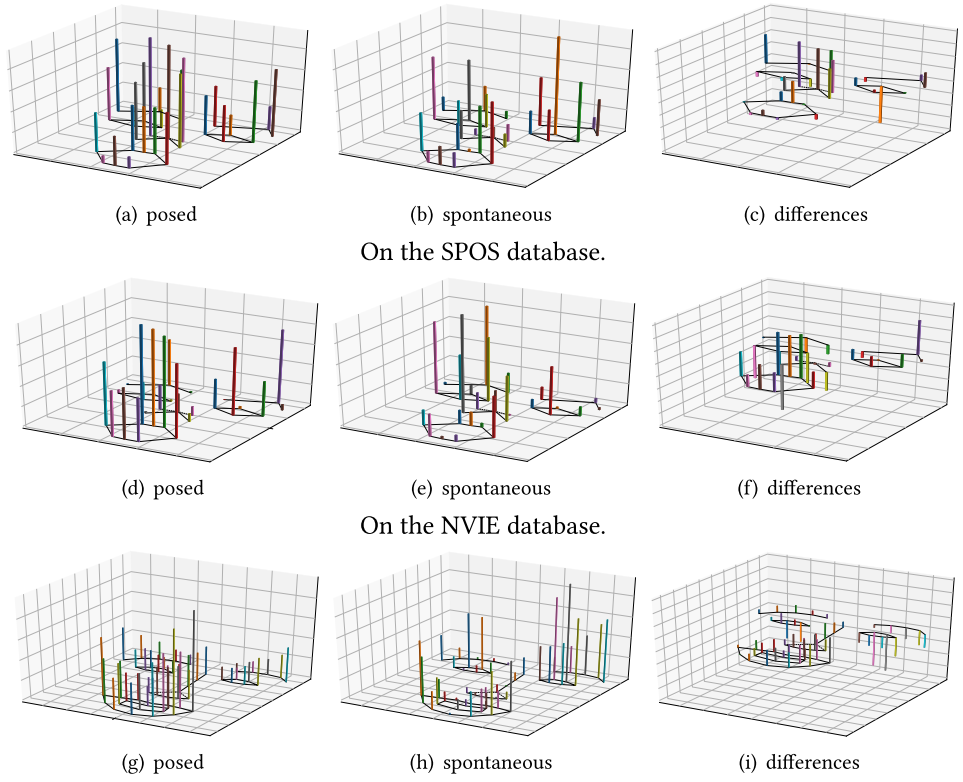


Fig. 4. The learned spatial patterns for posed and spontaneous expressions from three databases.

different spatial patterns. Specifically, for disgust, the Ws of points related to AU17 (i.e., points 25, 26, and 27 on the NVIE database and points 39, 40, 41, 42, 43, 47, 48, and 49 on the DISFA+ database) are slightly larger for posed models than spontaneous models. The Ws of points related to AU10 and AU12 (including points 20–24 on the NVIE database and points 32–38 and 44–46 on the DISFA+ database) are smaller for posed models than spontaneous models. This suggests that the upper lip raiser (AU10) as well as the lip corner puller (AU12) occur more frequently when an expression of disgust is spontaneous, and the chin raiser (AU17) is more frequent when the expression is posed. This observation is in accord with previous studies [23]. On the NVIE and DISFA+ databases, the Ws of points related to the zygomatic major and the orbicularis oculi parts are larger than other parts for spontaneous happiness models. However, for posed models, only the Ws of points related to the zygomatic major part are larger than other parts on the two databases. This observation is in accordance with previous studies [23]. For the surprise expression, the Ws of points related to AU1 (inner brow raiser), AU2 (outer brow raiser), AU5 (upper lid raiser), AU25 (lips part), and AU26 (jaw drop) are slightly larger for posed models than for spontaneous models on the NVIE and DISFA+ databases. These five AUs more frequently appear in posed expressions of surprise than in spontaneous surprise expressions. This observation is also in accord with previous studies [23]. It is difficult to distinguish between posed and spontaneous fear; the Ws for posed and spontaneous models are very similar on the NVIE and DISFA+ databases. For sadness, the Ws for the spontaneous model are slightly lower than for those of the posed model. This may suggest that compared to a posed expression of sadness, spontaneous sadness is individual or less

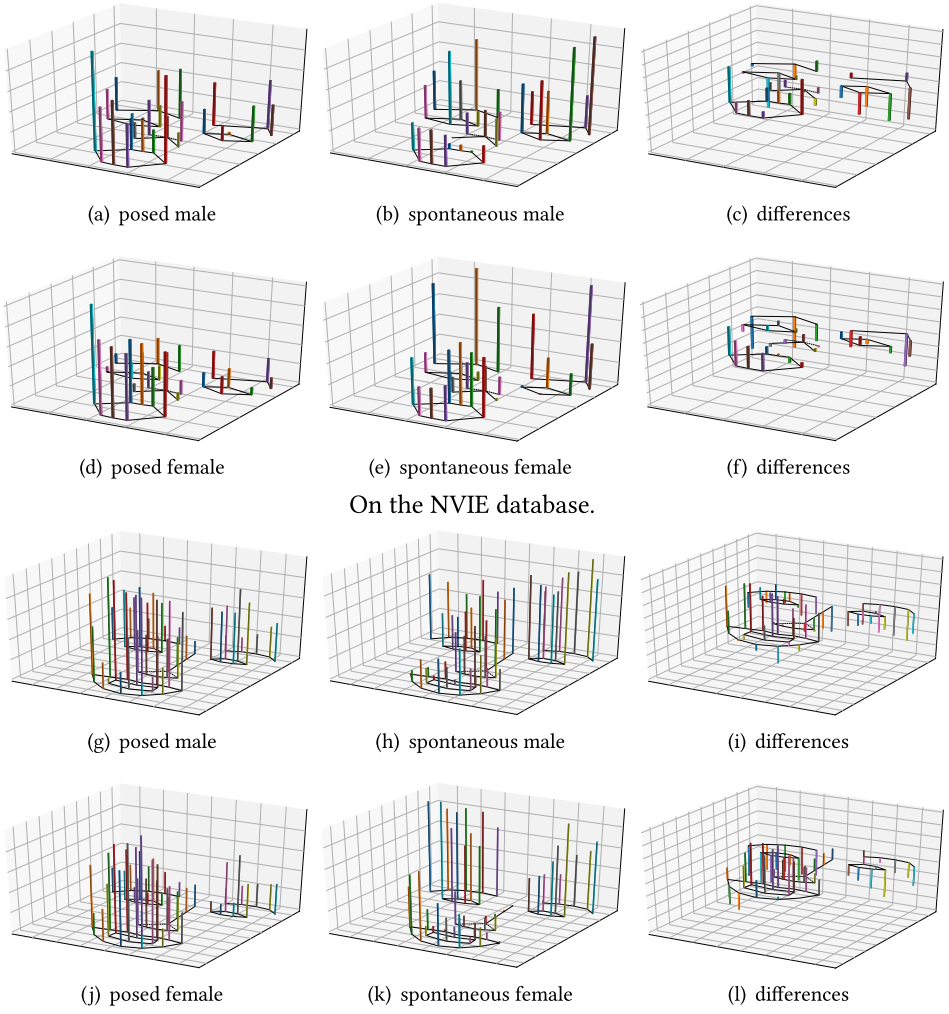


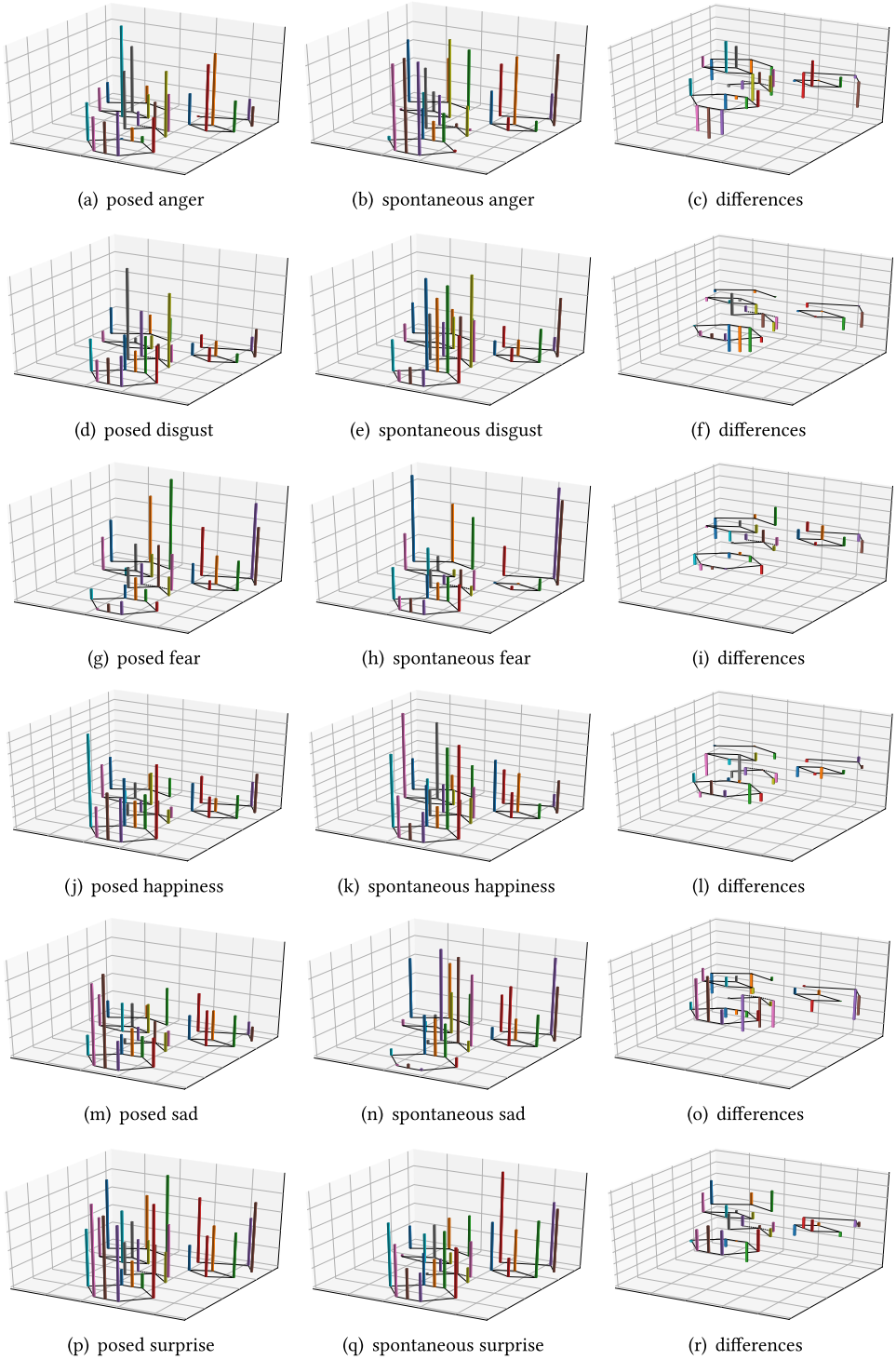
Fig. 5. The learned gender-specific spatial patterns for posed and spontaneous expressions from three databases.

likely to be observed. This is in accord with previous studies in References [23], [8], and [4]. For anger, the Ws for the spontaneous model are greater than for the posed model in most cases on the NVIE database. This may demonstrate that a spontaneous angry expression is stronger than a posed one.

4.3 Comparison to Related Work

To further validate our proposed method, we take a look at four other works that used either the SPOS database or the NVIE database to conduct experiments: Zhang et al.'s [44], Pfister et al.'s [28], Wang et al.'s [38], and Wang et al.'s [39]. The DISFA+ database was recently released, and thus no related works conduct posed and spontaneous expression distinction on it.

The first two methods are feature driven. Zhang et al. [44] utilize SIFT appearance-based features as well as FAP features to differentiate between posed expressions and spontaneous



(a) posed anger (b) spontaneous anger (c) differences

(d) posed disgust (e) spontaneous disgust (f) differences

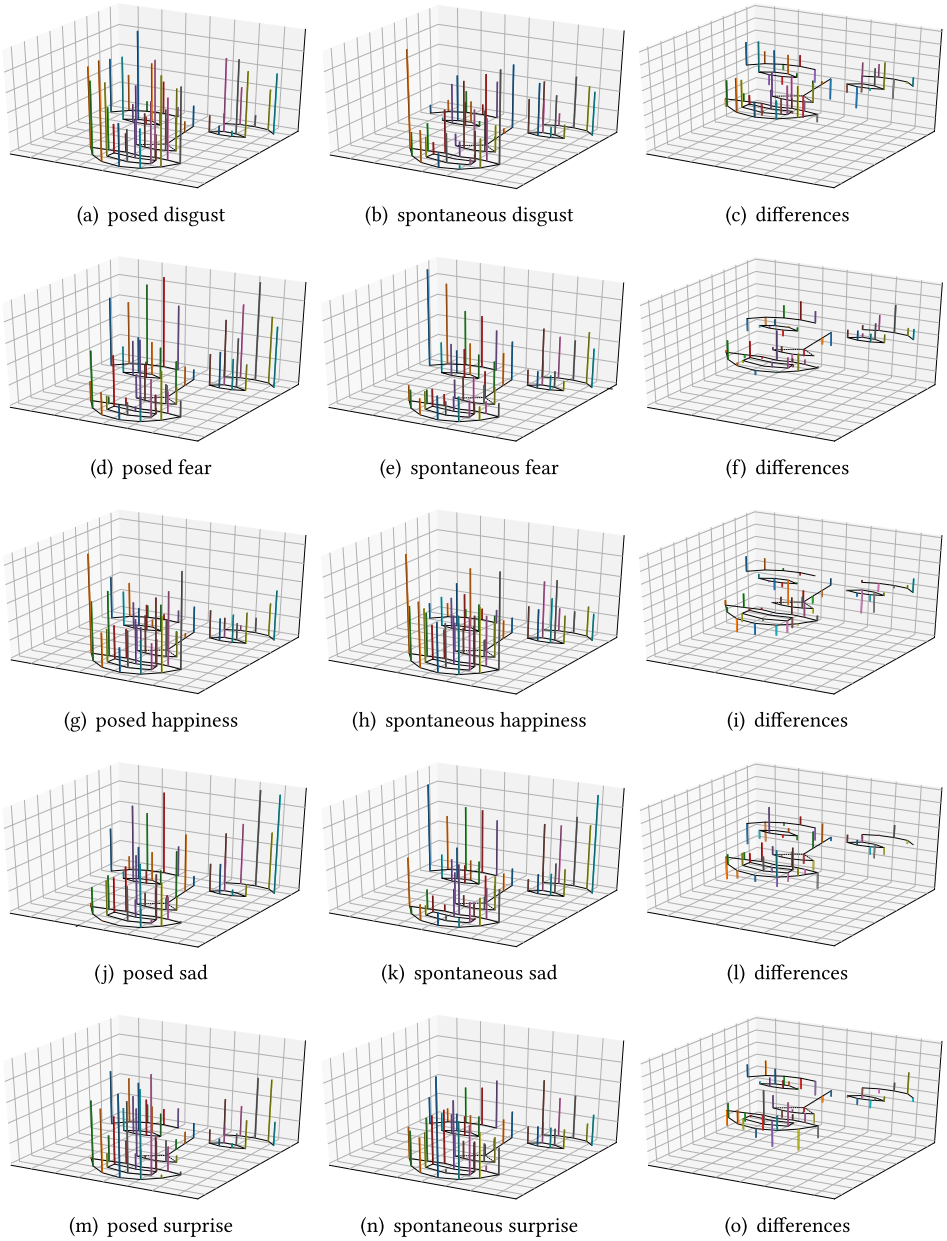
(g) posed fear (h) spontaneous fear (i) differences

(j) posed happiness (k) spontaneous happiness (l) differences

(m) posed sad (n) spontaneous sad (o) differences

(p) posed surprise (q) spontaneous surprise (r) differences

Fig. 6. The learned expression category-specific spatial patterns for posed and spontaneous expressions from the NVIE database.



(a) posed disgust (b) spontaneous disgust (c) differences

(d) posed fear (e) spontaneous fear (f) differences

(g) posed happiness (h) spontaneous happiness (i) differences

(j) posed sad (k) spontaneous sad (l) differences

(m) posed surprise (n) spontaneous surprise (o) differences

Fig. 7. The learnt expression category specific spatial patterns for posed and spontaneous expressions from the DISFA+ database.

expressions. They adopt SVM as a classifier. Their experiments were conducted on 3,572 posed and 1,472 spontaneous images from the NVIE database. Since selection criteria were not provided, we could not obtain identical samples. We show the results from their experiments for reference only. Pfister et al. [28] propose CLBP-TOP as well as a cascaded framework for differentiation between posed and spontaneous expressions. They conducted experiments using the SPOS database. The latter two studies use model-based techniques. They proposed multiple Bayesian

Table 4. Comparison with Related Works on the SPOS Database and the NVIE Database

	PS		PS_gender		PS_expression	
	accuracy (%)	F1 score	accuracy (%)	F1 score	accuracy (%)	F1 score
Comparison with related work on the SPOS database						
Pfister et al. [28]	72.00	/	/	/	/	/
Wang et al. [38]	74.79	**0.67**	/	/	/	/
Wang et al. [39]	76.07	0.64	/	/	/	/
Ours	**76.07**	0.64	/	/	/	/
Comparison with related works on the NVIE database						
L. Zhang et al. [44]	79.43	/	/	/	/	/
Wang et al. [38]	91.73	0.92	92.61	0.83	91.83	0.92
Wang et al. [39]	91.63	0.91	92.22	0.92	92.90	0.93
Ours	**98.74**	**0.99**	**99.32**	**0.99**	**98.35**	**0.98**

networks [38] and restricted Boltzmann machines [39] to capture inherent facial spatial patterns from feature points.

Table 4 shows the comparative results. It yields these observations:

First, for the PS model, the three model-based approaches achieve superior performance compared to the feature-driven techniques on all three databases. Although the latter works extract both appearance and geometric features, the model-based methods achieve better performance with geometric features only. From this, we can determine that the approaches based on models are successful at capturing and leveraging the spatial patterns inherent in posed and spontaneous expressions.

Second, when we examine the model-based methods, our proposed LRBN model performs best, achieving the highest F1 and accuracy scores in most cases. For the PS model on the NVIE database, the proposed LRBN model improves accuracy by 19.31% when compared to Pfister et al. [28], and it improves accuracy by 7.01% and F1 score by 0.07 when compared to Wang et al.'s work [38]. Our model improves accuracy by 7.11% and improves F1 score by 0.08 when compared to Wang et al.'s work [39]. On the SPOS database, the proposed LRBN improves accuracy by 4.07% and 1.28% compared to Pfister et al. [28] and Wang et al.'s work [38], respectively. On the NVIE database, the proposed PS_gender model improves accuracy by 6.71% and improves F1 score by 0.16 when compared to Wang et al.'s work [38], and it improves accuracy by 7.1% and F1 score by 0.07 when compared to Wang et al.'s work [39]. The proposed PS_expression model also achieves better performance than other methods on the NVIE database. Specifically, it outperforms Wang et al. [38] by 6.52% for accuracy and 0.06 for F1 score. It outperforms Wang et al. [39] by 5.45% for accuracy and 0.05 for F1 score. Unlike a BN, which is only able to model local rather than global dependencies among the variables, the proposed LRBN uses hidden units to obtain global probabilistic dependencies. Although an RBM can also represent global probabilistic dependencies among variables, hidden units are independent given the visible units. The proposed LRBN is able to leverage the probabilistic dependencies among latent variables given the observation. It is also able to capture global probabilistic dependencies among visible variables. These probabilistic dependencies are essential to faithfully represent the data, leading to the improved performance of the LRBN.

5 CONCLUSION

We propose several LRBNs designed to clearly model complex joint distributions over feature points, also called spatial patterns. These patterns are embedded in posed and spontaneous expressions and can be leveraged to distinguish between the two. Furthermore, considering the different

facial spatial patterns in gender and expression categories, we employ these categories as privileged information to enhance recognition performance. Specifically, we construct several LRBNs that are able to model embedded spatial patterns during training. During testing, the samples are classified as either posed or spontaneous according to the LRBN that has the largest likelihood. We also propose efficient learning and inference algorithms. The results of our experiments on three benchmark databases show the ability of the suggested models to capture spatial patterns, and demonstrate their superiority over existing techniques. Furthermore, the results on the NVIE and DISFA+ databases indicate that incorporating the privileged information, i.e., gender and expression categories, during training can help construct gender-specific and expression-specific spatial patterns, and are thus beneficial for posed and spontaneous expression distinction.

REFERENCES

[1] Hervé Abdi and L. J. Williams. 2010. Normalizing data. In *Encyclopedia of Research Design*. 935–938.

[2] Elisabeth Andre. 2013. Exploiting unconscious user signals in multimodal human-computer interaction. *ACM Trans. Multimedia Comput. Commun. Appl.* 9, 1s (2013), 48.

[3] Yoshua Bengio, Li Yao, and Kyunghyun Cho. 2014. Bounding the test log-likelihood of generative models. In *Proceedings of the International Conference on Learning Representations (Conference Track)*.

[4] George A. Bonanno and Dacher Keltner. 1997. Facial expressions of emotion and the course of conjugal bereavement. *J. Abnorm. Psychol.* 106, 1 (1997), 126.

[5] J. F. Cohn and K. L. Schmidt. 2004. The timing of facial motion in posed and spontaneous smiles. *Int. J. Wavelets Multires. Inf. Process.* 2, 02 (2004), 121–132.

[6] H. Dibeklioğlu, A. Salah, and T. Gevers. 2012. Are you really smiling at me? Spontaneous versus posed enjoyment smiles. In *Proceedings of the European Conference on Computer Vision (ECCV'12)*. Springer, 525–538.

[7] Hamdi Dibeklioglu, Roberto Valenti, Albert Ali Salah, and Theo Gevers. 2010. Eyes do not lie: Spontaneous versus posed smiles. In *Proceedings of the International Conference on Multimedia*. ACM, 703–706.

[8] Paul Ekman. 2003. Darwin, deception, and facial expression. *Ann. N. Y. Acad. Sci.* 1000, 1 (2003), 205–221.

[9] Paul Ekman and Wallace V. Friesen. 1982. Felt, false, and miserable smiles. *J. Nonverb. Behav.* 6, 4 (1982), 238–252.

[10] Paul Ekman, Joseph C. Hager, and Wallace V. Friesen. 1981. The symmetry of emotional and deliberate facial actions. *Psychophysiology* 18, 2 (1981), 101–106.

[11] Byron N. Fujita, Robert G. Harper, and Arthur N. Wiens. 1980. Encoding-decoding of nonverbal emotional messages: Sex differences in spontaneous and enacted expressions. *J. Nonverb. Behav.* 4, 3 (1980), 131–145.

[12] Quan Gan, Siqi Nie, Shangfei Wang, and Qiang Ji. 2017. Differentiating between posed and spontaneous expressions with latent regression Bayesian network. In *Proceedings of the Annual Cconference on Artifical Intelligence (AAAI'17)*. 4039–4045.

[13] Zhe Gan, Ricardo Henao, David Carlson, and Lawrence Carin. 2015. Learning deep sigmoid belief networks with data augmentation. In *Proceedings of the International Conference on Artificial Intelligence and Statistics* (2015).

[14] Karol Gregor, Andriy Mnih, and Daan Wierstra. 2014. Deep AutoRegressive networks. In *Proceedings of the 31st International Conference on Machine Learning* (2014).

[15] Geoffrey Hinton. 2010. A practical guide to training restricted Boltzmann machines. *Momentum* 9, 1 (2010), 926.

[16] Geoffrey Hinton and Ruslan Salakhutdinov. 2006. Reducing the dimensionality of data with neural networks. *Science* 313, 5786 (2006), 504–507.

[17] Geoffrey E. Hinton, Peter Dayan, Brendan J. Frey, and Radford M. Neal. 1995. The "wake-sleep" algorithm for unsupervised neural networks. *Science* 268, 5214 (1995), 1158–1161.

[18] Diederik P. Kingma and Max Welling. 2014. Auto-encoding variational bayes. In *Proceedings of the International Conference on Learning Representations (ICLR'14)*.

[19] C. Lithari, C. A. Frantzidis, C. Papadelis, A. B. Vivas, M. A. Klados, C. Kourtidou-Papadeli, C. Pappas, A. A. Ioannides, and P. D. Bamidis. 2010. Are females more responsive to emotional stimuli? A neurophysiological study across arousal and valence dimensions. *Brain Topogr.* 23, 1 (2010), 27–40.

[20] G. C. Littlewort, M. S. Bartlett, and K. Lee. 2009. Automatic coding of facial expressions displayed during posed and genuine pain. *Image Vis. Comput.* 27, 12 (2009), 1797–1803.

[21] Mohammad Mavadati, Peyten Sanger, and Mohammad H. Mahoor. 2016. Extended DISFA dataset: Investigating posed and spontaneous facial expressions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 1–8.

[22] Andriy Mnih and Karol Gregor. 2014. Neural variational inference and learning in belief networks. In *Proceedings of the 31st International Conference on Machine Learning* (2014).

[23] Shushi Namba, Shoko Makihara, Russell S. Kabir, Makoto Miyatani, and Takashi Nakao. 2017. Spontaneous facial expressions are different from posed facial expressions: Morphological properties and dynamic sequences. *Curr. Psychol.* 36, 3 (2017), 593–605.

[24] Siqi Nie, Yue Zhao, and Qiang Ji. 2016. Latent regression Bayesian network for data representation. In *Proceedings of the 23rd International Conference on Pattern Recognition (ICPR'16)*. IEEE, 3494–3499.

[25] E. Paul. [n.d.]. BBC-Dataset. Retrieved from http://www.bbc.co.uk/science/humanbody/mind/surveys/smiles/.

[26] S. Petridis, B. Martinez, and M. Pantic. 2013. The MAHNOB laughter database. *Image Vis. Comput.* 31, 2 (2013), 186–202.

[27] T. Pfister, X. Li, G. Zhao, and M. Pietikainen. 2011. Differentiating spontaneous from posed facial expressions within a generic facial expression recognition framework. In *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCV Workshops'11)*. IEEE, 868–875.

[28] Tomas Pfister, Xiaobai Li, Guoying Zhao, and Matti Pietikäinen. 2011. Differentiating spontaneous from posed facial expressions within a generic facial expression recognition framework. In *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCV Workshops'11)*. IEEE, 868–875.

[29] Danilo J Rezende, Shakir Mohamed, and Daan Wierstra. 2014. Stochastic backpropagation and approximate inference in deep generative models. In *Proceedings of the 31st International Conference on Machine Learning (ICML'14)*. 1278–1286.

[30] Fabien Ringeval, Björn Schuller, Michel Valstar, Jonathan Gratch, Roddy Cowie, and Maja Pantic. 2018. Introduction to the special section on multimedia computing and applications of socio-affective behaviors in the wild. *ACM Trans. Multimedia Comput. Commun. Appl.* 14, 1s (2018), 25.

[31] Herbert Robbins and Sutton Monro. 1951. A stochastic approximation method. *Ann. Math. Stat.* 22, 3 (1951), 400–407.

[32] Lawrence K. Saul, Tommi Jaakkola, and Michael I. Jordan. 1996. Mean field theory for sigmoid belief networks. *J. Artif. Intell. Res.* 4, 61 (1996), 76.

[33] M. Seckington. 2011. Using dynamic Bayesian networks for posed versus spontaneous facial expression recognition. *Master's Thesis, Department of Computer Science, Delft University of Technology* (2011).

[34] M. F. Valstar, M. Pantic, Z. Ambadar, and J. F. Cohn. 2006. Spontaneous vs. posed facial behavior: Automatic analysis of brow actions. In *Proceedings of the 8th International Conference on Multimodal Interfaces*. ACM, 162–170.

[35] V. Vapnik and A. Vashist. 2009. A new learning paradigm: Learning using privileged information. *Neur. Netw.* 22, 5–6 (2009), 544.

[36] Shangfei Wang, Longfei Hao, and Qiang Ji. 2019. Facial action unit recognition and intensity estimation enhanced through label dependencies. *IEEE Trans. Image Process.* 28, 3 (2019), 1428–1442.

[37] Shangfei Wang, Zhilei Liu, Siliang Lv, Yanpeng Lv, Guobing Wu, Peng Peng, Fei Chen, and Xufa Wang. 2010. A natural visible and infrared facial expression database for expression recognition and emotion inference. *IEEE Trans. Multimedia* 12, 7 (2010), 682–691.

[38] Shangfei Wang, Chongliang Wu, Menghua He, Jun Wang, and Qiang Ji. 2015. Posed and spontaneous expression recognition through modeling their spatial patterns. *Mach. Vis. Appl.* (2015), 1–13.

[39] Shangfei Wang, Chongliang Wu, and Qiang Ji. 2016. Capturing global spatial patterns for distinguishing posed and spontaneous expressions. *Comput. Vis. Image Understand.* 147 (2016), 69–76.

[40] Chongliang Wu and Shangfei Wang. 2016. Posed and spontaneous expression recognition through restricted boltzmann machine. In *MultiMedia Modeling*. Springer, 127–137.

[41] Alan L. Yuille. 2005. The convergence of contrastive divergences. In *Advances in Neural Information Processing Systems*. 1593–1600.

[42] S. Yunus and T. Christopher. 2006. Cascaded classification of gender and facial expression using active appearance models. In *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR'06)*. 393–400.

[43] Feifei Zhang, Qirong Mao, Xiangjun Shen, Yongzhao Zhan, and Ming Dong. 2018. Spatially coherent feature learning for pose-invariant facial expression recognition. *ACM Trans. Multimedia Comput. Commun. Appl.* 14, 1s (2018), 27.

[44] L. Zhang, D. Tjondronegoro, and V. Chandran. 2011. Geometry vs. appearance for discriminating between posed and spontaneous emotions. In *Neural Information Processing*. Springer, 431–440.

Received March 2019; revised March 2020; accepted March 2020