1-1-2022

CommuNety: deep learning-based face recognition system for the prediction of cohesive communities

Syed Afaq Ali Shah
Edith Cowan University, afaq.shah@ecu.edu.au

Weifeng Deng

Muhammad Aamir Cheema

Abdul Bais

10.1007/s11042-022-13741-y
Shah, S. A. A., Deng, W., Cheema, M. A., & Bais, A. (2022). CommuNety: deep learning-based face recognition system for the prediction of cohesive communities. Multimedia Tools and Applications. Advance online publication. https://doi.org/10.1007/s11042-022-13741-y

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-022-13741-y

CommuNety: deep learning-based face recognition system for the prediction of cohesive communities

Syed Afaq Ali Shah[1] · Weifeng Deng[2] · Muhammad Aamir Cheema[3] · Abdul Bais[4]

Received: 3 February 2021 / Revised: 8 June 2022 / Accepted: 29 August 2022

Abstract
Effective mining of social media, which consists of a large number of users is a challenging task. Traditional approaches rely on the analysis of text data related to users to accomplish this task. However, text data lacks significant information about the social users and their associated groups. In this paper, we propose CommuNety, a deep learning system for the prediction of cohesive networks using face images from photo albums. The proposed deep learning model consists of hierarchical CNN architecture to learn descriptive features related to each cohesive network. The paper also proposes a novel Face Co-occurrence Frequency algorithm to quantify existence of people in images, and a novel photo ranking method to analyze the strength of relationship between different individuals in a predicted social network. We extensively evaluate the proposed technique on PIPA dataset and compare with state-of-the-art methods. Our experimental results demonstrate the superior performance of the proposed technique for the prediction of relationship between different individuals and the cohesiveness of communities.

Keywords Deep learning · Social communities · Predictive modelling

1 Introduction

With the pervasiveness of low cost digital cameras and advent in computer vision and machine learning approaches, the collection and analysis of large image data has become a trivial task. As the value of photos is greatly determined by who appears in those photos (e.g., celebrity), labeling photos with their identities becomes an essential task [20, 21, 31, 33].

✉ Syed Afaq Ali Shah
afaq.shah@ecu.edu.au

[1] Centre for AI and Machine Learning, School of Science, Edith Cowan University, Joondalup, Australia

[2] The University of Western Australia, Perth, Australia

[3] Monash University, Melbourne, Australia

[4] University of Regina, Regina, Canada

The popularity of social applications and social networking services (SNS) such as Facebook, Twitter, LinkedIn, Weibo, MOMO and Flickr has led to the formation of online social networks of users on these sites. At present, analyzing online comments (e.g., tweets) is a popular approach to determine effective communities in social networks. While text data contains rich information, the existing methods are unable to utilise the text data to get sufficient information about the social users. In addition, the social networks need to be more comprehensive and accurate [36]. With the advent of imaging technology and the availability of portable high resolution cameras e.g., on smartphones, users can now upload their images and profiles to social media websites and share photos with other users who are part of their social community [18, 34]. Social media users upload countless photos of social activities each day, and the relationship among those who appear in these photos cannot be mined accurately only from text data. Hence, defining online social networks with user-uploaded images, and extraction of human features, such as faces or body from photos becomes an important procedure in building social networks [6, 17, 26, 38]. Note that the popular SNS applications have very large user bases. In 2018 alone, Facebook had 2.2 billion monthly active users. Flickr had over 90 million monthly users, and the number of monthly users of Weibo exceeded 0.44 billion. Therefore, the mining of a potential relationship between social network users is a challenging problem.

To overcome these challenges, this paper proposes a deep learning system, called CommuNety, which uses image data and face recognition for the prediction of comprehensive and cohesive communities. The proposed system complements existing approaches and helps in discovering communities where there are no explicit relationships (e.g., discover communities in an image database) or discovers communities when not all relationships are directly represented in the network e.g., two people may not be friends on social media and may have never interacted on the platform, however, if they appear together in some photos, they have a relationship which can only be discovered using face images.

Several deep learning algorithms have been developed in recent years and have achieved significant breakthroughs in image recognition tasks [11]. In 2014, Simonyan and Zisserman proposed a Deep Convolutional Neural Network (CNN) architecture and achieved an outstanding classification performance [28]. Parkhi et al. used the VGG (Visual Geometry Group) network structure for face recognition and achieved results comparable to other face recognition techniques [22]. Razavian et al. have demonstrated that the features extracted from CNN are powerful and the models trained using CNN features have superior performance. Such features can be used for visual recognition tasks [25].

Inspired by prior approaches, in this paper, we propose a deep learning-based face recognition model, which learns distinctive image features. The proposed model is then used to predict community network and its hierarchy that is centered at the target person in photo albums. In our proposed technique, every photo in the training set is also ranked using the term frequency inverse document frequency (TF-IDF) numerical statistics. Then, the strength of relationship between each pair of persons in the photos is represented by the sum of the TF-IDF values of their group photos. As a result, the community predicted by the proposed deep learning system contains all the persons who have direct or indirect relationship with the target person and different relationship strength among them.

Recognizing people from high-quality photos, which contain high-resolution facial images, is a trivial task for humans. However, well trained autonomous system still struggle with this challenging task. This is because of the variations in natural images, such as changes in illumination and viewpoint change or head rotation. Moreover, although some progresses have been made recently in frontal face recognition, non-frontal views are more common in social media photo albums. A few face recognition techniques perform face

detection as a preliminary step [9, 27, 37]. Note that face detection can be regarded as a two-class (face versus non-face) classification problem. However, these techniques cannot deal with significant variations in face images such as head rotation and view changes, etc. to detect and recognise faces. Other model-based approaches require that the initial locations of faces are known in advance [8] and then they perform face tracking to recognise individuals in the image data. This paper overcomes these challenges. The significance of this research is to recognise people from any viewpoint and associate them with established cohesive social communities [35].

The contributions of this paper can be summarised as follows:

- **First**, we propose a deep learning model to predict cohesive communities or networks from photo albums using image data and face recognition.
- **Second**, we propose a novel algorithm to calculate the relationship strength among people in the predicted network/community. We also present the final networks using data visualization techniques.
- **Third**, we propose novel features for image-generated networks compared with other social communities.
- **Four**, we perform extensive evaluation of the proposed technique. Our experimental results demonstrate the superior performance of the proposed system on the PIPA ("People In Photo Albums") dataset.

The rest of this paper is organised as follows. The next section discusses the prior works related to this research. Section 3 describes the proposed methodology and provides information about the proposed face recognition model, construction of social networks, and analysis of the predicted social networks. The PIPA dataset used for the evaluation and data pre-processing are discussed in Section 4. Section 5 presents our experimental results. The paper is concluded in Section 6.

2 Literature review

Kim et al., [13] proposed DiscFace to address the limitation of softmax-based models. One of the important issues of softmax-based methods is that the sample features around the corresponding class weight are similarly penalized in the training phase even though their directions are different from each other. This directional discrepancy, i.e., process discrepancy leads to performance degradation at the evaluation phase. To address this issue, they proposed minimum discrepancy learning that enforces directions of intra-class sample features to be aligned toward an optimal direction by using a single learnable basis.

Bah et al., [1] proposed an improved local binary pattern technique combined with histogram equalization, bilateral filter, and image blending for face recognition. They used their proposed method for an attendance system and it was shown to achieve very good face recognition performance. The limitation of their approach is that it has been evaluated on frontal face images.

Deep Unified Model (DUM) has been proposed for face recognition [10]. DUM is based on convolutional neural network and edge computing. They trained their model on publicly vailable labeled faces in the wild (LFW) dataset. The model was then tested for the student's attendance system using face recognition. Their proposed technique is shown to achieve good performance for frontal face images.

In a recent study by Liu et al., [16], the privacy of face recognition and influencing factors are analysed. The study collected 518 questionnaires through the Internet and SPSS

25.0 was used to analyze the data and also to evaluate its reliability. They used Cronbach's alpha coefficient to measure data in this study. The study demonstrates that when users perceive the risk of their private information being disclosed through face recognition, they have greater privacy concerns. However, most users will still choose to provide personal information in exchange for the services and applications they need.

Pfeil et al. [23] proposed a technique to estimate the age differences of users in online social communities. They extracted information from MySpace's user profile pages and divided the users into teenagers and older people communities. Users in the same community have common features, for example, teenagers have larger friends networks than older users.

Chen et al. [3] proposed a technique to identify family and non-family images, which were collected from social media and to predict the pairwise relationship of persons who were in the same family images. To categorize different group types or events, a bag-of-face-subgraphs (BoFG) was proposed. BoFG contained meaningful subgraphs, which represented a group photo, and the occurring frequency of these subgraphs was adequate to identify specific image types. The authors trained an SVM classifier using BoFG features and their technique achieved an accuracy of 89% on family image recognition. In addition, a Naive Bayesian classifier was used to predict the pairwise relationship by getting the image frequency of appearance of the informative subgraph in the image collections. Their proposed technique achieved good improvement over prior works, especially in image categorization area. However, there are still several limitations of their technique. For instance, the images used in the training and testing phases are frontal face images. Hence, if BoFG is applied to open world images, which contain large number of non-frontal facse, the performance would be significantly affected. Moreover, in their proposed method, the pairwise relationship is identified based on the gap of age and gender in a household. This special feature is not feasible for other types of relationship, which do not involve age and gender gap. This limits the application of their proposed technique on real world social network data.

Kim et al. [12] developed an associative network structure called Face Co-occurrence Networks (FCON), which was used to recommend reliable social friends and explore relationships among people based on tagged personal photos. FCON consists of vertices (V) and edges (E), where V is a set of faces, which appeared in photos and E is a set of links between each pair of faces (aka. co-occurrences of faces), both V and E are accumulated. Converting all photos into a global FCON, the weights of V and E in the network were obtained by accumulating V and E in each subnetwork. Subsequently, parts of weights which were related to the target user were calculated to get a set of scores and compare these scores with a pre-set threshold. Finally, using the vertices which have scores higher than the threshold to establish a target user-centric relationship network. Besides, the authors also develop a web-based system named VizFaceCo for data visualization. An aspect that is obviously worth improving is that their technique does not include face recognition. The photos are manually annotated with corresponding names before building FCON. In contrast, in our proposed technique, automatic face recognition is used as a core technology.

Li et al. [15] proposed hybrid method for person recognition in photo albums. In their approach, both the deep convolutional and hand-crafted features extracted from every person's image. These multi-modality features are then fused by a weighted average method and classified by a pre-trained SVM in the recognition procedure. Their experimental results show the effectiveness of the proposed method. However, their technique is not computationally efficient as it requires the computation of hand-crafted and deep convolutional hybrid features for good performance.

Oh et al. [19] proposed an optimized model for person recognition, called naeil2, which can handle large variations in person images. naeil2 consists of seventeen cues (including

five vanilla regional cues, two head cues, ten attribute cues) and DeepID2+ face recognition module. All the cues were obtained from the seventh layer (fc7) of AlexNet [19], and concatenated together. Finally, these cues and DeepID2+ using L2 normalization was combined to build the final naeil2 model. naeil2 has been shown to achieve an outstanding person recognition performance. However, this model relies on multiple features such as several body cues to identify persons. In most of the social media photos, multiple body cues are hard to capture and therefore their proposed technique fails in these situations.

Li et al. [14] has proposed a dual-glance technique to recognize social relationship. In their proposed technique, the first glance fixates at the person of interest and the second glance deploys attention mechanism to exploit contextual cues. To demonstrate the effectiveness of their technique, a large scale People in Social Context dataset is developed that contains 23,311 images and 79,244 person pairs with annotated social relationships.

Dong et al. [4] proposed a method for human age identification. They proposed CNN based DeepID architecture. In their method, the loss function for classification was modified and a distance term was added to the loss function to emphasize on the relationships between labels. They used different parts of face images to train multiple classifiers, and by comparing the accuracy of an exact match (AEM), the eye region was found to be the most significant feature, which can reflect the age of the person. To further improve AEM, different models were combined, and the best model combination was shown to achieve good performance.

They also described in detail the transfer learning strategy adopted in their work that used fewer data samples to train their model to achieve good performance. Concretely, they used large-scale data sets to train a face recognition model, then transferred the parameters of convolutional layers to another network which had same architecture, but the parameters in fully connected layers were randomly initialized. This new network was fine-tuned using the small-scale dataset to get the desired age classification model.

The limitation of their technique is that the performances of the face recognition models were not outstanding compared to the state-of-the-art face recognition techniques e.g., the lowest error rate for DeepID is 0.4. In other words, the accuracy of the best model is 60% [4]. One of the reason is that the architecture of CNN used in DeepID is relatively simple, for instance, the DeepID only has four convolutional layers. Their model was not able to handle the image data complexity, therefore it was under-fitting. The accuracy of recognition can be improved by using a more complex CNN architecture and more training data [5].

In this paper, we overcome the limitations of the prior methods and propose a novel technique for the prediction of communities using images and calculate the relationship strength between the connected persons.

3 Proposed methodology

In this section, we describe our proposed deep learning-based system, CommuNety, to predict social community centered on the input image. Figure 1 shows the block diagram of our

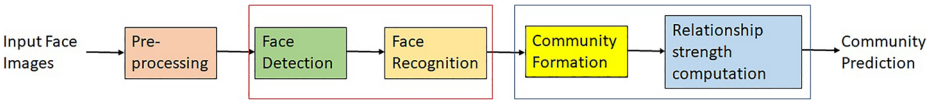

Fig. 1 Block diagram of our proposed methodology for community prediction using face images. Red rectangle: Face recognition phase. Blue Rectangle: Community Formation phase

proposed methodology. The proposed deep learning system consists of two phases including the face recognition phase (red rectangle in Fig. 1) and the community formation phase (blue rectangle in Fig. 1). In face recognition phase, our proposed deep learning model first detects faces in the input images and is then trained to perform accurate face recognition. In the community formation phase, a novel Face Co-occurrence Frequency algorithm is developed to form communities and calculate the relationship strength to predict cohesive communities. We propose a novel algorithm to calculate the relationship strength between people in the predicted social community. In addition, we also propose novel features for the predicted communities by analyzing their properties.

3.1 Face recognition phase

To establish accurate and cohesive networks, the most challenging task is to accurately recognize persons in given photos. We first detect faces in the input images by using the Viola and Jones algorithm [32]. The outcome of face detection (i.e., the bounding box of faces) is validated by the annotations provided in [35]. The detected face images along with their labels are then fed to our deep learning based face recognition system. These face images are used for the training and testing of our proposed deep neural network, which is discussed in the following section.

3.1.1 Deep neural network architecture

We propose a deep face recognition architecture to extract discriminating and distinctive features for face recognition task. The proposed deep learning architecture is composed of sixteen blocks. The first eleven blocks consist of convolutional layers. Each block is followed by one non-linear activation function ReLU, and five max-pooling layers are interspersed between blocks to reduce computational load. The last three blocks are the Fully Connected (FC) layers. The last layer is a softmax layer for multi-class classification and its dimension is equal to the number of class labels in task.

3.1.2 Deep network training and testing

The neural network is trained as a multi-class classifier to recognize persons using their face images. The class probability is computed using the following equation, which computes probability in the range between 0 and 1 for each class:

$$y_j = \frac{e^{x_j}}{\sum_{i=1}^{N} e^{x_i}} \tag{1}$$

where N represents the number of classes and y_j is the probability of class j. x_j is the output of the *jth* neuron in the soft-max layer. Its role is to increase the probability of true class label. In addition, we use cross-entropy loss function as in (2) for the softmax layer:

$$L = -\sum_{j=1}^{N} 1[al = j] \log y_j \tag{2}$$

where *al* is the actual label of input.

During testing, given a test face image, the network then predicts the class label for the input test image. The output of face recognition is then used in the subsequent modules and to predict the social community as discussed below.

3.2 Community prediction and formation phase

Once the faces have been successfully recognised in the input images, the next phase is to predict communities (i.e., persons who have direct or indirect relationship with the target person and different relationship strength among them) using facial images. We propose two algorithms to predict social communities using our face recognition system, and compute relationship strength for each pair of connected nodes in the communities.

3.2.1 Recursive face co-occurrence frequency

In this section, we describe our proposed recursive face co-occurrence frequency technique for the prediction of communities. The proposed technique is similar to FCON [12], however, it is recursive in nature. A dictionary is first defined to store face co-occurrence frequencies as follows:

$$F_i = \{P_1 : f_1, P_2 : f_2, ..., P_k : f_k\} \tag{3}$$

where i is the target candidate, key P_k is the class of the kth person in the dataset, and key value f_k represents the number of times the kth person appears in the given album. The key values are initially zero.

Given an input face image of target person, our proposed face recognition system recognizes and collects all the photos of the target class. Next, comparison of the photo labels of other classes (persons) with the collected target photo labels is performed. The images of other classes are inputted to face recognition system to predict the class label of these input images. When the input person's class is predicted, the key value corresponding to the person's name in the co-occurrence frequency dictionary is incremented by one. After assigning all the matched photos to the dictionary, persons whose key values are larger than zero are considered as directly connected with the target in the social network. Below, we provide a definition of elements contained in the predicted social network.

Definition Root, nodes, and layers:

1. Initial target is the root of its social network, meanwhile, it is on layer 0.
2. Other people in the social network are nodes. Nodes that are directly connected with the root are on layer 1; similarly, the nodes on the second layer are connected to the nodes on the first layer.
3. Because multiple persons may occur in a group photo, therefore the nodes on the same layer may be connected.

Each person on layer 1 is treated as a new target and the same method (as stated above) is followed to build their corresponding single-layer community network. The new community network is then integrated with the previous network to build a 2-layer social network. We only add people who do not already exist in the previous network to the second layer. This process is repeated until no more new person can be found to join the network, and finally a complete community centered on the initial target person is set up. Figure 2 shows an example of community formation process. Person A is the root, which is directly connected to persons B, C and D. The final network has two layers, as only person E is on the latest layer and E does not connect to any new person who is not in the current network.

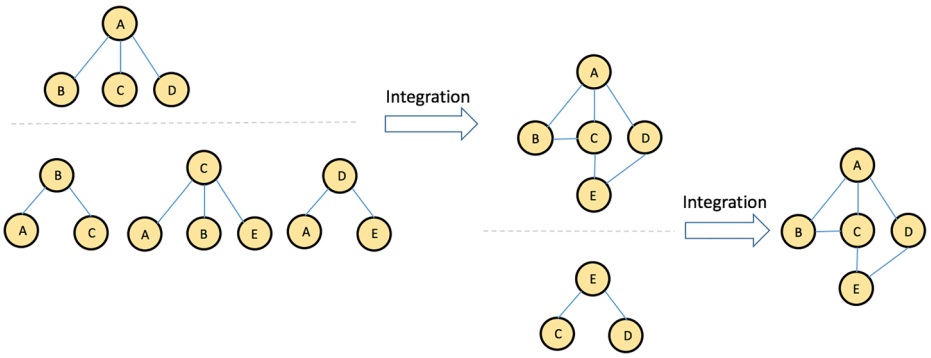

Fig. 2 Community formation process. Left: Person A is the root, which is directly connected to persons B, C and D. Right: The final network has two layers, as only person E is on the latest layer and E does not connect to any new person who is not in the current network

3.2.2 Prediction of relationship strength in communities

To predict the relationship strength of persons in predicted communities, we propose an image ranking algorithm to assign scores to the images that determine the relationship strength in a community. To achieve this, TF-IDF is used for ranking in the predicted community.

TF-IDF is a statistical analysis technique for weighing that reflects the importance of a word for the documents in a corpus. This importance is obtained by comparing the relative frequency of a word in a particular document with the inverse proportion of the word in the entire corpus [24]. In the proposed technique, the corpus consists of all the group photos, where each photo represents a document in the corpus, and the words are replaced by persons in the group photos.

In the proposed technique, the formula for TF-IDF is defined as follows. Given the group photo set G, a candidate c, and a single group photo $g \in G$, the TF-IDF is represented as:

$$TF - IDF_{c,g} = f_{c,g} * \log(|G|/f_{c,G}) \tag{4}$$

where $f_{c,g}$ is the number of times c appears in g, |G| represents the size of the group photo set, and $f_{c,G}$ equals the number of group photos in which c appears in G.

The TF-IDF formula can be separated into two terms TF and IDF as follows:

$$TF_{c,g} = f_{c,g} \tag{5}$$

$$IDF_{c,G} = \log(|G|/f_{c,G}) \tag{6}$$

Since a given person in each photo can only appear once, therefore $TF_{c,g} = 1$. Meanwhile, each candidate has their own fixed IDF value, as the number of times they appear in the entire photo collection is fixed. The more a person appears in the photo collection, the smaller the IDF they receive, and is considered as the lesser important in a specific photo.

The group photos are ranked by calculating the averages of the TF-IDFs of all candidates in each photo:

$$Score_g = \frac{\sum_{i=1}^{k} TF - IDF_{i,g}}{k} \tag{7}$$

where k represents the number of persons in g.

Intuitively, the score of a photo is related to the IDF values of the persons in the photo. For example, if the persons in a photo appear only a few times in the entire collection, the score of this photo is high. On the contrary, if most people in a photo appear in the photo collection many times, then the IDFs of these people are small, and the significance of this photo is low.

Ultimately, the strength of the relationship between each pair of connected people in the network is represented by the sum of the scores of all of their photos in which both persons appear. Each edge in the network is assigned a weight representing the strength of the relationship (larger the better) between the two persons connected by the edge.

4 Image data for evaluation

4.1 Dataset

The proposed technique is evaluated on People In Photo Albums (PIPA) dataset. PIPA dataset contains 37107 Flickr personal photo album images, with 63188 head annotations of 2356 identities, all the images have Creative Commons Attribution License [35]. We used the same experimental protocol as in [19].

The train, val and test sets in the original dataset contain distinct identities i.e., the class labels in training and test set were totally different. Therefore these image data cannot be used directly for the proposed technique. Besides, the number of photos from different identities varies significantly, e.g., the minimum number was only 5, and such a small data size is not enough to train the proposed model. Therefore, we pre-process the data for our deep learning model.

4.2 Data pre-processing

In data pre-processing stage, data cleaning, redistribution, and data augmentation are performed.

First, we crop all the face images of identities in the train, val and test sets. All instances are then resized to 224x224 to fit the input size of the proposed model. Second, data cleaning is performed. The cropped images have different appearances, including the front face, the side face, and even the back of heads. The back of head does not contribute in recognition and could affect the performance of our deep network, therefore, these images are removed from the dataset. An example of these images/instances is shown in Fig. 3. Third, in our

Fig. 3 An example of poor quality instances i.e., head images

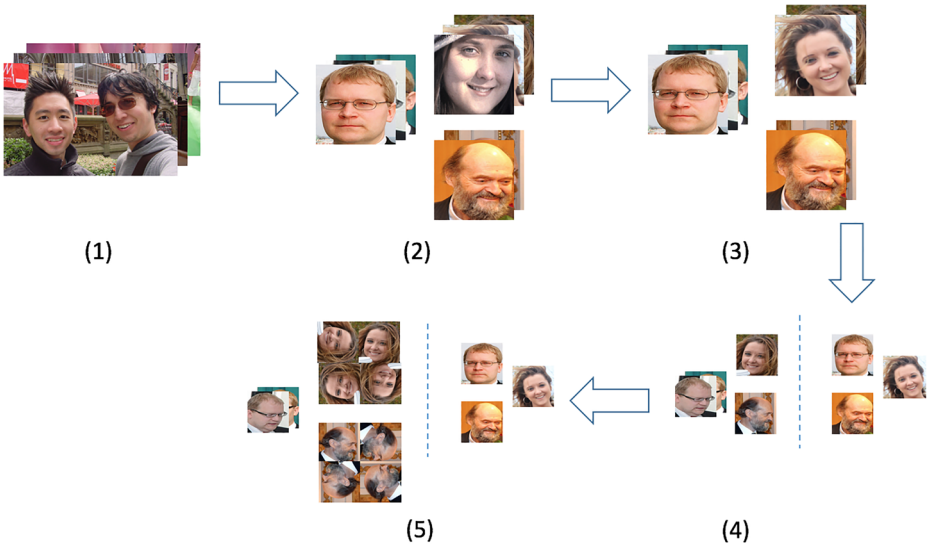

Fig. 4 Data pre-processing. (1) Personal photos in PIPA Dataset. (2) Re-sized head images cropped from original photos. (3) Good quality head images. (4) Training and test data. (5) Augmented training data

experiments the training and test images are randomly selected with the proportion 80% and 20%, respectively. Because the instance size of each class is different, the stratified sampling method is used for data allocation to avoid significantly biased results [5]. The last step is to perform data augmentation. We set 8 as the minimum number of instances per class. For the classes with insufficient instances, we perform data augmentation by rotating their instances by different angles, flipping and scaling. As a result of this, the numbers of instances in those classes are expanded. Figure 4 presents different steps involved in our data pre-processing.

After pre-processing and data augmentation, the dataset has 2356 classes, training set contains 41533 face images out of which 8613 of them are augmented. The test set contains 8230 face images. The distribution of images in the pre-processed dataset is shown in Table 1.

5 Experimental results

In the following, we first train the proposed model for the face classification task, then construct the desired community using using images.

Table 1 Statistics of the pre-processed dataset

Split	All	Train	Test
Instances (augmentation)	49763	41533 (8613)	8230
Identities	2356	2356	2356
Average identity	21.12	17.63	3.49

During training, we use Stochastic Gradient Descent (SGD) and back propagation to decrease the loss function. SGD randomly chooses one training instance at each step and calculates the gradients based only on that single instance. This speeds up the algorithm as it only manipulates little data at each iteration, especially on huge training sets [5]. In addition, to find the most satisfactory gradient, the learning rate is set to gradually decrease in the range of 0.005 to 0.00001 as the number of epochs increases. The learning rate changes every 30 epochs on average.

The model is trained to solve the multi-class classification task. It is assessed by top 1 error rate of classification. We compared the highest probability class of each sample with the actual classes, the top-1 error reflects the proportion of the number of incorrectly predicted samples to the total number of input samples.

5.1 Comparison with state-of-the-art

We compare our proposed technique with the state-of-the-art methods including the deep learning model naeil2 [19], DeepFace [30], VGG-19 [29], DiscFace [13], Improved LBP (ILBP) [1] and Deep Unified Model (DUM) [10]. We used the original parameter settings for these methods. Our experimental results are reported in Table 2.

As can be noted, classification accuracy of the proposed model is 86.87%, i.e., the top-1 error rate is 1-0.8687 = 0.1313. naeil2 [19], which fine-tuned the pre-trained AlexNet model using head images in PIPA Dataset achieved an accuracy of 83.88% [19]. DeepFace, VGG-19, DiscFace, ILBP and DUM achieved an accuracy of 76.66%, 84.23%, 82.58%, 81.86% and 85.78%, respectively on PIPA dataset. These results demonstrate the superior performance of the proposed technique, which relies on face images to predict social communities.

5.2 Implementation details

Our technique is developed in MATLAB. All our experiments have been performed on a machine with Intel Corei5 CPU and 16GB RAM.

5.3 Community prediction and formation

For community formation task, a complete community prediction system is devised that is built on top of our face recognition model. The proposed system is evaluated on PIPA dataset. The input to community prediction system is a face image of the target person, and a predicted network graph starts with the target person as its node as shown in Fig. 5.

Table 2 Comparison of the proposed technique with the state-of-the-art methods

Method	Accuracy (%)
naeil2 [19]	83.88
DeepFace [30]	76.66
VGG-19 [29]	84.23
DiscFace [13]	82.58
ILBP [1]	81.86
DUM [10]	85.78
Proposed Technique	86.85

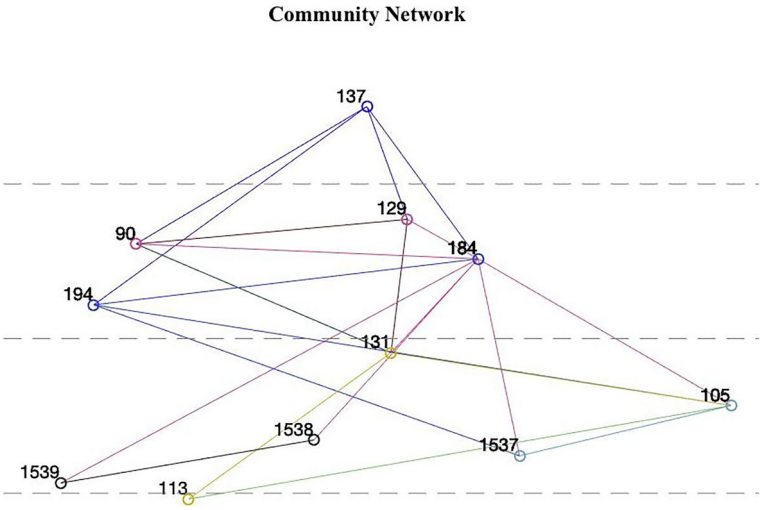

Fig. 5 An example of a predicted community

Figure 5 is divided into three parts by the dotted line, where each part represents a layer. The numbers beside the nodes are people's identity, e.g., person 137 is the target, and they are also the root of this network. Moreover, to enhance visibility, the edges emitted from the same node have the same color.

5.3.1 Performance of community formation task

Precision and Recall criterion is used for evaluating all the predicted networks. For each predicted label (predicted person name) in a network, it can only be judged whether it is consistent with the true label, no matter which class it belongs to. Thus, the multi-class classification tasks are converted into binary ones. Every predicted label is recorded as one of true negatives (TN), false positives (FP), false negatives (FN), or true positives (TP) base on the classification result.

Precision is the accuracy of the positive predictions [5], the equation is shown as:

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

where TP represents the number of people who exist in the networks and are correctly predicted. On the contrary, FP are wrongly classified person into the networks. However, precision is deceptive in some cases, for example, predicting one person as TP and to ensure that is correct, the precision is equal to 100%, but the network could not be constructed by the single person. Hence, precision is necessarily utilized along with recall, a.k.a. true positive rate (TPR). As (9) shows, FN represents the number of persons who should be in the communities and are not there.

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

The precision may be improved by setting a minimum face frequency. However, this negatively affects recall. In Fig. 5, we study this trade-off and observe that precision does not improve much whereas recall is severely affected. Thus, we set minimum face frequency

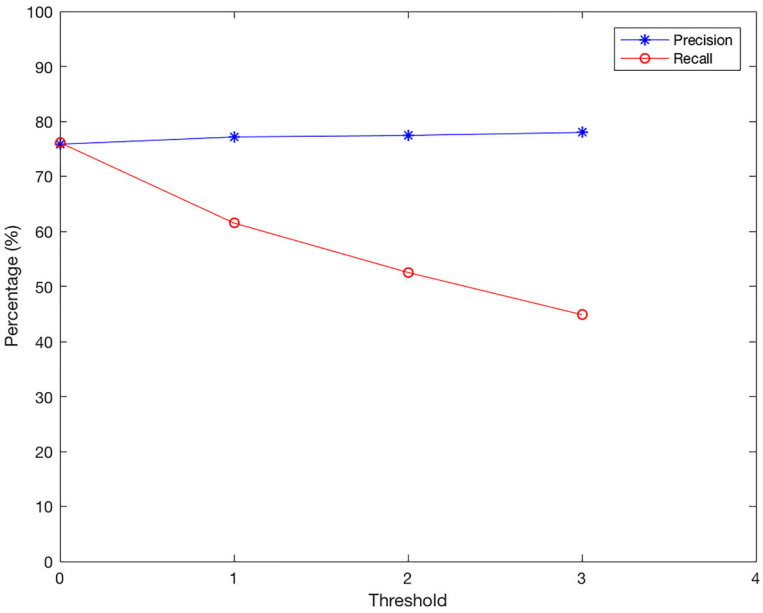

Fig. 6 Precision and recall versus the minimum frequency threshold

to be zero for the rest of the experiments. Only those individuals whose face frequency is greater than this threshold are classified into the corresponding community.

Figure 6 shows the precision and recall for different thresholds and frequencies. When the threshold increases, the precision is not significantly improved, however, the recall is greatly reduced. We therefore empirically set the threshold to 0. The achieved precision with this threshold is 75.84%, and the recall is 76.13%.

5.4 Prediction of relationship strength between candidates and analysis of network properties

To explore the relationship strength between candidates, we first calculate the IDF of each person using (6). The score of a photo is represented by the average of IDFs of all candidates in the photo. Once all the IDFs have been computed, the number of persons in each photo is calculated. The statistics of IDF and photo score is shown in Table 3.

As discussed in Section 3.2, the relationship strength between two candidates is obtained by summing the scores of photos in which both appear together. Hence, we improve our social network prediction system by enabling it to record the face co-occurrence frequencies and corresponding photo names simultaneously. Then, find the scores of those photos from the previously defined photo score library to calculate the relationship strength.

Table 3 Statistics of IDF and photo score

Type	Min	Max	Average
IDF	2.45	4.35	3.31
Photo Score	2.45	4.35	3.08

All scores that reflect the relationship strengths are then displayed in the final network graph. The example plots are shown in Figs. 7 and 8. These networks are built with two different target persons, respectively. The nodes are replaced by candidates' face images for visualisation, moreover, in addition to the scores on the edges, the thickness of the edges also reflects different relationship strength.

5.5 Analysis of predicted community

In this section, we analyze the whole community set after inputting all the test data to our proposed system and keeping all the distinct communities. By counting the size distribution of communities and the density of the communities, the cohesiveness of these predicted communities using image data is achieved. Community size and density are two of the primary network properties. Community size represents the number of nodes in a community. Community density refers to the Actual Connection-Maximum Connection ratio of a community as in (10):

$$D(s) = \frac{m_s}{n_s(n_s - 1)/2} \tag{10}$$

where m_s is the number of edges in network s, n_s represents the amount of nodes in the network. The more the edges, the denser is the community. A community is more cohesive when it has larger density and smaller size [2].

To further explore the networks, all the candidates are integrated into communities with different size. Figure 9 shows the size distribution of all communities. Although the largest

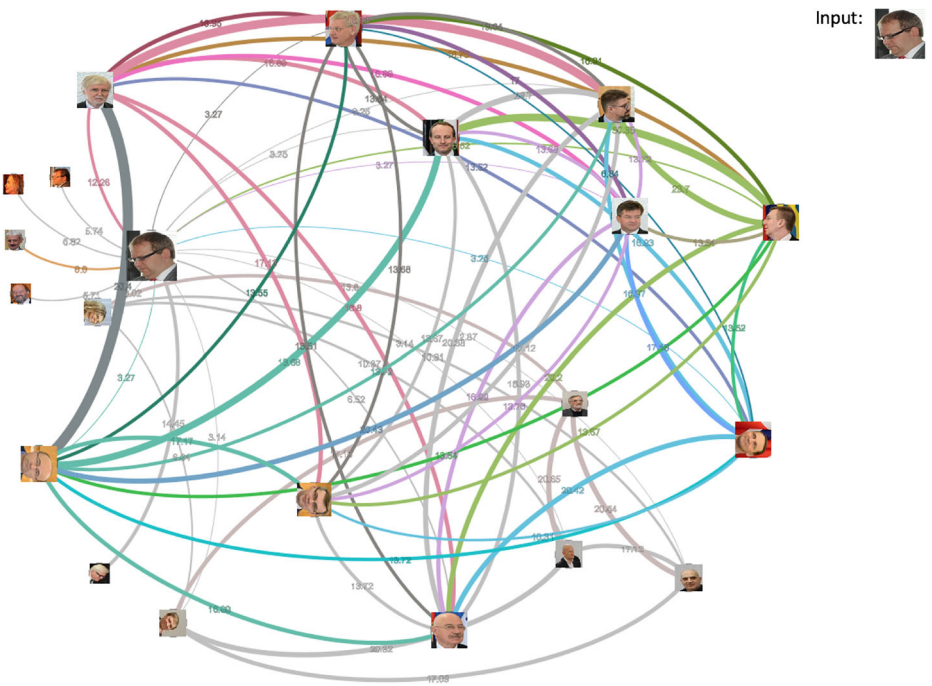

Fig. 7 Social network centered at person ID 1

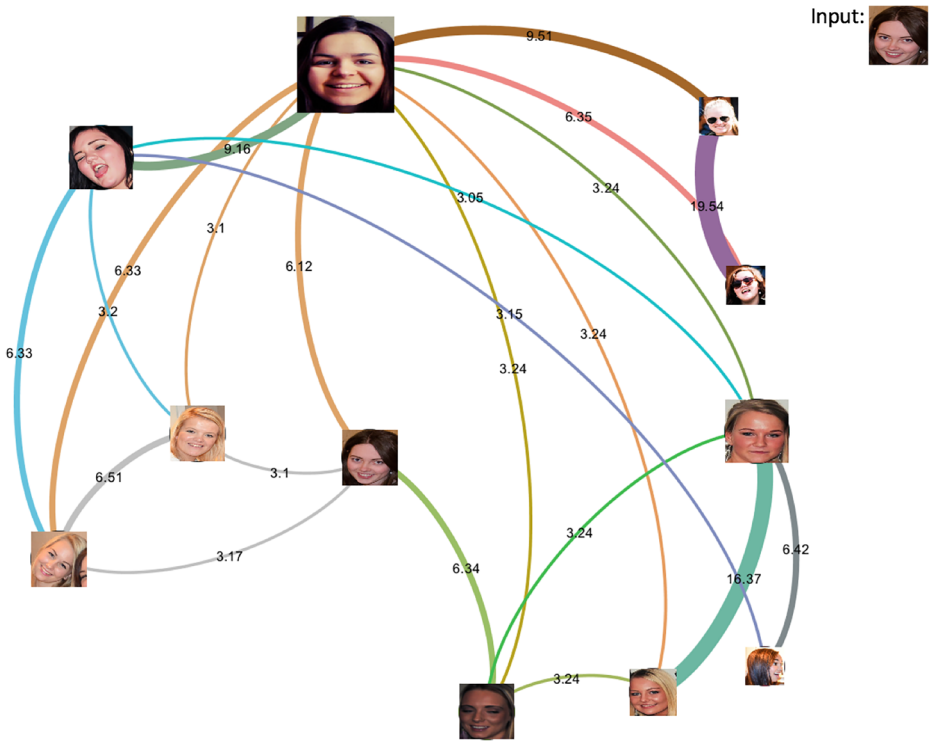

Fig. 8 Social network centered at person ID 137

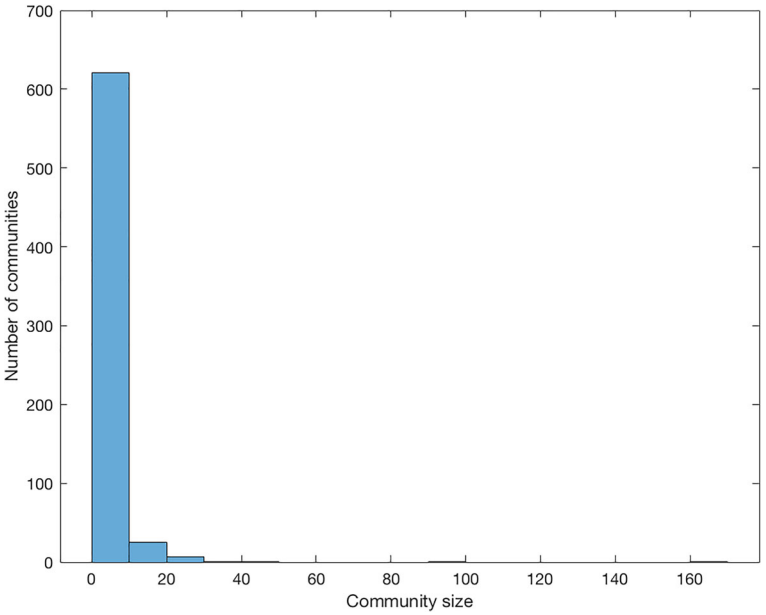

Fig. 9 Size distribution of communities

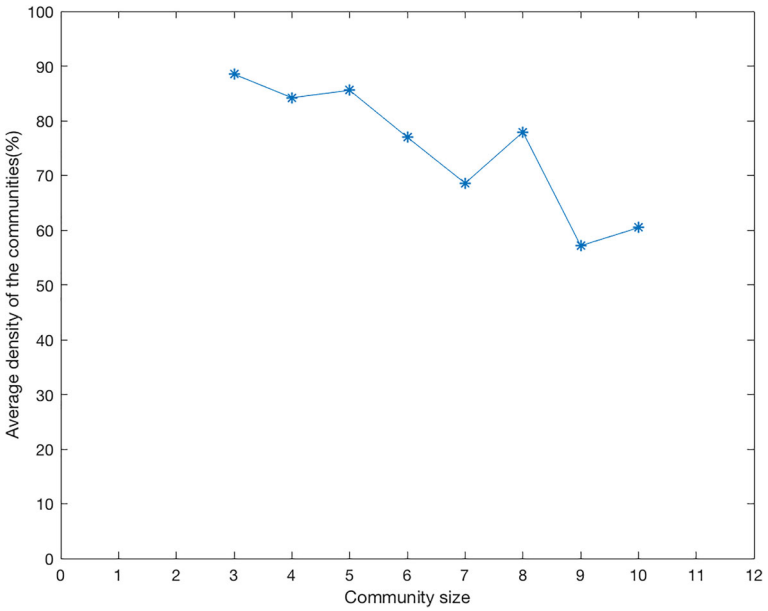

Fig. 10 Average network density

network size is 167, most of the network sizes predicted using image data contain fewer than 10 persons. Therefore, the analysis of network density focuses on the network size of 3 to 10. The average network density of each size is shown in Fig. 10. Although the network density gradually decreases as the network size expands, the minimum network density is still 57.14%.

To explore the features of image-generated community, a comparison between the communities predicted by the proposed technique and other social network communities built in [7] is also conducted. These communities include Twitter Friendship Network, Epinions Social Network, Wikipedia Vote Network and EU Email Communication Network. These four networks were constructed using textual data, such as user profiles, emails, and questionnaires.

Table 4 shows the statistics of network properties calculated on our image-generated community set and four text-generated networks. Intuitively, the communities predicted using PIPA dataset have smaller sizes and higher densities than other four networks. Because small network size and high network density lead to cohesive communities, this indicates that the predicted communities are cohesive.

Table 4 Statistics for comparison of image-generated network and text-generated network

Property	Twitter	Epinions	Wekipedia	Email	Image-generated
Size	500	500	500	500	3-10
Edges	3099	13739	11672	2396	2-45
Density	6.18	27.47	23.34	4.79	88.51-57.14

6 Conclusion and future work

In this paper, we propose a deep learning based social network prediction system, CommuNety. The input to our deep neural network is an image of a target person, and the output is a target-centered predicted community, which also presents the relationship strength of persons in the predicted network.

Due to lack of labeled image data and hardware limitation, data augmentation is used for the training of the proposed face recognition model. The training data is augmented by image rotation and all CNN features of images are fed to three fully connected layers to train the face recognition system. To predict and build social networks, face co-occurrence frequency technique is proposed to recognize people in the dataset who are directly or indirectly related to the target, and at the same time, use the face recognition model to classify each person's identity. As the deep neural network is the core of the proposed community prediction system, hence its classification accuracy affects the performance of the system. We also propose a photo ranking algorithm to rank photos in the data set based on the TF-IDFs of persons in the same photos. Consequentially, relationship strength of identities in network depends not only on the number of group photos, but also on the scores of these photos. This information is more valuable than simply constructing a social network. In addition, the networks predicted using image data are smaller and more cohesive than other social networks.

In our future work, we aim to optimize and further improve our community prediction system. We will consider using more complex deep learning model architectures and generate more training examples. Moreover, we also intend to explore other valuable information such as text and location of individuals from social networks and use them as additional features to improve the prediction of our proposed CommuNety system. We also intend to evaluate the effectiveness of the proposed technique on social media data, which contains additional information such as profiles, connections (e.g., friends, communities, followers) and communication (e.g., text, photos, share and likes).

Funding Open Access funding enabled and organized by CAUL and its Member Institutions

Data Availability The dataset analysed during the current study is available in the Github repository, https://github.com/coallaoh/PIPA_dataset.

Declarations

Conflict of Interests The authors have no conflict of interests to declare.

References

1. Bah SM, Ming F (2020) An improved face recognition algorithm and its application in attendance management system. Array 5:100014
2. Brunelli R, Falavigna D (1995) Person identification using multiple cues. IEEE Trans Pattern Anal Mach Intell 17(10):955–966
3. Chen YY, Hsu WH, Liao HYM (2012) Discovering informative social subgraphs and predicting pairwise relationships from group photos. In: Proceedings of the 20th ACM international conference on multimedia, pp 669–678. https://doi.org/10.1145/2393347.2393439
4. Dong Y, Liu Y, Lian S (2016) Automatic age estimation based on deep learning algorithm. Neurocomputing, 4–10. https://doi.org/10.1016/j.neucom.2015.09.115
5. Geron A (2017) Hands-on machine learning with scikit-learn and tensorflow. O'Reilly Media Inc, 2017
6. Guidi B, Michienzi A, De Salve A (2019) Community evaluation in facebook groups. Multimed Tools Appl, 1–20
7. Hashmi A, Zaidi F, Sallaberry A, Mehmood T (2012) Are all social networks structurally similar? In: 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, IEEE, pp 310–314
8. Hsu R, Abdel-Mottaleb M, Jain AK (2002) Face detection in color images. IEEE Trans Pattern Anal Mach Intell, 696–706. https://doi.org/10.1109/34.1000242
9. Hu W, Hu H (2019) Disentangled spectrum variations networks for NIR–VIS face recognition. IEEE Transactions on Multimedia 22(5):1234–1248
10. Khan MZ, Harous S, Hassan SU, Khan MUG, Iqbal R, Mumtaz S (2019) Deep unified model for face recognition based on convolution neural network and edge computing. IEEE Access 7:72622–72633
11. Khan S, Rahmani H, Shah SAA, Bennamoun M (2018) A guide to convolutional neural networks for computer vision. Synth Lect Comput Vis 8(1):1–207
12. Kim HN, Saddik AE, Jung JG (2012) Leveraging personal photos to inferring friendships in social network services. Expert Syst Appl, 6955–6966
13. Kim I, Han S, Park SJ, Baek JW, Shin J, Han J, Choi C (2020) Discface: minimum discrepancy learning for deep face recognition. In: Proceedings of the Asian conference on computer vision
14. Li J, Wong Y, Zhao Q, Kankanhalli MS (2020) Visual social relationship recognition. Int J Comput Vis, 1–15
15. Li S, Huang L, Zhang W, Tang B (2020) Hybrid feature fusion for person recognition in photo albums. In: MIPPR 2019: pattern recognition and computer vision, vol 11430, p 114300R. International Society for Optics and Photonics
16. Liu T, Yang B, Geng Y, Du S (2021) Research on face recognition and privacy in china—based on social cognition and cultural psychology. Front Psychol 12. https://doi.org/10.3389/fpsyg.2021.809736, https://www.frontiersin.org/article/10.3389/fpsyg.2021.809736
17. Mohapatra D, Patra MR (2019) Anonymization of attributed social graph using anatomy based clustering. Multimed Tools Appl 78(18):25455–25486
18. Nadeem U, Shah SAA, Bennamoun M, Togneri R, Sohel F (2021) Real time surveillance for low resolution and limited data scenarios: an image set classification approach. Inform Sci 580:578–597
19. Oh SJ, Benenson R, Fritz M, Schiele B (2017) Person recognition in social media photos. arXiv preprint arXiv: 1710.03224
20. Oro E, Pizzuti C, Procopio N, Ruffolo M (2017) Detecting topic authoritative social media users: a multilayer network approach. IEEE Trans Multimed 20(5):1195–1208
21. Ortiz EG, Becker BC (2014) Face recognition for web-scale datasets. ELSEVIER Comput Vis Image Underst 118:153–170
22. Parkhi OM, Vedaldi A, Zisserman A (2015) Deep face recognition. Br Mach Vis Conf
23. Pfeil U, Arjan R, Zaphiris P (2009) Age differences in online social networking – a study of user profiles and the social capital divide among teenagers and older users in myspace. Comput Hum Behav 643–654. https://doi.org/10.1016/j.chb.2008.08.015
24. Ramos J et al (2003) Using tf-idf to determine word relevance in document queries. In: Proceedings of the first instructional conference on machine learning. Citeseer, vol 242, no 1, pp 29–48
25. Sharif Razavian A, Azizpour H, Sullivan J, Carlsson S (2014) CNN features off-the-shelf: an astounding baseline for recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 806–813
26. Shah SAA, Bennamoun M, Boussaid F (2016) Iterative deep learning for image set based face and object recognition. Neurocomputing 174:866–874

27. Shah SA, Nadeem U, Bennamoun M, Sohel F, Togneri R (2017) Efficient image set classification using linear regression based image reconstruction. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 99–108
28. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition CoRR. arXiv: 1409.1556
29. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. International Conference on Learning Representations
30. Taigman Y, Yang M, Ranzato M, Wolf L (2014) Deepface: closing the gap to human-level performance in face verification. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1701–1708
31. Tseng WY, Chen KH, Huang JW (2019) Crowdsourced object-labeling based on a game-based mobile application. Multimed Tools Appl 78(13):18137–18168
32. Viola P, Jones MJ (2004) Robust real-time face detection. Int J Comput Vis 57(2):137–154
33. Xu X, Shimada A, Nagahara H, Taniguchi RI (2016) Learning multi-task local metrics for image annotation. Multimed Tools Appl 75(4):2203–2231
34. Xu L, Bao T, Zhu L, Zhang Y (2018) Trust-based privacy-preserving photo sharing in online social networks. IEEE Trans Multimed 21(3):591–602
35. Zhang N, Paluri M, Taigman Y, Fergus R, Bourdev L (2015) Beyond frontal faces: Improving person recognition using multiple cues. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4804–4813
36. Zhao Z, Yang Q, Lu H, Weninger T, Cai D, He X, Zhuang Y (2017) Social-aware movie recommendation via multimodal network learning. IEEE Trans Multimed 20(2):430–440
37. Zhang Z, Han J, Coutinho E, Schuller B (2018) Dynamic difficulty awareness training for continuous emotion prediction. IEEE Trans Multimed 21(5):1289–1301
38. Zhang F, Li S, Yu Z (2019) The super user selection for building a sustainable online social network marketing community. Multimed Tools Appl 78(11):14777–14798