International Journal of Automation and Computing 17(5), October 2020, 637-651
DOI: 10.1007/s11633-020-1244-1

DiscoStyle: Multi-level Logistic Ranking for Personalized Image Style Preference Inference

Zhen-Wei He Lei Zhang Fang-Yi Liu

School of Microelectronics and Communication Engineering, Chongqing University, Chongqing 400044, China

Abstract: Learning based on facial features for detection and recognition of people′s identities, emotions and image aesthetics has been widely explored in computer vision and biometrics. However, automatic discovery of users′ preferences to certain of faces (i.e., style), to the best of our knowledge, has never been studied, due to the subjective, implicative, and uncertain characteristic of psychological preference. Therefore, in this paper, we contribute to an answer to whether users′ psychological preference can be modeled and computed after observing several faces. To this end, we first propose an efficient approach for discovering the personality preference related facial features from only a very few anchors selected by each user, and make accurate predictions and recommendations for users. Specifically, we propose to discover the style of faces (DiscoStyle) for human′s psychological preference inference towards personalized face recommendation system/application. There are four merits of our DiscoStyle: 1) Transfer learning is exploited from identity related facial feature representation to personality preference related facial feature. 2) Appearance and geometric landmark feature are exploited for preference related feature augmentation. 3) A multi-level logistic ranking model with on-line negative sample selection is proposed for on-line modeling and score prediction, which reflects the users′ preference degree to gallery faces. 4) A large dataset with different facial styles for human′s psychological preference inference is developed for the first time. Experiments show that our proposed DiscoStyle can well achieve users′ preference reasoning and recommendation of preferred facial styles in different genders and races.

Keywords: Facial preference, feature representation, logistic regression, face recommendation, transfer learning.

1 Introduction

Facial features, as one kind of important biometrics, can explicitly and implicitly represent the objective and subjective facial attributes (e.g., eyes, nose, mouth) and personal characteristics (e.g., identity, age, gender, races, emotion, beauty, personal character and hobbies). Learning facial features for detection and recognition of person identity, age, gender, races, expression, emotion, and beauty has been widely developed in computer vision and biometrics[1–10] which has also greatly promoted the industrial applications of artificial intelligence. Currently, face recognition has been used in security inspection, access control system, video surveillance, etc. Additionally, age, emotion and beauty analysis have been used for multimedia, social and internet interaction. However, to the best of our knowledge, there is no research on exploration of a user′s psychological and emotional preference to different facial image styles towards recommendation applications.

Generally, it is very challenging to infer and reason about the implicit, fine-grained, subjective and common facial preference features that attract users from a very few selected face images of different styles (i.e., anchors) by the users. That is, if we could discover the facial preference features that the user internally and subjectively pays more attention to, then we can compute and predict the user′s personality preference via probabilistic models. After all, interesting applications with advanced emotional analysis, robot services[11], and automatic personalized image recommendation can be promoted by the discovered facial preference characteristic. It is worth noting that there have been a number of research works in facial beauty and attractiveness prediction, which, however, is essentially different from the proposed user specific preference inference and recommendation of different facial image styles in the following aspects.

1) Facial preference is less relevant to facial beauty that can be modeled with a universal criterion[12, 13], while preference is user-specific and highly relevant to external facial styles (e.g., hairstyle, eye, nose, lips, glass).

2) Facial preference is also relevant to internal character reflected from faces (e.g., temperament, lovely, elegant), which is also user specific and even comprehensive for modeling users′ preference.

3) Due to the person-specific property of facial preference, the preference model parameters are dynamic and vary from person to person, while the facial beauty model is generally fixed and not person-specific. In other words, a face of highly beauty does not mean a high degree of a user′s preference, due to the users′ emotional dif-

Research Article
Manuscript received March 2, 2020; accepted June 30, 2020; published online September 9, 2020
Recommended by Associate Editor Bin Luo

ference.

Deep learning (DL)[14–16], as a kind of supervised learning method originated from large-scale image recognition, has witnessed a huge success in multiple vertical fields, such as computer vision, pattern recognition, text analysis and speech recognition. Recently, transfer learning (TL)[17–21], as a weakly-supervised cross-domain learning technique, has successfully promoted the horizontal development of DL in learning methodologies and applications. With the seamless connection between the supervised DL and the weakly-supervised TL, there is no doubt that DL and TL greatly stimulate the progress of artificial intelligence in many horizontal weakly-supervised research areas, such as medical image analysis[22,23], remote sensing image analysis[24,25], satellite image analysis[26,27], kinship verification[28,29], computer vision[30,31], load forecasting[32], fault diagnosis[33,34], etc. Generally, DL aims to obtain a universal and generalized knowledge representation model in a supervised manner, while TL aims to connect and propagate the DL knowledge to more weakly-supervised domains and tasks w/o fine-tune or partial re-training, where the data and labels are not completely or accurately prepared and deployed.

In this paper, as shown in Fig. 1, we are dedicated to the inference and reasoning analysis and modeling of a user′s psychological preference of facial image styles based on very few selected anchors (e.g., 10 images) by the user, which is undoubtedly a subjective, implicative, and weakly-supervised task. Therefore, a DL and TL inspired preference feature representation method is exploited for knowledge transfer from a large-scale supervised face recognition task to a single user-specific weakly-supervised face preference reasoning task. Further, probabilistic learning is used in the reasoning stage based on very few selected anchor faces (labeled as preferred faces by a user). Then, the model can be used to compute the psychological preference score for each gallery facial image, and the score value can successfully represent the degree of a user′s preference with respect to each gallery face. The main contributions of this paper are four-fold.

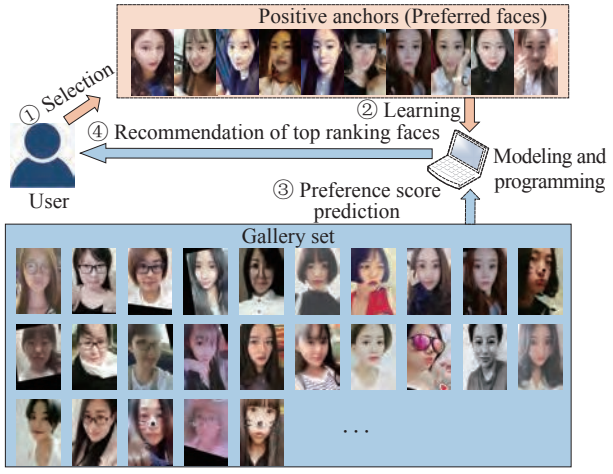


Fig. 1 User-specific facial preference reasoning and prediction tasks, which shows the user′s preference reasoning and recommendation to female faces. Notably, the anchors represent the preferred faces selected by users. Color versions of the figures in this paper are available online.

1) We propose an efficient DiscoStyle approach for user-specific facial preference reasoning and computation, by looking at very few anchors with more glances, which achieves automatic preference prediction and recommendation, which, to the best of our knowledge, is the first work for users′ preference and face recommendation.

2) A deep transfer learning paradigm is proposed for facial preference related feature representation, based on a pre-trained face representation deep network, which comprehensively integrates the appearance features and geometric landmark feature for fully reflecting the facial style.

3) A multi-level logistic ranking (MLR) model with a novel on-line negative sample selection (ONSS) strategy is proposed in DiscoStyle for preference reasoning, which can predict the preference score and objectively define one user′s degree of preference to each gallery face and the faces of high preference degree are recommended.

4) A large facial style dataset (i.e., StyleFace) is developed for the first time for facial preference prediction, which includes a facial style subset for style attribute vector learning, an anchor subset for probabilistic reasoning and a gallery subset for preferred faces recommendation.

The rest of paper is organized as follows. In Section 2, we review the related work in deep learning and face analysis. In Section 3, we present the proposed DiscoStyle framework with feature representation, negative sample selection and preference reasoning model. In Section 4, the experiments with the developed dataset, evaluation results and discussions are presented. Finally, Section 5 concludes this paper.

2 Related work

Our DiscoStyle approach involves deep feature representation learning and face modeling. Therefore, in this section, we first briefly review the deep learning methods, and then the work related to face modeling and face analysis are presented.

2.1 Deep learning

Inspired by the depth of the biological brain[35], in recent years, deep learning has attracted much attention and neural networks get deeper and deeper for achieving better feature representation, especially in computer vision tasks[36–39]. In early work, Krizhevsky et al.[14] proposed a convolutional neural network (CNN) of 8 layers and achieved a great success in large-scale image classification task (i.e., ImageNet), which is an essential problem in computer vision. Since then, very deep CNN struc-

tures such as ResNet and DenseNet have been proposed[15,40,41] for state-of-the-art competition performance.

Benefiting from the great success in image classification task, deep learning has also been used in many other computer vision tasks, such as object detection, face recognition, video surveillance, etc. For object detection task, some typical techniques such as faster-RCNN[42], single-shot detector (SSD)[43] and you only look once (YOLO)[44] have shown greater detection performance than traditional handcrafted features. For face recognition tasks, as a very important issue, deep learning has also achieved a better accuracy than human in particular datasets. Schroff et al.[1] proposed a triplet loss based CNN structure, Sun et al.[3] proposed a joint classification and verification loss supervised CNN for face recognition. Deep learning has also achieved great success in other computer vision tasks, such as pose estimation[37,45], person re-identification[46,47].

Upon the significant progress of deep learning, the strength of DL lies in its high-level feature representation ability. To this end, the feature representation of DL is also used in our DiscoStyle approach for implicit facial feature description that can reflect users′ preference.

2.2 Face analysis

Human faces contain a large amount of personal information such as identity, age, expression and emotion, which can be distinguished by facial images. Face recognition is a basic but important task in artificial intelligence, which is widely used in many areas and scenes, such as public security, law enforcement, commercial contexts, etc. Besides the recognition problem, face aging[48,49] is another important task which is widely used in cross-age verification or searching for a missing child[50]. Different from face identification, kinship verification, which aims to mine implicit kin-relations from facial images, has also been widely studied[28,51–53]. Facial beauty analysis toward attractiveness assessment application, as an emerging topic, has also attracted an amount of research[6,12,13,54]. Benefiting from the rich information carried by human faces, all of these topics have achieved great success in recent years. However, to our best knowledge, users′ preference analysis of a facial image and face recommendation have never been studied. To this end, we propose a novel face analysis work, DiscoStyle, for users′ preference prediction towards face recommendation from a gallery face database.

3 Proposed DiscoStyle approach

In the proposed DiscoStyle approach, four key stages are included: 1) preference oriented facial feature representation (PFR) for appearance and geometric feature extraction; 2) on-line negative samples selection (ONSS) for selective negative anchors; 3) multi-level logistic ranking (MLR) based probabilistic reasoning for users′ preference prediction; 4) on-line preference score computation (PSC) of galleries for face recommendation. The process of the proposed DiscoStyle framework for a user can be described in the following steps:

1) In the feature extraction stage, the deeply represented appearance features that reflect the abstract facial representation and landmark geometric features that represent the facial shape jointly formulate the PFR module.

2) In the negative sample selection stage, the proposed ONSS algorithm is used to find a few negative anchors (the non-preferred faces of users) from the gallery face database for each user, such that the positive anchors and negative anchors can be used for subsequent learning.

3) In the preference reasoning stage, the proposed on-line MLR model is trained on the positive and negative anchors of each user for user-specific model parameter learning.

4) In the on-line prediction and recommendation stage, the gallery face database is scored by the user-specific PSC via the model parameters, and the galleries with the highest scores are recommended for the user.

3.1 Preference oriented facial feature representation

Preference oriented facial feature representation (PFR) aims at extracting the abstract and implicit features that the user most likely pays attention to. With a full survey on different users′ attention to their preferred faces, we summarize some universal preference related characteristics including explicit attribute features (e.g., hair style, face type, eyes, nose, mouth, lips, skin color) and implicit features (e.g., temperament type, lovely type, elegant type, gentle type). Specifically, for explicit but objective features, facial appearance features and geometric landmark feature are proposed. For implicit but subjective features, we established two groups of labeled datasets for facial style classification and the classification scores are used as implicit features.

3.1.1 ROI detection

Considering that the preference is not only related to face regions but also hair style, therefore, the region of interest (ROI) detection in this work is around the faces from head to the neck as shown in Fig. 2 (the aligned facial image), which can provide the hair style information in facial preference reasoning. Technically, the ROI detection is implemented based on the multi-task cascaded convolutional networks (MT-CNN) framework[55], which shows clearly the face alignment and detection but not the ROI detection in this paper. Therefore, we obtain the ROI by adaptively enlarging the facial bounding box of MT-CNN.

3.1.2 Appearance feature

The face recognition problem aims at distinguishing

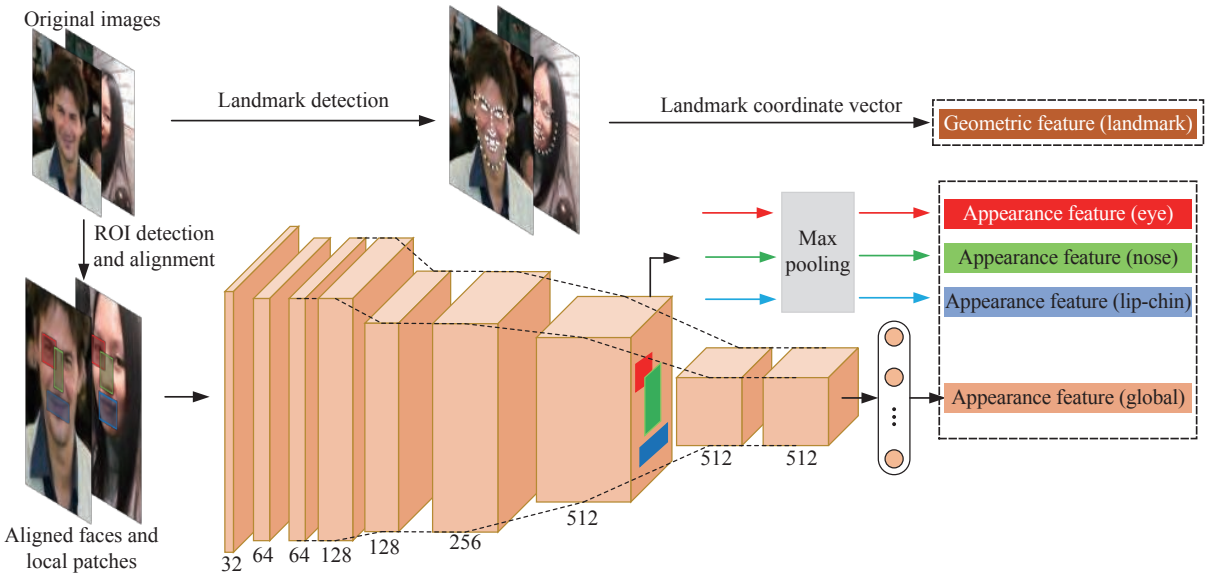


Fig. 2 The paradigm for our PFR method for feature representation, which shows the ROI detection and alignment, appearance feature extraction with three local regions in red (eye part), green (nose part) and blue (lip-chin part), and the geometric landmark coordinate feature vector (shape). In total, five kinds of features including 1 geometric feature, 3 local appearance features and 1 global appearance feature.

different faces with different identity, which depends on the high-level and abstract appearance features of faces. As we know that deep feature representation that reflects the discriminative identity characteristic of the face has greatly benefitted to the face recognition tasks. To this end, in this work, the deep feature representation network established for face recognition is used in our DiscoStyle framework. Appearance feature is closely related to the basic texture and local description of faces. In this paper, for appearance feature extraction, the global and local representations are jointly considered, which are shown as follows.

1) The global appearance feature in the ROI region is extracted by using an off-the-shelf CNN, which is trained from scratch on a large-scale CASIA (Institute of Automation, Chinese Academy of Sciences) WebFace dataset[56]. The process for feature representation in this work is deployed in Fig. 2. The deep feature achieves a global representation of the appearance feature in ROI region.

2) Additionally, consider the local parts sensitivity of users' preference for a face, the fine-grained features that users most likely to pay attention to have also been specially extracted, including the eye part (red), nose part (green) and lip-chin part (blue) as shown in Fig. 2 (the aligned facial image). In our experiment, the feature map in the last convolutional layer of CNN with respect to the three local parts is formulated, respectively, for local part representation. Also, the high-level and discriminative feature vector in the last fully-connected layer is used for global representation. In total, there are 4 kinds of appearance features (i.e., eye, nose, lip-chin, and global) jointly learned in subsequent modeling. The specific structure of our CNN framework is shown in Table 1, in which the global representation feature in FC1 layer is extracted in our model and the local appearance features (eye, nose, lip-chin) in Conv4_x layer are extracted with max pooling for feature vector formulation.

Table 1 Convolutional neural network architecture of our PFR part

Layers	Filters	Output size
Conv1_x	$3\times3, 32$ $3\times3, 32$	108×92
Conv2_x	$\begin{bmatrix}3\times3, 64\\3\times3, 64\end{bmatrix}\times1$ $3\times3, 128$	52×44
Conv3_x	$\begin{bmatrix}3\times3, 128\\3\times3, 128\end{bmatrix}\times2$ $3\times3, 256$	24×20
Conv4_x	$\begin{bmatrix}3\times3, 256\\3\times3, 256\end{bmatrix}\times5$ $3\times3, 512$	10×8
Conv5_x	$\begin{bmatrix}3\times3, 512\\3\times3, 512\end{bmatrix}\times3$	5×4
FC1	512 (feature dimension)	
FC2	10 575 (IDs)	
Softmax	—	

3.1.3 Geometric landmark feature

In the appearance feature, the geometric information that fully describes the shape of faces is missed out. However, the geometric feature is also important explicit

and objective characteristics for face style learning in facial attractiveness study[6], which is also a factor related to a user's preference for a face. Therefore, in this paper, by following the findings in [6], 68 landmark points are detected by the ensemble of regression tree (ERT) algorithm[57] and the resulting 136-dimensional coordinate vector (68×2) is used as the geometric feature for preference reasoning. In order to normalize the coordinate vector of face pictures of different sizes, the width and height of the detected bounding box are used as the denominator in ratio normalization. To this end, in total 5 kinds of normalized features including 1 kind of global appearance feature, 3 kinds of local part appearance feature and 1 kind of geometric feature are formulated as the input of our DiscoStyle framework.

3.2 On-line negative sample selection algorithm

In real-world face recommendation systems, the general application scene is that the preferred faces (i.e., positive anchors) are available because of the users' independent choice when they are connecting the Internet. However, the non-preferred faces may not be obtained due to the users' uncertainty. Therefore, for subsequent reasoning and learning, we propose an on-line negative sample selection (ONSS) algorithm for selecting the non-preferred anchors, which greatly benefit the subsequent learning and reasoning task.

In our proposed ONSS algorithm, two important aspects are considered: 1) on-line selection; 2) fast selection. Specifically, on-line selection is owing to the user-specific preference reasoning model characteristic, because each user's selected preferred faces (positive anchors) in their on-line internet access are also different. Fast selection is owing to the requirement of real-time preference reasoning and face recommendation. Since that our ONSS is Euclidean distance based, it is not allowed to find the potential negative samples by traversing the whole gallery database that can be very large in quantity in real-world application.

Therefore, we propose a guessing strategy, by on-line guessing the possible negative samples in the real-time streaming data with an efficient voting mechanism based on five kinds of features. Once the required quantity (i.e., the same as the quantity of positive anchors) of negative samples is achieved, the ONSS algorithm is automatically terminated. The schematic of the proposed ONSS algorithm is visually shown in Fig. 3. Suppose that there are S feature modalities, then the voting number V $(0 \leq V \leq S)$ of the gallery face x being a negative sample is computed as

$$V = \sum_{i=1}^{S} L(\boldsymbol{x}_i - \boldsymbol{c}_{i2} > R_i)$$
$$\text{s.t.} \quad \boldsymbol{c}_i = \frac{1}{n}\sum_{j=1}^{n} \boldsymbol{p}_i^j \tag{1}$$

where $\boldsymbol{x}_i$ is the i-th feature modality (appearance feature or geometric feature) of the query picture, $\boldsymbol{p}_i^j$ stands for the i-th feature from the j-th positive anchor selected by users, and $\boldsymbol{c}_i$ represents the center of the i-th feature modality. $R_i = \max(\|\boldsymbol{p}_i^j - \boldsymbol{c}_i\|_2),\ j = 1, \cdots, 10$ is the threshold with respect to the i-th feature modality computed as the farthest distance value between the positive anchor $\boldsymbol{p}_i^j$ and the center $\boldsymbol{c}_i$. $L(\cdot)$ is an indicator function, where $L(f) = 1$ if f is true, otherwise, $L(f) = 0$. In this paper, the gallery face x is recognized to be a negative face if $V/S > 0.5$. Since the centers of each kind of feature modality can be easily pre-computed on-line, our ONSS algorithm is self-terminated if only the required number of negative samples is achieved. Specifically, the implementation process of the proposed ONSS is shown in Algorithm 1.

Algorithm 1. The proposed ONSS

Input: Gallery faces set $\boldsymbol{X}$, the feature subset $\boldsymbol{p}_i$ of the i-th feature modality with respect to the m positive anchors (preferred faces) selected by users, the number S of feature modalities and the required number m of the

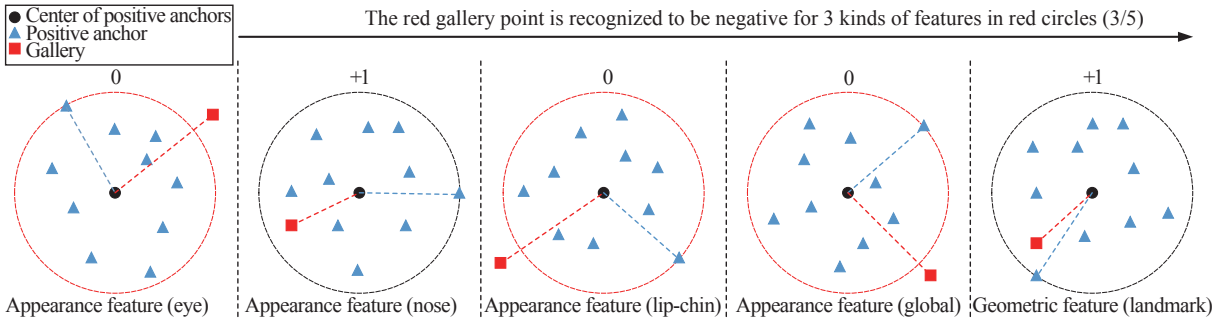


Fig. 3 The schematic of our proposed ONSS approach, which shows how to select the potential negative samples (i.e., non-preferred faces) on-line without traversing all the gallery faces, because in real-application the gallery faces are infinity. In this figure, the gallery face (red square) is recognized to be potential negative sample, because it shows the maximum distance to the center of positive anchors in red circle with a higher proportion $(3/5 > 0.5)$. That is, the gallery face is recognized to be negative sample under 3 feature modalities rather than 2. Notably, for balancing between positive samples (i.e., preferred faces or anchors) and negative samples (i.e., non-preferred faces), the ONSS program is automatically stopped when the number of selected negative samples achieves to the number of anchors.

potential negative samples.

Output: Negative anchor set Θ of m negative samples.

1: **Initialization.** Compute the R_i and c_i, the number n of the selected negative samples is set as $n = 0$, and the set of negative samples is defined as $\Theta = \varnothing$;

2: **repeat**

3: Take a gallery face x from the $\boldsymbol{X}$;

4: The i-th modality feature representation $\boldsymbol{x}_i$ of x is based on the proposed PFR method in Section 3;

5: Compute V of the gallery face x using (1);

6: **if** $V/S > 0.5$ **then**

7: $x \in \Theta$;

8: $n = n + 1$;

9: **end if**

10: **until** $n = m$

3.3 Multi-level logistic ranking (MLR) based reasoning

In this paper, we propose a multi-level logistic ranking model for on-line user preference reasoning and face recommendation. Considering the user-specific characteristic of preference, the MLR reasoning and recommendation should be implemented on-line based on a very few positive anchors (i.e., preferred faces selected by a user) and the same number of negative anchors (i.e., non-preferred faces selected by ONSS).

3.3.1 Facial style coarse score prediction

The appearance and geometric features are explicit and objective characteristics of faces. However, for the style attribute, it is implicit but subjective that it may be different from person to person. That is, the style attribute score cannot be directly reflected from the image, but depends more on the automatic evaluation score of each style, computed by using pre-trained coarse classifiers such as support vector machine (SVM), multilayer perceptron (MLP), based on the appearance feature. Therefore, in this paper, we establish two groups of dataset with different styles.

1) Group 1: A facial subset of 3 styles (long versus short hair style, lovely versus temperament type, round versus thin face type);

2) Group 2: A facial subset of 5 styles (single versus double eyelids, pointed versus round chin, long versus short hair style, big versus small mouth, straight versus flat nose).

For the former 3-style subset, three binary SVM classifiers are trained. For the latter 5-style subset, a multi-label multilayer perceptron model is trained. For the 3-style model, for each gallery face, a score value (i.e., 0, 1/3, 2/3 or 1) is calculated for each style based on the anchors and binary classifiers, and a 3-dimensional score vector $\boldsymbol{S}_{svm}$ is formulated. For the 5-style model, the 5-dimensional multi-label output of MLP model is recognized as the score vector defined as $\boldsymbol{S}_{mlp}$. Note that the score values from SVM and the score vector from MLP are used in the 2nd level LR model together with the 5 logistic ranking scores from the 1st level LR model for ultimate preference reasoning, as shown in Fig. 4. Specially, the principle of MLR model is formulated as follows.

3.3.2 Details of MLR model

As we know, user′s preference for a face cannot be simply recognized as a classification or recognition problem with coarse labels. Instead, user′s preference modeling should be a confidence estimation problem. Logistic regression is a typical probabilistic model for revealing the relation and importance among variables, which tends to provide a probability for an instance $\boldsymbol{x}$ with a group of variables $[x_1, x_2, \cdots, x_d]$ to be positive or negative. In this paper, based on the logistic regression model, we propose a multi-level logistic ranking model for revealing the user′s degree of preference to a gallery face $\boldsymbol{x}$ and providing the probability score ranking given a gallery set.

Suppose that a random variable ξ comes from a binomial distribution (0−1 distribution or Bernoulli distribution), i.e., $\xi \sim B(0, 1)$, and $p(\xi)$ represents the probability density function (PDF). Let $P(Y|X)$ be the conditional probability of response Y given an input variable X, then there is

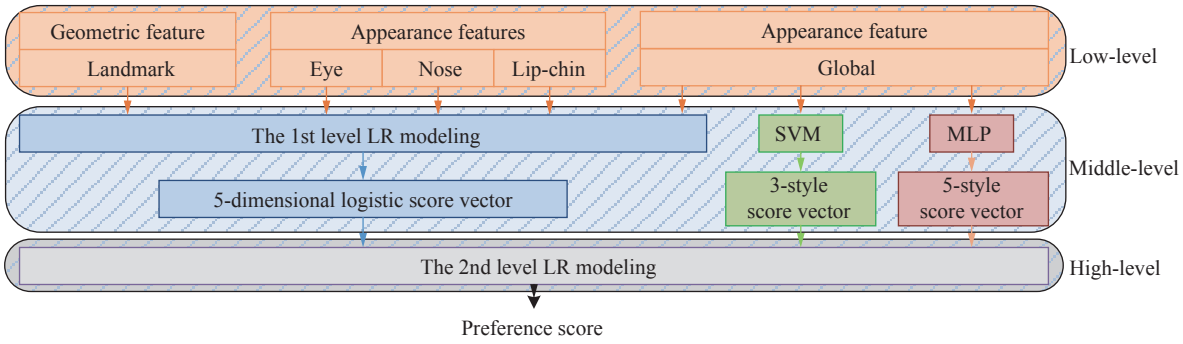


Fig. 4 The flowchart of our MLR model. The geometric features and appearance features are fed into the logistic regression model respectively in the 1st LR level. The global appearance features are also used for SVM and MLP to get 3-style score and 5-style score vector. All results of the 1st LR level are then fed into the 2nd LR level, and we can get the final preference score of the input face.

$$P(Y=1|X=\boldsymbol{x})=p(\boldsymbol{x};\boldsymbol{w})$$
$$P(Y=0|X=\boldsymbol{x})=1-p(\boldsymbol{x};\boldsymbol{w}) \quad (2)$$

where $\boldsymbol{w}$ represents the model parameter being learned.

According to (2), given an observation $\boldsymbol{x}_i$ (a gallery), the conditional probability of $Y=y_i$ can be written as

$$P(Y=y_i|X=\boldsymbol{x}_i)=p(\boldsymbol{x}_i;\boldsymbol{w})^{y_i}(1-p(\boldsymbol{x}_i;\boldsymbol{w}))^{1-y_i} \quad (3)$$

where $y_i=1$ if $\boldsymbol{x}_i$ is true (i.e., preferred face), otherwise, $y_i=0$ (i.e., non-preferred face).

Further, for preference reasoning, we focus on the likelihood of the whole dataset. For convenience, we impose a hypothesis of independence of N observations, then, the joint PDF (i.e., data likelihood) based on the instance-level likelihood function in (3) can be easily formulated as

$$\prod_{i=1}^{N}P(Y=y_i|X=\boldsymbol{x}_i)=\prod_{i=1}^{N}p(\boldsymbol{x}_i;\boldsymbol{w})^{y_i}(1-p(\boldsymbol{x}_i;\boldsymbol{w}))^{1-y_i}. \quad (4)$$

Through the log-likelihood maximization of (4), the model parameter $\boldsymbol{w}$ can be easily solved. Then, with the activation of sigmoid function, the logistic probabilistic score $P(y=1|\boldsymbol{x})$ of the gallery face $\boldsymbol{x}$ belonging to a preferred face can be computed as

$$P(y=1|\boldsymbol{x})=p(\boldsymbol{x};\boldsymbol{w})=\frac{1}{1+\mathrm{e}^{-h(\boldsymbol{x})}} \quad (5)$$

where $h(x)$ is the linear representation of the variables $[1,x_1,\cdots,x_d]^{\mathrm{T}}$ of the gallery face x by the model parameter $w=[w_0,\cdots,w_d]^{\mathrm{T}}$, which can be written as

$$h(\boldsymbol{x})=w_0+w_1x_1+\cdots+w_dx_d=\boldsymbol{w}^{\mathrm{T}}x \quad (6)$$

Generally, from (6), MLR is a linear model for prediction.

The multi-level LR model consists of two level LR learning and computation, which are shown as follows.

1) The 1st level LR

In the 1st level, five feature modalities including 3 kinds of local part appearance features, 1 kind of global appearance feature and 1 kind of geometric feature of anchors (i.e., 10 positive anchors and 10 negative anchors) are fed into the LR model in (4), respectively, for parameter learning. Then, 5 scores for each anchor are computed and concatenated as a 5-dimensional score vector, as shown in Fig. 4. For convenience, the expression of the probability score for each feature can be summarized as

$$(s_1,\boldsymbol{w}_1^1)=\arg\max MLR(\boldsymbol{x}_{landmark},y)$$
$$(s_2,\boldsymbol{w}_1^2)=\arg\max MLR(\boldsymbol{x}_{eye},y)$$
$$(s_3,\boldsymbol{w}_1^3)=\arg\max MLR(\boldsymbol{x}_{nose},y)$$
$$(s_4,\boldsymbol{w}_1^4)=\arg\max MLR(\boldsymbol{x}_{lip-chin},y)$$
$$(s_5,\boldsymbol{w}_1^5)=\arg\max MLR(\boldsymbol{x}_{global},y) \quad (7)$$

where $\boldsymbol{x}_{eye},\boldsymbol{x}_{nose},\boldsymbol{x}_{lip-chin},\boldsymbol{x}_{global}$ stand for the appearance feature of eye, nose, lip-chin and global, respectively, $\boldsymbol{x}_{landmark}$ is the geometric feature formulated by the coordinate vector, and y represents the label of anchors ($y=1$ for positive anchors, otherwise, $y=0$). By concatenating the five scores together, a 5-dimensional score vector that would be fed into the 2nd level can be formulated as

$$\boldsymbol{S}_{mlr}=[s_1,s_2,s_3,s_4,s_5]. \quad (8)$$

2) The 2nd level LR

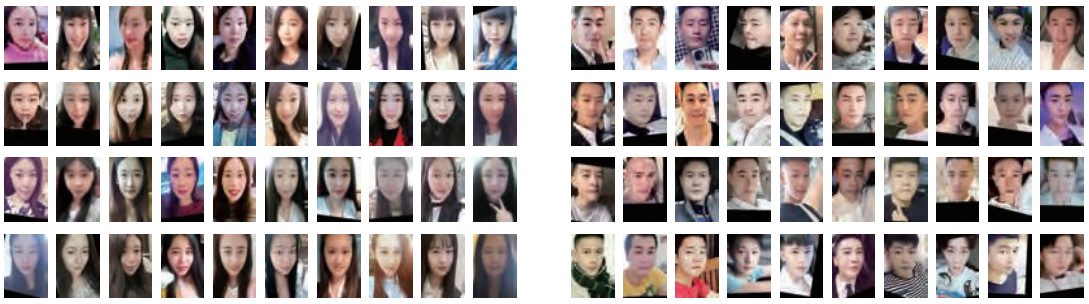
For the 2nd level, the input is the concatenated final score vector $\boldsymbol{S}$ by the three score vectors $\boldsymbol{S}_{svm}$, $\boldsymbol{S}_{mlp}$, and $\boldsymbol{S}_{mlr}$ in (8), shown as

$$\boldsymbol{S}=[\boldsymbol{S}_{mlr},\boldsymbol{S}_{svm},\boldsymbol{S}_{mlp}]. \quad (9)$$

Then, the MLR is retrained for learning the 2nd level parameter $\boldsymbol{w}_2$, and there is

$$\boldsymbol{w}_2=\arg\max MLR(\boldsymbol{S},y). \quad (10)$$

The flowchart of the proposed DiscoStyle framework is clearly shown in Fig. 4, and the procedure of DiscoStyle is described in Algorithm 2. The score ranking is then used to guide the face recommendation. Visually, we provide the score ranking of two users as shown in Fig. 5.



(a) Female images (b) Male images

Fig. 5 Examples of the preferred faces from two users in our StyleFace dataset. For each user, 10 anchors (the first row) are used for training and the remaining 30 faces (the last three rows) are used for testing. All images have been detected and aligned.

4 Experiments

In order to study the facial preference prediction problem by using the proposed DiscoStyle, we develop a preference dataset of eastern faces from the Internet including male and female faces, which is, to the best of our knowledge, the first facial style dataset in the world. For convenience, we define the dataset as StyleFace in this work. In the evaluation stage, several basic methods are exploited on our StyleFace dataset as baselines, and a comparison to our DiscoStyle framework is conducted. Then, the visualization of the proposed method is shown for analysis and discussion. Further, we test our method on the public face recognition dataset of western faces, i.e., labeled faces in wild (LFW)[58].

Algorithm 2. The proposed DiscoStyle approach

Input: Gallery faces set $\boldsymbol{X}$, the positive anchors $\boldsymbol{P}$ selected by a user, the 3-style subset 1 and the 5-style subset 2.

Output: The preference scores of the gallery faces $\boldsymbol{X}$ with respect to the user and the recommended faces.

1: **Feature representation via PFR.** 1) Compute the features of the gallery faces set $\boldsymbol{X}$ and the positive anchors $\boldsymbol{P}$ by using the PFR method; 2) Compute the appearance features of the 3-style and 5-style subsets.

2: **Negative anchors selection via ONSS.** Based on the positive anchors $\boldsymbol{P}$, the negative samples are selected from the gallery set by using the ONSS Algorithm 1.

3: **The 1st level MLR.** MLR is learned on 5 feature modalities of the anchors, respectively, and 5 parameter sets $\boldsymbol{w}_1^1, \cdots, \boldsymbol{w}_1^5$ are obtained. Finally, 5-dimensional score vector is resulted by using (7) and (8).

4: **Facial style coarse score prediction.** SVM and MLP are trained on the global appearance feature of 3-style and 5-style subsets, respectively, for 3-dimensional and 5-dimensional coarse style score vector computation.

5: **The 2nd level MLR.** MLR is learned by feeding the concatenated score vector (13-dimensional) in Steps 3 and 4 as input for ultimate reasoning of $\boldsymbol{w}_2$.

6: **Face recommendation.** Compute the probability scores of the gallery set using (5) and (6) based on the model parameter $\boldsymbol{w}_2$ from Step 5. The faces with top preference scores are recommended for the user.

4.1 Description of our developed StyleFace dataset

We collect and annotate a large-scale StyleFace dataset for the facial preference reasoning task. Without loss of generality, both male and female faces are studied, respectively, for reasoning and recommendation. All the images in the StyleFace dataset are collected from the Internet (e.g., Baidu, Momo). In total, 6 055 facial images with different sizes and scales including 3 028 female images and 3 027 male images are collected in our StyleFace dataset. Then, the users are asked to annotate the facial images. A volunteer who is invited as our user is required to annotate the faces based on their preference. In total, 106 volunteers (74 for the female images and 32 for the male images) are invited to participate the labeling stage and their preferred faces are specially labeled for reliable modeling. Specifically, 40 images after preference ranking by each user are finally selected as their preferred faces. That is, in the developed dataset, for each user, 40 preferred faces will be divided into training and testing sets for modeling and evaluation. In this paper, the first 10 faces are defined as training set (positive anchors), and the remaining 30 samples are used as test set. Note that, in the labeling process, we are not asking the users to label negative faces (i.e., their non-preferred faces) because of ethics.

Visually, the labeled 40 preferred facial images for female and male by a user are shown in Figs. 5(a) and 5(b), respectively, from which we can observe the users′ preference of facial styles. The claim that the facial preference model should be user-specific is further confirmed, due to the user′s personality difference. Then, customized recommendation is feasible.

4.2 Evaluation protocol and metrics

In this section, the experimental protocol and the evaluation metrics are clearly described.

4.2.1 Evaluation protocol

In experiments, the training process of the model parameters is conducted on the first 10 faces out of the 40 preferred faces by each user and the automatically selected 10 negative faces by the proposed ONSS algorithm. The training process follows the procedure in Algorithm 2. In the test process, due to that the preferred face, recommendation task in this paper is generally a retrieval task instead of classification. Therefore, the following 3 schemes are considered based on different numbers of testing samples: 1) 10 testing samples out of 30 samples; 2) 20 testing samples out of 30 samples; 3) all the 30 samples are used as testing data. With the 3 evaluation schemes, the evaluation metrics such as top−1, top−5, top−10, and mean average precision (mAP) are computed by score ranking of the whole dataset excluding the training data.

4.2.2 Evaluation metrics

In evaluation, the preference scores of the testing samples in the whole dataset are computed by using the trained model (i.e., 3 018 female samples and 3 017 male samples after excluding the 10 training samples). Then, after score ranking, the top 10 faces with the highest scores from all testing samples are recorded. The hit rate of all users that is computed as the rank−k accuracy is used for evaluating the proposed method. Specifically, the accuracy of rank−1, rank−5, and rank−10 is used as the evaluation metric, respectively. In detail, rank−1 means that the sample with the top 1 score (i.e., the highest score) should fall into the testing set. Similarly, for

rank−5 and rank−10, it means that there is at least one sample in the top 5 or top 10 falling into the testing set. It is clear that the larger the size of the test set is, the higher the accuracy is. Additionally, considering the retrieval task, mAP is also used as a metric for evaluating the proposed method. Specifically, all the samples excluding the 10 training anchors are used as gallery set for evaluation. Therefore, the gallery sizes for the female and male data are 3018 and 3017, respectively.

4.3 Compared methods

For evaluating the effectiveness and superiority of the proposed DiscoStyle method on the developed StyleFace database, several popular methods including one feature descriptor and 3 regressors have been conducted. For the texture descriptor, the popular histogram of oriented Gradient (HOG)[59] is exploited for facial feature extraction and representation, which is a competitor of our PFR method. For regressor, OneclassSVM[60], traditional support vector machine (SVM)[61] and extreme learning machine (ELM)[62] are implemented as competitors of our MLR model. Considering the different combinations between 2 feature descriptors and 4 modeling methods, in total 8 methods are formulated and implemented for experimental evaluation and comparison. Specifically, the 8 methods include: HOG+OneclassSVM, HOG+SVM, HOG+ELM, HOG+MLR, PFR+OneclassSVM, PFR+SVM, PFR+ELM, and DiscoStyle (PFR+MLR). Obviously, the proposed DiscoStyle consists of the proposed preference feature representation (PFR) and the proposed multi-level logistic ranking (MLR). Note that the OneclassSVM[60] belongs to a one-class algorithm which only depends on the positive samples during the training process. Also, for each method, the preference prediction score for each face is computed for final ranking.

4.4 Experimental results

In this section, by following the experimental protocol described above, the testing results of rank−1, rank−5, rank−10 and mAP on the StyleFace database by using different methods are presented. First, the evaluation is conducted when the number of testing samples is set as 10, and the results are reported in Table 2 for both female and male samples. From the results, we have the following observations:

1) The OneclassSVM[60] shows the worst results by comparison to other models. The reason is that in this model, only the positive samples (preferred faces) are used for training. However, in other models, the negative samples (non-preferred faces) are automatically selected by using the proposed ONSS algorithm. Therefore, the advantage of our ONSS method is clearly shown.

2) By comparing the feature representation methods between PFR and HOG, we observe that for OneclassSVM, SVM and ELM models, the proposed PFR can achieve significant superiority in retrieval performance, which clearly shows the strength of the proposed PFR in revealing the knowledge of facial preference.

3) Under the image representation of HOG descriptor, we can see that the proposed MLR model outperforms another three models in retrieval of preferred faces. This demonstrates that the proposed MLR can successfully understand the implicit preference feature.

4) Finally, we can see that our proposed DiscoStyle method (PFR+MLR) achieves a huge breakthrough in the preferred faces retrieval task, and shows the state-of-the-art performance among all the presented methods.

Additionally, we have also presented the evaluation results when the number of testing samples is 20, and the results are shown in Table 3. From the results, we can also observe that the proposed DiscoStyle with ONSS, PFR, and MLR outperforms other methods. The proposed DiscoStyle method still ranks the 1st position among all the methods. It is reasonable that with an increasing number of testing samples, the retrieval performance is improved. Similar findings can be observed in Table 4, which is achieved with 30 testing samples for each user. We can see that for our DiscoStyle, the rank−1 accuracy is 67.6% and the rank−5 accuracy can achieve 97.3% for female data while achieving 68.75% in rank−1

Table 2 Performance comparisons of all methods under the setting of 10 testing samples

	Female				Male			
Methods	Rank−1 (%)	Rank−5 (%)	Rank−10 (%)	mAP (%)	Rank−1 (%)	Rank−5 (%)	Rank−10 (%)	mAP (%)
HOG+OneclassSVM	1.35	5.40	12.16	6.89	3.12	6.25	12.50	1.16
PFR+OneclassSVM	1.35	16.21	35.13	3.76	15.62	81.25	90.62	31.87
HOG+SVM	2.70	36.48	60.81	9.52	15.62	62.50	75.00	15.18
PFR+SVM	21.62	54.05	77.02	17.02	21.87	71.87	90.62	30.89
HOG+ELM	10.81	41.89	64.86	11.00	25.00	68.75	71.87	19.12
PFR+ELM	21.62	59.45	75.67	16.41	28.12	96.87	96.87	34.05
HOG+MLR	39.18	79.72	83.79	36.52	**37.50**	75.00	90.62	**40.32**
DiscoStyle (PFR+MLR)	**41.89**	**82.43**	**90.54**	**48.45**	34.37	**90.62**	**100**	39.72

Table 3 Performance comparisons of all methods under the setting of 20 testing samples

	Female				Male			
Methods	Rank−1 (%)	Rank−5 (%)	Rank−10 (%)	mAP (%)	Rank−1 (%)	Rank−5 (%)	Rank−10 (%)	mAP (%)
HOG+OneclassSVM	4.05	13.51	25.67	7.59	3.12	21.87	25.00	2.48
PFR+OneclassSVM	6.75	39.18	62.16	5.57	31.25	81.25	93.75	19.12
HOG+SVM	13.51	70.27	83.78	13.96	21.87	78.12	84.37	16.69
PFR+SVM	35.13	75.67	91.18	20.31	40.62	87.50	93.75	33.70
HOG+ELM	25.67	72.97	85.13	15.20	40.62	78.12	90.62	20.15
PFR+ELM	39.18	77.02	91.89	19.75	28.12	87.50	96.87	34.05
HOG+MLR	55.40	91.18	**95.94**	38.95	56.25	**93.75**	**100**	38.30
DiscoStyle (PFR+MLR)	**60.81**	**95.59**	**95.94**	**44.29**	**59.37**	**93.75**	**100**	**38.56**

Table 4 Performance comparisons of all methods under the setting of 30 testing samples

	Female				Male			
Methods	Rank−1 (%)	Rank−5 (%)	Rank−10 (%)	mAP (%)	Rank−1 (%)	Rank−5 (%)	Rank−10 (%)	mAP (%)
HOG+OneclassSVM	4.05	18.91	31.08	8.13	12.50	59.37	62.50	5.28
PFR+OneclassSVM	12.16	55.40	77.02	7.40	37.50	93.75	96.87	21.78
HOG+SVM	47.29	93.24	97.29	20.21	28.12	90.62	90.62	17.65
PFR+SVM	56.75	95.94	**100**	23.12	59.37	96.87	96.87	33.39
HOG+ELM	51.35	93.24	**100**	20.18	59.37	96.87	**100**	20.18
PFR+ELM	55.40	95.94	98.64	23.12	50.00	96.87	96.87	33.80
HOG+MLR	66.21	94.59	**100**	36.94	65.62	93.75	**100**	**35.19**
DiscoStyle (PFR+MLR)	**67.56**	**97.29**	98.64	**39.02**	**68.75**	**100**	**100**	35.01

and 100% for rank−5 for male data. This demonstrates that our proposed method can successfully realize accurate and reliable recommendation of preferred faces for users.

Similar to the retrieval tasks, the precision-recall (PR) curves of all the compared methods for both female and male samples are also presented in Fig. 6, from which we can observe the superior performance of our DiscoStyle (PFR+MLR) to others. The mAPs for female and male data under different settings are shown in Tables 2–4. We can see that the proposed DiscoStyle model shows state-of-the-art performance. The superiority of the proposed PFR and MLR is clearly shown.

4.5 Visualization of the StyleFace dataset

Visualization of the feature/image distribution. For better insight of the proposed PFR feature representation method, we have shown the feature distribution (displayed with images) in Fig. 7 for user a (female samples) and user b (male samples), which is implemented by running the t-SNE based dimension reduction algorithm[63] on our PFR feature. The appearance feature of the detected ROI region from the deep transfer network is experimented in distribution visualization. From Fig. 7, we can observe that for both users, the top 25 faces in green color with the highest preference scores are in the same cluster as the training set (preferred faces) in blue color. Additionally, the bottom 25 faces in red color with the lowest preference scores are in the same cluster as the negative samples in yellow color selected by our ONSS algorithm. Notably, the numbers below the bounding boxes in Fig. 7 represent the score ranking of each face. The visualization clearly shows the effectiveness of the proposed feature representation in reflecting the essence of users′ preference to facial images.

4.6 Visualization of the LFW dataset

In order to further test the effectiveness of the proposed DiscoStyle model in public western faces, we present a study on the LFW[58] dataset. We divide the LFW dataset into female and male parts. Then 10 preferred anchor images from female and male parts are separately selected by a user for model training. The top 10 recommended faces and bottom 10 non-recommended faces for both female and male parts are visualized in Figs. 8(a) and 8(b), respectively. We can observe that the recommended faces show similar style with the preferred anchors and the non-recommended faces show non-similar style. Also, the proposed method is gender and race-insensitive, and shows good generalization and reliability.

4.7 Discussions

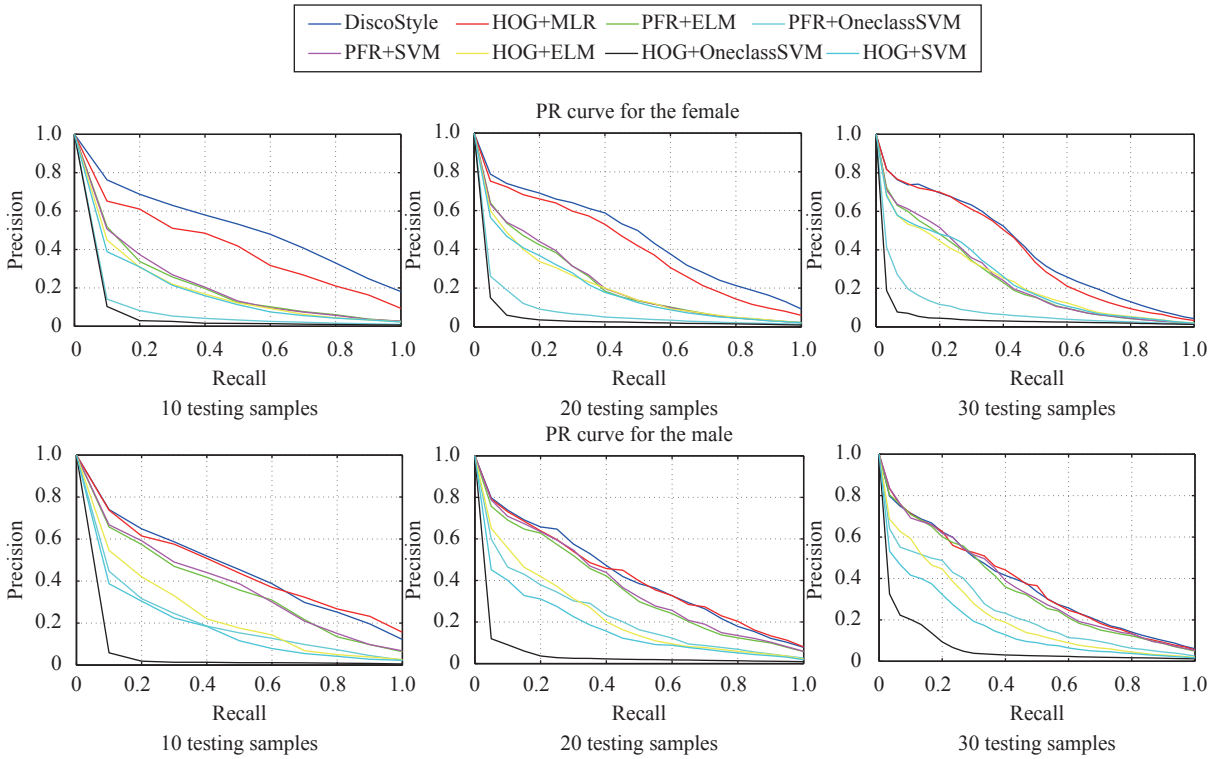


Fig. 6 The PR curves of all the compared methods based on three evaluation protocols. The 1st and the 2nd rows show the PR curves of female and male, respectively. The online color figure can get a better view.

Fig. 7 The feature distribution of training set, high-score samples and low-score samples for female (a) and male (b). The numbers below the red and green bounding boxes denote the rankings of the preference scores predicted by the proposed DiscoStyle. The online color figure can get a better view.

A number of experiments on the developed StyleFace and LFW[58] databases have shown the effectiveness of the proposed DiscoStyle in predicting users′ preference for faces. The success of our DiscoStyle for this challenging task, to our knowledge, lies in three important aspects. 1) An MLR model is proposed for predicting the possibility of a face to be preferred by users, such that the style similarity can be finely analyzed by scoring strategy. 2) A PFR is proposed with local appearance features (e.g., eye, nose, and lip-chin parts that users generally pay more attention to), global appearance feature and geometric shape feature, such that the facial features that

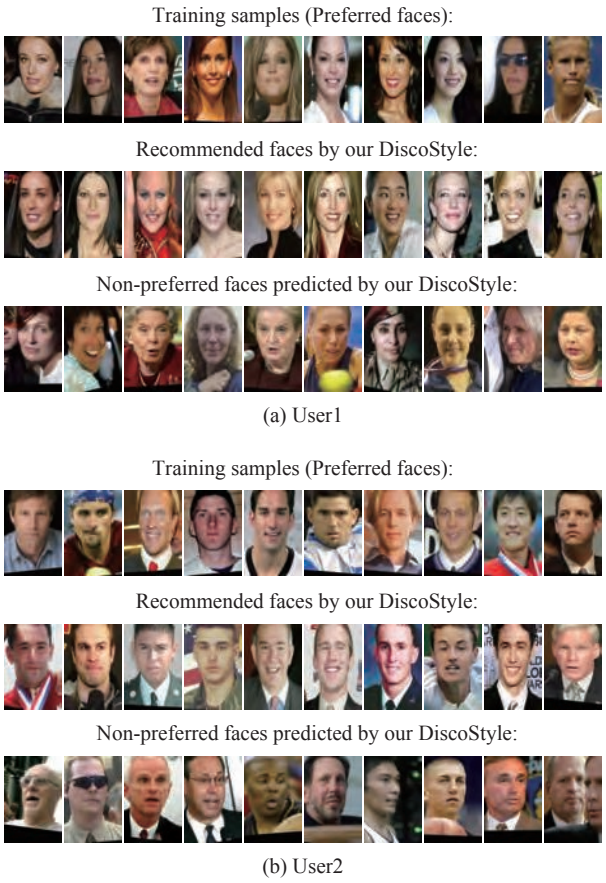


Fig. 8 Visualization of the anchors, recommended faces and non-recommended faces for female and male data in LFW dataset. For each user, the 1st row shows the training samples (i.e., preferred anchors) selected by the user, the 2nd row shows the recommended faces (most probably preferred faces) with the highest preference scores, and the 3rd row shows the faces (most probably non-preferred faces) with the lowest preference scores by using the proposed DiscoStyle method.

mostly reflect the users′ preference can be extracted. 3) An ONSS algorithm is proposed, such that a generalized model can be learned by trading off between positive samples and negative samples, which benefits the algorithm in score ranking.

Although the proposed DiscoStyle, to the best of our knowledge, is the first work in this challenging task and has shown a significant success toward the users′ preference modeling of faces in different genders and races, the essence of users′ preference to faces is still complex and not explicit. The preference is closely related to several personal aspects such as hobbies, character, psychological, etc. However, in this paper, we aim at proposing intelligent and automatic approach by concentrating on the external facial image and feature without considering other internal and complicated factors.

5 Conclusion and future work

In this paper, we introduce a novel and efficient DiscoStyle approach and a StyleFace database for modeling and inferring human′s psychological preference to faces. To our knowledge, this is the first work for preference inference of image styles with a StyleFace database, towards preferred faces recommendation. The proposed DiscoStyle framework is formulated by three important modules including the preference oriented feature representation method, the on-line negative sample selection algorithm, and the multi-level logistic ranking model. Extensive experiments on the developed StyleFace database show the superior efficiency of the proposed DiscoStyle over other general methods in accurately inferring the users′ psychological preference to facial image styles.

In future work, we would like to deploy the proposed DiscoStyle on more interesting applications for recommendation, such as human′s psychological preference inference to many diverse image styles (e.g., landscapes, animals, paintings) rather than only human faces. This research will open up the emotional world of humans by using artificial intelligence.

Acknowledgments

This work was supported by National Natural Science Fund of China (No. 61771079), Chongqing Natural Science Fund (No. cstc2018jcyjAX0250) and Chongqing Youth Talent Program. The authors would like to thank the volunteers for their contribution in labeling the StyleFace for preferences modeling.

References

[1] F. Schroff, D. Kalenichenko, J. Philbin. FaceNet: A unified embedding for face recognition and clustering. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Boston, USA, pp. 815–823, 2015. DOI: 10.1109/CVPR.2015.7298682.

[2] Y. Taigman, M. Yang, M. A. Ranzato, L. Wolf. Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Columbus, USA, pp. 1701–1708, 2014. DOI: 10.1109/CVPR.2014.220.

[3] Y. Sun, X. G. Wang, X. O. Tang. Deep learning face representation from predicting 10, 000 classes. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Columbus, USA, pp. 1891–1898, 2014. DOI: 10.1109/CVPR.2014.244.

[4] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009. DOI: 10.1109/TPAMI.2008.79.

[5] J. K. Chen, Z. H. Chen, Z. R. Chi, H. Fu. Facial expression recognition in video with multiple feature fusion. *IEEE Transactions on Affective Computing*, vol. 9, no. 1, pp. 38–50, 2018. DOI: 10.1109/TAFFC.2016.2593719.

[6] L. Zhang, D. Zhang, M. M. Sun, F. M. Chen. Facial beauty analysis based on geometric feature: Toward attractiveness assessment application. *Expert Systems with Applications*, vol. 82, pp. 252–265, 2017. DOI: 10.1016/j.eswa.

2017.04.021.

[7] Y. Fu, G. D. Guo, T. S. Huang. Age synthesis and estimation via faces: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 1955–1976, 2010. DOI: 10.1109/TPAMI.2010.36.

[8] E. Eidinger, R. Enbar, T. Hassner. Age and gender estimation of unfiltered faces. *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2170–2179, 2014. DOI: 10.1109/TIFS.2014.2359646.

[9] Z. Lian, Y. Li, J. H. Tao, J. Huang, M. Y. Niu. Expression analysis based on face regions in real-world conditions. *International Journal of Automation and Computing*, vol. 17, no. 1, pp. 96–107, 2020. DOI: 10.1007/s11633-019-1176-9.

[10] H. S. Du, Q. P. Hu, D. F. Qiao, I. Pitas. Robust face recognition via low-rank sparse representation-based classification. *International Journal of Automation and Computing*, vol. 12, no. 6, pp. 579–587, 2015. DOI: 10.1007/s11633-015-0901-2.

[11] H. Wu, Z. W. Chen, G. H. Tian, Q. Ma, M. L. Jiao. Item ownership relationship semantic learning strategy for personalized service robot. *International Journal of Automation and Computing*, vol. 17, no. 3, pp. 390–402, 2020. DOI: 10.1007/s11633-019-1206-7.

[12] D. Zhang, Q. J. Zhao, F. M. Chen. Quantitative analysis of human facial beauty using geometric features. *Pattern Recognition*, vol. 44, no. 4, pp. 940–950, 2011. DOI: 10.1016/j.patcog.2010.10.013.

[13] F. M. Chen, X. H. Xiao, D. Zhang. Data-driven facial beauty analysis: Prediction, retrieval and manipulation. *IEEE Transactions on Affective Computing*, vol. 9, no. 2, pp. 205–216, 2018. DOI: 10.1109/TAFFC.2016.2599534.

[14] A. Krizhevsky, I. Sutskever, G. E. Hinton. ImageNet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems*, Lake Tahoe, USA, PP.1097–1105, 2012.

[15] K. Simonyan, A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *Proceedings of the 3rd International Conference on Learning Representations*, San Diego, USA, 2015. https://arxiv.org/abs/1409.1556.

[16] W. Y. Liu, Y. D. Wen, Z. D. Yu, M. Yang. Large-margin softmax loss for convolutional neural networks. In *Proceedings of the 33rd International Conference on Machine Learning*, New York, USA, 2016.

[17] S. J. Pan, Q. Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010. DOI: 10.1109/TKDE.2009.191.

[18] K. Saenko, B. Kulis, M. Fritz, T. Darrell. Adapting visual category models to new domains. In *Proceedings of the 11th European Conference on Computer Vision*, Springer, Heraklion, Greece, 2010. DOI: 10.1007/978-3-642-15561-1_16.

[19] M. S. Long, H. Zhu, J. M. Wang, M. I. Jordan. Unsupervised domain adaptation with residual transfer networks. In *Proceedings of the 30th Conference on Neural Information Processing Systems*, Barcelona, Spain, pp. 136–144, 2016.

[20] L. Zhang, W. M. Zuo, D. Zhang. LSDT: Latent sparse domain transfer learning for visual adaptation. *IEEE Transactions on Image Processing*, vol. 25, no. 3, pp. 1177–1191, 2016. DOI: 10.1109/TIP.2016.2516952.

[21] L. Zhang, S. S. Wang, G. B. Huang, W. M. Zuo, J. Yang, D. Zhang. Manifold criterion guided transfer learning via intermediate domain generation. *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 12, pp. 3759–3773, 2019. DOI: 10.1109/TNNLS.2019.2899037.

[22] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, J. M. Liang. Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, 2016. DOI: 10.1109/TMI.2016.2535302.

[23] H. C. Shin, H. R. Roth, M. C. Gao, L. Lu, Z. Y. Xu, I. Nogues, J. H. Yao, D. Mollura, R. M. Summers. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016. DOI: 10.1109/TMI.2016.2528162.

[24] D. Marmanis, M. Datcu, T. Esch, U. Stilla. Deep learning earth observation classification using imagenet pretrained networks. *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 1, pp. 105–109, 2016. DOI: 10.1109/LGRS.2015.2499239.

[25] X. W. Yao, J. W. Han, G. Cheng, X. M. Qian, L. Guo. Semantic annotation of high-resolution satellite images via weakly supervised learning. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 6, pp. 3660–3671, 2016. DOI: 10.1109/TGRS.2016.2523563.

[26] M. Xie, N. Jean, M. Burke, D. Lobell, S. Ermon. Transfer learning from deep features for remote sensing and poverty mapping. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence*, AAAI, Phoenix, USA, 2015.

[27] N. Jean, M. Burke, M. Xie, W. M. Davis, D. B. Lobell, S. Ermon. Combining satellite imagery and machine learning to predict poverty. *Science*, vol. 353, no. 6301, pp. 790–794, 2016. DOI: 10.1126/science.aaf7894.

[28] Q. Y. Duan, L. Zhang, W. M. Zuo. From face recognition to kinship verification: An adaptation approach. In *Proceedings of IEEE International Conference on Computer Vision Workshops*, IEEE, Venice, Italy, pp. 1590–1598, 2017. DOI: 10.1109/ICCVW.2017.187.

[29] L. Zhang, Q. Y. Duan, D. Zhang, W. Jia, X. Z. Wang. Advkin: Adversarial convolutional network for kinship verification. *IEEE Transactions on Cybernetics*, *published online*, 2020. DOI: 10.1109/TCYB.2019.2959403.

[30] C. Q. Hong, J. Yu, J. Zhang, X. N. Jin, K. H. Lee. Multimodal face-pose estimation with multitask manifold deep learning. *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 3952–3961, 2019. DOI: 10.1109/TII.2018.2884211.

[31] Q. C. Zhu, Z. H. Chen, Y. C. Soh. A novel semisupervised deep learning method for human activity recognition. *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 3821–3830, 2019. DOI: 10.1109/TII.2018.2889315.

[32] Y. D. Yang, W. Li, T. A. Gulliver, S. F. Li. Bayesian deep learning-based probabilistic load forecasting in smart grids. *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4703–4713, 2020. DOI: 10.1109/TII.2019.2942353.

[33] L. Zhang, D. Zhang. Efficient solutions for discreteness, drift, and disturbance (3D) in electronic olfaction. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 2, pp. 242–254, 2018. DOI: 10.1109/TSMC.2016.

2597800.

[34] L. Zhang, P. L. Deng. Abnormal odor detection in electronic nose via self-expression inspired extreme learning machine. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 10, pp. 1922–1932, 2019. DOI: 10.1109/TSMC.2017.2691909.

[35] T. Serre, G. Kreiman, M. Kouh, C. Cadieu, U. Knoblich, T. Poggio. A quantitative theory of immediate visual recognition. *Progress in Brain Research*, vol. 165, pp. 33–56, 2007. DOI: 10.1016/S0079-6123(06)65004-8.

[36] D. Cheng, Y. H. Gong, S. P. Zhou, J. J. Wang, N. N. Zheng. Person re-identification by multi-channel parts-based CNN with improved triplet loss function. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Las Vegas, USA, pp. 1335–1344, 2016. DOI: 10.1109/CVPR.2016.149.

[37] E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, B. Schiele. DeeperCut: A deeper, stronger, and faster multi-person pose estimation model. In *Proceedings of the 14th European Conference on Computer Vision*, Springer, Amsterdam, The Netherlands, 2016. DOI: 10.1007/978-3-319-46466-4_3.

[38] Y. Li, H. Z. Qi, J. F. Dai, X. Y. Ji, Y. C. Wei. Fully convolutional instance-aware semantic segmentation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Honolulu, USA, pp. 4438–4446, 2017. DOI: 10.1109/CVPR.2017.472.

[39] C. Dong, C. C. Loy, K. M. He, X. O. Tang. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2016. DOI: 10.1109/TPAMI.2015.2439281.

[40] K. M. He, X. Y. Zhang, S. Q. Ren, J. Sun. Deep residual learning for image recognition. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Las Vegas, USA, pp. 770–778, 2016. DOI: 10.1109/CVPR.2016.90.

[41] G. Huang, Z. Liu, L. van der Maaten, K. Q. Weinberger. Densely connected convolutional networks. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Honolulu, USA, pp.2261–2269, 2017. DOI: 10.1109/CVPR.2017.243.

[42] S. Q. Ren, K. M. He, R. Girshick, J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *Proceedings of Advances in Neural Information Processing Systems 28*, Montreal, Canada, 2015.

[43] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, A. C. Berg. SSD: Single shot multibox detector. In *Proceedings of the 14th European Conference on Computer Vision*, Springer, Amsterdam, The Netherlands, pp. 21–37, 2016. DOI: 10.1007/978-3-319-46448-0_2.

[44] J. Redmon, A. Farhadi. Yolo9000: Better, faster, stronger. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Honolulu, USA, pp. 6517–6525, 2017. DOI: 10.1109/CVPR.2017.690.

[45] Z. Cao, T. Simon, S. E. Wei, Y. Sheikh. Realtime multi-person 2D pose estimation using part affinity fields. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Honolulu, USA, pp.1302–1310, 2017. DOI: 10.1109/CVPR.2017.143.

[46] X. L. Wang, T. T. Xiao, Y. N. Jiang, S. Shao, J. Sun, C. H. Shen. Repulsion loss: Detecting pedestrians in a crowd. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Salt Lake City, USA, pp.7774–7783, 2018. DOI: 10.1109/CVPR.2018.00811.

[47] Z. X. Feng, J. H. Lai, X. H. Xie. Learning view-specific deep networks for person re-identification. *IEEE Transactions on Image Processing*, vol. 27, no. 7, pp. 3472–3483, 2018. DOI: 10.1109/TIP.2018.2818438.

[48] L. Q. Liu, C. Xiong, H. W. Zhang, Z. H. Niu, M. Wang, S. C. Yan. Deep aging face verification with large gaps. *IEEE Transactions on Multimedia*, vol. 18, no. 1, pp. 64–75, 2016. DOI: 10.1109/TMM.2015.2500730.

[49] Z. F. Li, D. H. Gong, X. L. Li, D. C. Tao. Aging face recognition: A hierarchical learning model based on local patterns selection. *IEEE Transactions on Image Processing*, vol. 25, no. 5, pp. 2146–2154, 2016. DOI: 10.1109/TIP.2016.2535284.

[50] U. Park, Y. Y. Tong, A. K. Jain. Age-invariant face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 947–954, 2010. DOI: 10.1109/TPAMI.2010.14.

[51] H. Dibeklioglu, A. A. Salah, T. Gevers. Like father, like son: Facial expression dynamics for kinship verification. In *Proceedings of IEEE International Conference on Computer Vision*, IEEE, Sydney, Australia, pp. 1497–1504, 2013. DOI: 10.1109/ICCV.2013.189.

[52] R. G. Fang, K. D. Tang, N. Snavely, T. Chen. Towards computational models of kinship verification. In *Proceedings of IEEE International Conference on Image Processing*, IEEE, Hong Kong, China, pp. 1577–1580, 2010. DOI: 10.1109/ICIP.2010.5652590.

[53] H. B. Yan, J. W. Lu, X. Z. Zhou. Prototype-based discriminative feature learning for kinship verification. *IEEE Transactions on Cybernetics*, vol. 45, no. 11, pp. 2535–2545, 2015. DOI: 10.1109/TCYB.2014.2376934.

[54] D. I. Perrett, K. A. May, S. Yoshikawa. Facial shape and judgements of female attractiveness. *Nature*, vol. 368, no. 6468, pp. 239–242, 1994. DOI: 10.1038/368239a0.

[55] K. P. Zhang, Z. P. Zhang, Z. F. Li, Y. Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2016. DOI: 10.1109/LSP.2016.2603342.

[56] D. Yi, Z. Lei, S. C. Liao, S. Z. Li. Learning face representation from scratch. https://arxiv.org/abs/1411.7923, 2014.

[57] V. Kazemi, J. Sullivan. One millisecond face alignment with an ensemble of regression trees. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Columbus, USA, pp. 1867–1874, 2014. DOI: 10.1109/CVPR.2014.241.

[58] G. B. Huang, M. Ramesh, T. Berg, E. Learned-Miller. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments, Technical Report, 07–49, Department of Computer Science, University of Massachusetts, USA, 2007.

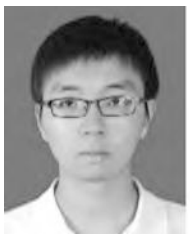
[59] N. Dalal, B. Triggs. Histograms of oriented gradients for human detection. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, San Diego, USA, pp. 886–893, 2005. DOI: 10.1109/CVPR.2005.177.

[60] B. Schölkopf, R. Williamson, A. Smola, J. Shawe-Taylor, J. Platt. Support vector method for novelty detection. In *Proceedings of the 12th International Conference on Neural Information Processing Systems*, Denver, USA, pp. 582–588, 1999.

[61] C. C. Chang, C. J. Lin. Libsvm: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, Article number 27, 2011. DOI: 10.1145/1961189.1961199.

[62] G. B. Huang, H. M. Zhou, X. J. Ding, R. Zhang. Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, vol. 42, no. 2, pp. 513–529, 2012. DOI: 10.1109/TSMCB.2011.2168604.

[63] L. van der Maaten, G. Hinton. Visualizing data using T-SNE. *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.



Zhen-Wei He received B. Eng. degree in information engineering from Tianjin University, China in 2014. From July 2014 to June 2016, he worked in Chongqing Cable Network Inc., China. Now, he is a Ph. D degree candidate in Chongqing University, China.

His research interests include deep learning and computer vision.

E-mail: hzw@cqu.edu.cn
ORCID iD: 0000-0002-6122-9277

Lei Zhang received the Ph. D degree in circuits and systems from the College of Communication Engineering, Chongqing University, China in 2013. He worked as a Post-Doctoral Fellow with Hong Kong Polytechnic University, China from 2013 to 2015. He is currently a professor/distinguished research fellow with Chongqing University, China. He has authored more than 90 scientific papers in top journals and top conferences. He serves as associate editors for *IEEE Transactions on Instrumentation and Measurement*, *Neural Networks*, etc. He is a senior member of IEEE.

His research interests include machine learning, pattern recognition, computer vision and intelligent systems.

E-mail: leizhang@cqu.edu.cn (Corresponding author)
ORCID iD: 0000-0002-5305-8543

Fang-Yi Liu received the B. Eng. degree in communication engineering from Guangxi University, China in 2017. Since September 2017, he is a master student in information and communication engineering in Chongqing University, China.

His research interests include person re-identification and deep learning.

E-mail: fangyiliu@cqu.edu.cn
ORCID iD: 0000-0001-8815-0254