

# Facial Components-based Representation for Caricature Face Recognition

Qiang Ma<sup>\*</sup> and Qingshan Liu

*School of Information and Control, Nanjing University of Information Science and Technology, Nanjing, 210044, China*

---

## Abstract

Caricature face recognition is an interesting but also difficult task due to the huge exaggeration between two different face modalities, photos, and caricatures. Therefore, we propose a new representation for recognition that is fused by the representation learned from photos, caricatures, and generated faces. Each generated face contains four main facial components. Photos, caricatures, and generated faces are sent to Photo-ResNet, Caricature-ResNet, and Generated-ResNet to learn specific representations. Then, the learned three representations are sent to a fully connected layer. We adopt Softmax loss and Center Loss for training, which can reduce the distance of intra-class. To test the performance of our proposed representation, we build a new dataset for caricature face recognition, which consists of 259 subjects, with 6490 caricatures and 8143 photos. The dataset we build is the biggest available caricature dataset. Several basic methods are used for caricature face recognition. To test the discrimination of our proposed representation, two more experiments are fulfilled, including searching photos according to the selected caricature (CTP) and searching caricatures according to the selected photo (PTC), and our proposed method performs better than other convolutional neural network (CNN)-based representations.

**Keywords:** caricature face recognition; generated face; facial components; Center Loss; caricature dataset

(Submitted on October 17, 2018; Revised on November 23, 2018; Accepted on December 26, 2018)

© 2019 Totem Publisher, Inc. All rights reserved.

---

## 1. Introduction

With the rapid development of face recognition, more and more attention is being paid to difficult conditions, such as pose, illumination, resolution, occlusion, and expression [1]. Caricature face recognition is a hard but not popular problem. The characteristic of a caricature is that the facial components may be exaggerated beyond possible or even replaced by other forms by the artist. Owing to this exaggeration or replacement, the identity of the face is easy for humans to recognize but difficult for computers.

Caricature face recognition is an interesting topic that researchers have attempted to deal with by creating a common subspace among several modalities, including facial sketches. In [2], facial sketches were divided in three kinds owing to the different degree of exaggeration compared with corresponding photos, such as viewed and forensic sketches and caricatures. Much research has been done to solve the recognition of viewed and forensic sketches.

Different from viewed sketches and forensic sketch face recognition, caricature face recognition is more different because the similarity of a photo and a caricature is much lower than viewed and forensic sketches. Of course, the exaggeration or replacement of specific facial components directly leads to the difficulty of caricature face recognition. Most existing caricature face recognition methods can be summed into two classes. The first class is getting different kinds of attributes from caricatures and then using them for recognition. Klare et al. [3] defined a series of face qualitative features. Abaci et al. [4] represented a face and a caricature with a set of characteristic features. Another category is learning caricature face representation using CNN. Crowley et al. [5] compared Fisher Vector and CNN representation for caricature face recognition. Huo et al. [6] used VGG [7] to obtain face representation for recognition. Considering caricatures are an exaggeration of facial components, we decide to use learned representation as the identity of a subject for caricature face

<sup>\*</sup> Corresponding author.

E-mail address: 20161221637@nuist.edu.cn

recognition. It is worth noticing that we use four exaggerated facial components to generate a new face image and learn representation from the generated face, which is additionally fused with representations learned from photos and caricatures.

In order to conduct more research on caricature face recognition, it is essential to have a comprehensive caricature dataset. Part of the existing caricature dataset can be seen in Figure 1. Most existing caricature datasets are too small; there are only 196 pairs of photos and caricatures in [3] and 200 in [4], which means methods based on deep learning are limited. Recently, a caricature dataset consisting of a total of 5974 photos and 6042 caricatures was proposed by Huo [6]. Different from [6], the dataset we propose consists of images with more complex head poses, even profiles.



Images from the four lines are separately built by Klare [3], Abaci [4], Crowley [5], and Huo [6].  
Figure 1. Part images from four previous available caricature datasets

In this paper, we first propose a new representation that is fused by the representations learned from photos, caricatures, and generated faces. The generated face is generated by four facial components, which is extracted from the caricatures according to the five face landmarks. Considering most available caricatures are too small, which limits the methods based on deep learning, we build a new caricature dataset that consists of 259 subjects, with 6490 caricatures and 8143 photos, and it will be available soon. Two groups of basic experiments are fulfilled, including recognition only using caricatures and recognition using caricatures and photos. Additionally, to prove the advantages of our proposed representation, two more experiments are fulfilled, including searching photos according to the selected caricature (CTP) and searching caricatures according to the selected photo (PTC).

## 2. Related Work

Caricature face recognition is an interesting but also difficult task for computers to implement. Limited research has been done to solve caricature face recognition, and limited datasets have been provided for caricature research.

Earlier studies have experimented to match photo with viewed sketches and forensic sketches. Viewed sketch face recognition is easy to fulfill because viewed sketches have high similarity with their corresponding photos. Viewed sketch face recognition has achieved good performance; for example, Liu et al. [8] adopted the KNDA (a nonlinear version of LDA)-based nonlinear discriminative classifier for sketch recognition to better describe nonlinear variations, and the accuracy reached 99% at rank-8. To overcome more difficult sketch face recognition problems, researchers began to shift their attention to forensic sketches. Tang et al. [9] proposed an Eigen transformation-based method to transform a photo into a sketch before matching. Klare et al. [10] cropped forensic sketches and photos into many patches and computed the Euclidean distances. Bhatt et al. [11] proposed Multi-scale Circular Weber's Local Descriptor (MCWLD) to encode the local face regions, and a memetic optimization was used to assign optimal weights to each region. Kotha et al. [12] used SURF to extract features in the form of 64 variable vectors for each image, and then the SURF descriptor vector was formed.

Compared with viewed sketches and forensic sketches, caricatures are harder for computers to recognize. Most existing caricature face recognition can be divided into two categories. The first category involves obtaining the different kinds of attributes from caricatures and then using them for recognition. Klare et al. [3] defined a series of face qualitative features, which were divided into two levels, general and component features, to label a caricature within 25 qualitative attributes.

Also, they proposed a caricature dataset that consists of 196 pairs of caricatures and corresponding photos, which is partly shown in the first line of Figure 1. What should be noticed is that all the caricatures and photos are corresponding, which means the pairs have almost the same pose, illumination, occlusion, or even expression.

Similar with [3], Abaci et al. [4] proposed a way to match caricatures to photos by representing a subject with a set of characteristic features, which can be automatically extracted from photos but manually from caricatures. A bigger caricature dataset was built, which has 200 pairs, part of which is shown in the second line of Figure 1. All the caricatures in [4] are gray level, hand-drawn caricatures. Similar with [3], all the photos are selected with the possible maximum similarity to the caricature.

Also, Ouyang et al. [2] accomplished caricature face recognition by learning a model to find the facial attributes independently. Different from all the methods mentioned above, a synergy was created by learning an embedding subspace using canonical correlation analysis (CCA) [13].

Another category is representation-based caricature face recognition, which means plenty of images are needed for learning caricature face representation. Crowley et al. [5] combined Fisher Vector and CNN-based representation with discriminative dimensionality reduction (DDR) or SVM classifiers for caricature face recognition. To learn the representation by CNNs, a much bigger dataset is built, part of which is shown in the third line of Figure 1. It is worth noticing that images in [5] are photos and paintings, which are similar with caricatures, and have similarity with corresponding photos to a certain extent.

Recently, a bigger caricature dataset was proposed by Huo et al. [6], and it has 252 subjects with a total of 5974 photos and 6042 caricatures and is partly shown in the fourth line of Figure 1. Each caricature has 17 landmarks. Different from the caricature labeling progress, landmarks in each photo were detected by Face++ [14]. They also used some basic methods, such as Gary, LBP, Gabor, SIFT [15], and VGG for caricature face recognition. More details about the mentioned caricature datasets can be seen in Table 1.

Table 1. More details about the mentioned five caricature datasets.

Caricature dataset	Subjects	Images	Features
Klare et al. [3]	196	239	Attributes
Abaci et al. [4]	200	400	Two level attributes
Crowley et al. [5]	DEV	1088	Fisher Vector and CNN
	NPG	188	
	Train	496	
Huo et al. [6]	252	12016	Gary, LBP, Gabor, SIFT, and CNN
Our	259	14633	Gary, LBP, VGG16, and ResNet

### 3. Facial Components-based Representation

Most existing work is focused on extracting facial attributes for caricature face recognition. These methods may perform well on small level caricature datasets under the condition of high similarity. When the number of caricatures becomes bigger and the similarity becomes lower, the performance of these methods will be worse. To accomplish more difficult caricature face recognition tasks, we propose a new representation that is fused by representation learning from different modalities, including generated faces, which are generated by four facial components. Before generating the faces, five landmarks (two eyes, nose, and two mouth corners) will be marked by face alignment in photos and manually in caricatures.

After alignment, four facial components in caricatures can be extracted, which will be used to generate a face with these components fixed to a specific position. After generating the caricature faces, photos, caricatures, and generated faces are sent to an independent network for feature learning. Three networks have the same structure and the learned representations have the same size, which benefits from the same size of input images. After learning the representation from three modalities, three representations are connected as one, which will be sent to a fully connected layer. To speed up the training progress and accelerate the speed of convergence, we adopt Softmax loss along with Center Loss, which can effectively reduce the distance of intra-class. The framework of our work can be seen in Figure 2.

#### 3.1. Facial Components

It is obvious that the eyes, nose, and mouth play an important role in recognition. Additionally, caricatures involve the exaggeration or replacement of the main facial components. Therefore, we propose a new representation learned from these

components. What we need to do first is to extract the facial components using the five landmarks. We can get the four facial components according to the five landmarks. Then, we pick out all the wrong parts and label again for accurate extraction. Different from the labeling process mentioned above, eight new points are labelled, which means each component has two points, ensuring the area of each component. Then, all the components can be extracted according to the new eight landmarks.

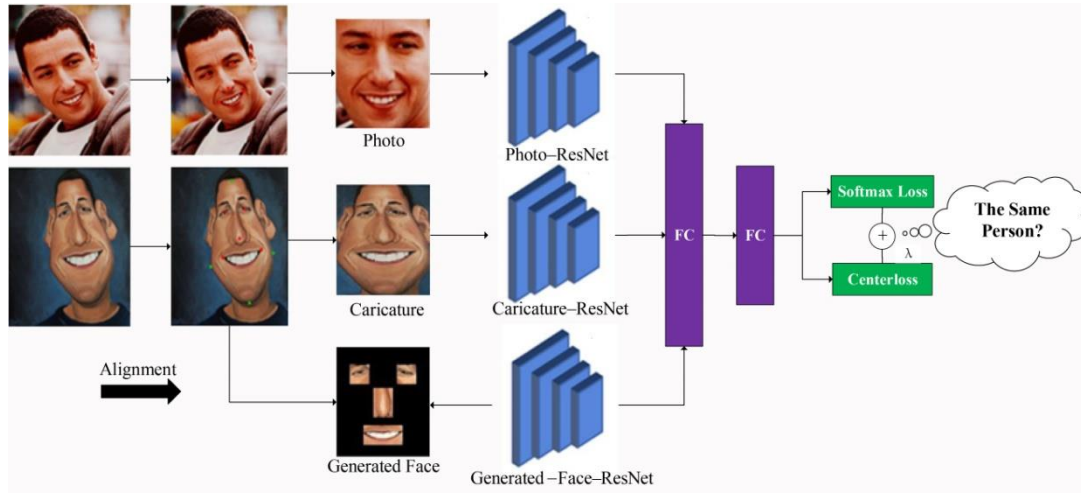


Figure 2. The framework of the caricature face recognition progress. FC represents fully connected layer

In order to use these four facial components to enhance the feature representation and make the fused representation more representative and discriminative, we restructure the extracted four facial components to generate a new face. Thus, after extracting all the face components, we resize them to a unified size. Then, we put those facial components to certain positions. For example, the middle of the left eye and right eye, nose, left corner of mouth, and right corner of mouth are put at the points of (24,28), (72,28), (48,64), (28,102), and (68,102) respectively in a  $112 \times 96$  black background image. An example of a generated face can be seen in the third line of Figure 2. Obviously, the generated face looks like a real face because it is generated using only four facial components, and the construction of all the components is the same as in a photo.

### 3.2. Implementation Details

We adopt ResNet [16] to extract representations of photos and caricatures separately. The structure of our ResNet can be seen in Table 2.

Table 2. The structure of one of the ResNet in our framework

Layer	Filter / Stride / Padding	Output Size
Input		$3 \times 112 \times 96$
Conv1a	$3 \times 3 / 1 / 0$	$32 \times 110 \times 94$
Conv1b	$3 \times 3 / 1 / 0$	$64 \times 108 \times 92$
Pool1	(MAX: $2 \times 2 / 2 / 0$ )	$64 \times 54 \times 46$
Conv2	$3 \times 3 / 1 / 0$	$128 \times 52 \times 44$
Pool2	(MAX: $2 \times 2 / 2 / 0$ )	$128 \times 26 \times 22$
Conv3	$3 \times 3 / 1 / 0$	$256 \times 24 \times 20$
Pool3	(MAX: $2 \times 2 / 2 / 0$ )	$256 \times 12 \times 10$
Conv4	$3 \times 3 / 1 / 0$	$512 \times 10 \times 8$
Pool4	(MAX: $2 \times 2 / 2 / 0$ )	$512 \times 5 \times 4$
Res5		$1536 \times 5 \times 4$
FC6		$1 \times 1 \times 2048$
FC7		$1 \times 1 \times 259$

It is worth noticing that there is one residual part between Pool1 and Conv2, two between Pool2 and Conv3, five between Pool3 and Conv4, and three between Pool4 and Res5. After each convolution layer, there is a PReLU layer. Moreover, we adopt Center Loss [17] for training our model to reduce the intra difference of each subject. In our paper, ResNet is used for representation learning, and ResNet for different modalities does not share parameters with each other, so the final fusion can be more representative for each subject. With the help of Center Loss, the model can be trained more suitably for cross-modality recognition. As the size of the input image is  $112 \times 96$  and the input images are three channels,

each ResNet can learn a representation of the size of  $512 \times 5 \times 4$ . After fusion, the representation is sent to a fully connected layer, and the size of the feature will be  $1 \times 1 \times 2048$ .

Considering limit work has been done to solve the caricature face recognition task, we train our model with all weights in residual parts initialized from a zero-centered normal distribution with a standard deviation of 0.01, and the bias is initialized to 0. We use the way of multi-step to train our model, and the base learning rate is 0.01. The decay rate is 0.1, and the learning rate will decrease after every 20000 iterations. The momentum is 0.9, and the decay of weight is 0.0005.

## 4. Caricature Dataset

### 4.1. Implementation Details

As we can see in Table 1, most available caricature datasets are too small. Therefore, we build a new caricature dataset first, which will be made available. Before collecting all the images, we first list the names of famous people involved in different kinds of jobs, such as singers, film stars, politicians, and athletes, and these people are from different countries. According to the list of names, we download images from Google Image search and Pinterest, which contain a great quantity of different types of images. After downloading all the images, we remove the images that are unrecognizable or do not belong to the corresponding person and reorder the images. Part of our dataset can be seen in Figure 3(a). From (a), we can see that both the faces in photos and the faces in caricatures have variable poses. Meanwhile, the similarity between photos and caricatures is lower, which makes the caricature face harder to recognize.



(a) Part images of our caricature dataset



(b) The landmarks on photo and caricature

Figure 3. Part of our caricature dataset and the labeling examples

### 4.2. Caricature Face Labeling and Alignment

As all the images are checked, we use Multitask CNN (MTCNN) [18] to accomplish face detection and face alignment for photos. Unlike photos, landmarks on caricatures cannot be localized easily because the structure of a caricature face is unstable. Thus, for the purpose of accurate labeling, caricatures are labeled manually, which may cause a bit of deviation. For each face in a photo, there are five landmarks, numbered 1 to 5, representing the two eyes, nose, and mouth. Unlike photos, face detection for caricatures is hard to accomplish because all the caricatures are drawn by different artists, so the style of exaggeration can be quite different. Before labeling in caricatures, we first rotate the face according to the two eyes, ensuring the two eyes are on the same horizontal line. After that, four more landmarks, numbered 6 to 9, are labeled in the caricature and can be used to extract the caricature face. One example of the landmark localization can be seen in Figure 3(b).

After labeling all the images, what we need to do next is to crop the face image for caricature face recognition. With the help of the five labeled landmarks, numbered 1 to 5, we can easily crop the faces from photos, owing to the high similarity of the structure between human faces. Face alignment in caricatures is harder, so we cannot extract the face using the same way

as in photos. After rotating the caricature faces, we use four landmarks, numbered 6 to 9, to ensure the region of a caricature face. Considering the exaggerated part may be the chin, forehead, or cheek, we cannot fix the location of the two eyes.

## 5. Experiments and Results

Before caricature face recognition, we need to conduct some basic experiments for comparison. To compare with the proposed representation, Gary, LBP, VGG16, and ResNet are used to learn the representation. We classify our experiments into two groups: 1) recognition using only caricatures and 2) recognition using caricatures and photos. Both are the baseline experiment of our proposed dataset. After finishing these basic experiments, we use the trained model for searching photos according to the selected caricature (CTP) and searching caricatures according to the selected photo (PTC).

### 5.1. Experiments

In the first group of experiments, we only use caricatures to accomplish caricature face recognition tasks. Gary, LBP, VGG16, and ResNet are used to learn representation. Before training, all the images are cropped. For photos, we use MTCNN to detect the face, while we do it manually for caricatures. All the images are cropped at the size of  $112 \times 96$ . The learned representation can be classified into two categories. The first category is hand-crafted representations, such as Gary and LBP. The other category is the representation learned by CNNs, including ResNet and VGG16 [19].

For hand-crafted representations, we first turn all the images into grayscale. After that, we use the gray map of caricatures to be the Gary representation, while LBP is used to learn the LBP representation. Considering the hand-crafted representation is low-level and is carefully designed, we use multi-SVM [20] to solve the multi-classification problem. It is noteworthy that the rate of training and testing samples is 9:1.

For CNN-based feature representation, we choose VGG16 and ResNet to learn representations because of their excellent performance in face recognition. Different from the commonly used VGG16 and ResNet, we use Center Loss combined with Softmax loss to train the model, which can reduce the intra difference compared with only using Softmax loss and accelerate the speed of convergence. What differs from hand-crafted representations is that the photos and caricatures of each person are paired with each other for data augmentation. The total number of training samples is more than 140 thousand, and the testing samples are more than 4 thousand.

In the second group of experiments, we fuse the representations learned from photos and caricatures. After getting the learned representations from photos and caricatures separately, we fuse these two kinds of representations and use the fused representation as the final representation of the input images to train the classifier. Taking two categories of representation into consideration, we train the model in the same way of the first group of experiments.

Additionally, to test the discrimination and advantages of our proposed representation, two more experiments are done, including searching the corresponding photos for a specific caricature (CTP) and searching the corresponding caricatures of a specific photo (PTC). Considering that the CNN-based representations can learn deeper information than hand-crafted representations, we just use VGG16, ResNet, and the method we proposed to accomplish the searching task. For example, if we are doing the PTC experiments, we first select several photos randomly among all the photos, and each of the selected photos is paired with all the caricatures as the testing set. All the pairs are sent to the trained model to get a probability representing the possibility that the paired caricature belongs to the subject of the selected photo. According to the testing result, we compare the result with the actual corresponding caricatures to calculate the top-1, top-5, and top-10 accuracy.

### 5.2. Results

The accuracy of the two groups can be seen in Table 3, and the number in the brackets indicates which group it belongs to. Compared with the two groups, it is obvious that fusing representation learned from the photos can improve the performance of caricature face recognition. It is worth noticing that most of the existing caricature face recognition methods use the representation fused by caricatures and photos.

Moreover, compared with representation fused by hand-crafted representations, representation based on CNNs performed better than hand-crafted representations. This is not only because the CNNs can learn a much deeper representation, but also because the representations learned by CNNs are more discriminative for each subject, with the help of Center Loss and a large number of samples.

Table 3. The performance of the two groups of experiments

Methods	Top-1(%)	Top-5(%)	Top-10(%)
Gary (1)	0.7790	2.7264	3.9922
LBP (1)	0.1947	1.5579	3.7975
VGG16 (1)	33.8851	56.8647	64.5570
ResNet (1)	34.0798	57.6435	66.7965
Gary (2)	6.7268	20.5906	30.4348
LBP (2)	1.0664	5.2502	8.2855
VGG16 (2)	58.6881	76.2531	82.3308
ResNet (2)	67.6065	82.0384	86.8421
Our	<b>71.9089</b>	<b>85.9023</b>	<b>89.5155</b>

From Figure 4, we can obviously see that compared with the performances of those basic experiments, our proposed method, whose representation is fused by photo face, caricature face, and the generated face, performs better than other baseline methods. From Table 3, the top-1, top-5, and top-10 performances of our proposed method are 71.91%, 85.90%, and 89.52% respectively. Also, we can see that the method we proposed improved by 4% in the top-1 performance compared with the best baseline performance, ResNet. As we mentioned above, caricatures are locally exaggerated images where most of the exaggerated parts are facial components. Applied with these features learned from facial components, the fused features can be more discriminative for computers to recognize.

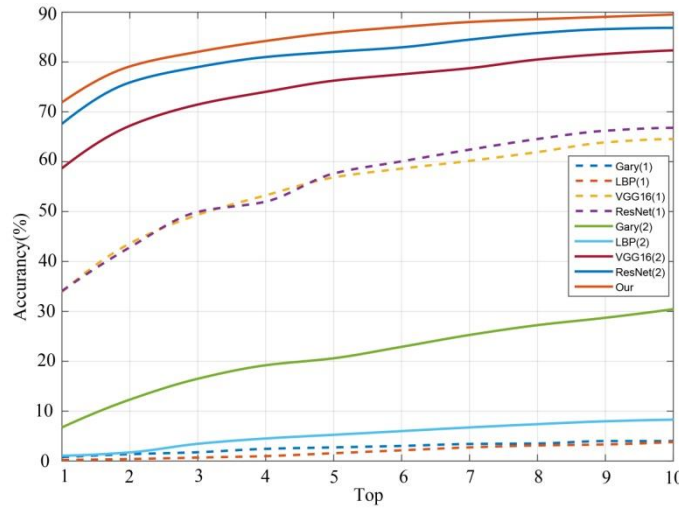


Figure 4. The Top-1 to top-10 accuracy of the two groups of experiments

After training the model for caricature face recognition, we can use the trained models to accomplish the tasks of searching photos according to the selected caricature (CTP) and searching caricatures according to the selected photo (PTC). The performance of CTP and PTC tasks can be seen in Table 4 and Figure 5.

From Figure 5, it is obvious that the performance of our proposed method is better than that of the other CNN-based methods in both CTP and PTC tasks. From Table 4, we can see that for the CTP task, our proposed representation is 1.5% higher than ResNet, and it is 2% higher for the TCP task. Additionally, the top-1 to top-10 performance of our proposed method is always higher than others. Therefore, the representation combined with facial components is more discriminative than others. It is remarkable that the accuracy of CTP is higher than that of PTC, which to some degree means that the representation learned from caricatures is more representative than that learned from photos.

Table 4. The performance of CTP and PTC on the proposed caricature dataset

Method	CTP			PTC		
	Top-1(%)	Top-5(%)	Top-10(%)	Top-1(%)	Top-5(%)	Top-10(%)
VGG16	64.7059	78.2933	80.8616	60.3700	68.2571	76.7283
ResNet	69.3455	79.3703	82.6843	65.4333	72.4440	79.4547
Our	<b>70.8368</b>	<b>80.7788</b>	<b>82.8500</b>	<b>67.5754</b>	<b>76.5336</b>	<b>82.4732</b>

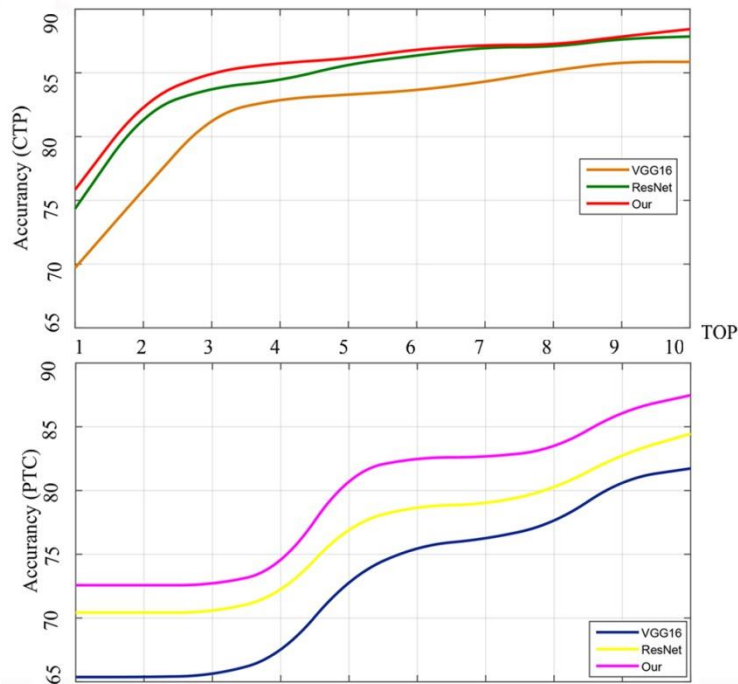


Figure 5. The performance of CTP and PTC on our proposed dataset

## 6. Conclusions

In this paper, we first release a new caricature dataset for caricature face recognition that consists of 259 subjects, with 6490 caricatures and 8143 photos. Each image has five landmarks representing four facial components, and each caricature has four more landmarks for extracting the caricature face. Then, we propose a new representation for recognition, which is fused by the learned representation learned from photos, caricatures, and the corresponding generated faces.

Two groups of basic experiments are fulfilled to prove the advantages of our proposed representation. Our method achieves 71.91% recognition success on the proposed caricature dataset. Also, two more experiments, CTP and PTC, are conducted for the verification of the discrimination of our proposed method. From the results, the representation we proposed is more discriminative for caricature face recognition. For further research, we will focus more on caricature face detection and landmark localization.

## References

1. Z. Yu, Q. Liu, and G. Liu, "Deeper Cascaded Peak-Piloted Network for Weak Expression Recognition," *The Visual Computer*, Vol. 34, No. 12, pp. 1691-1699, 2017
2. S. Ouyang, T. Hospedales, Y. Z. Song, and X. Li, "Cross-Modal Face Matching: Beyond Viewed Sketches," in *Proceedings of Asian Conference on Computer Vision (ACCV)*, pp. 210-225, Springer, Cham, 2014
3. B. F. Klare, S. S. Bucak, A. K. Jain, and T. Akgul, "Towards Automated Caricature Recognition," in *Proceedings of 2012 5th IAPR International Conference on Biometrics (ICB)*, pp. 139-146, IEEE, 2012
4. B. Abaci and T. Akgul, "Matching Caricatures to Photographs," *Signal, Image and Video Processing*, Vol. 9, No. 1, pp. 295-303, December 2015
5. E. J. Crowley, O. M. Parkhi, and A. Zisserman, "Face Painting: Querying Art with Photos," in *British Machine Vision Conference (BMVC)*, pp. 65.1-65.13, 2015
6. J. Huo, W. Li, Y. Shi, and H. Yin, "WebCaricature: A Benchmark for Caricature Face Recognition," *arXiv preprint arXiv:1703.03230*, 2017
7. K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv preprint arXiv:1409.1556*, 2014
8. Q. Liu, X. Tang, H. Jin, and S. Ma, "A Nonlinear Approach for Face Sketch Synthesis and Recognition," in *Proceedings of CVPR 2005 and IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1005-1010, IEEE, 2005
9. X. Tang and X. Wang, "Face Photo Recognition using Sketch," in *Proceedings of International Conference on Image Processing*, Vol. 1, pp. I-I, IEEE, 2002
10. B. Klare, Z. Li, and A. K. Jain, "Matching Forensic Sketches to Mug Shot Photos," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 3, pp. 639-646, 2011
11. H. S. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa, "Memetic Approach for Matching Sketches with Digital Face Images,"

*IEEE Transactions on Information Forensics and Security*, Vol. 7, No. 5, pp. 1522-1535, 2012

12. D. K. Kotha and S. Rath, "Forensic Sketch Matching using SURF," in *Proceedings of International Conference on Advances in Computing*, pp. 527-537, Springer, New Delhi, 2013
13. Y. Gong, Q. Ke, M. Isard, and S. Lazebnik, "A Multi-View Embedding Space for Modelling Internet Images, Tags, and their Semantics," *International Journal of Computer Vision*, Vol. 106, No. 2, pp. 210-233, 2014
14. E. Zhou, Z. Cao, and Q. Yi, "Naive-Deep Face Recognition: Touching the Limit of LFW Benchmark or Not?" *arXiv preprint arXiv:1501.04690*, 2015
15. C. Geng and X. Jiang, "Face Recognition using Sift Features," in *Proceedings of 16th IEEE International Conference on Image Processing (ICIP)*, pp. 3313-3316, 2009
16. K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778, 2016
17. Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A Discriminative Feature Learning Approach for Deep Face Recognition," in *Proceedings of European Conference on Computer Vision (CVPR)*, pp. 499-515, Springer, Cham, 2016
18. K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint Face Detection and Alignment using Multitask Cascaded Convolutional Networks," *IEEE Signal Processing Letters*, Vol. 23, No. 10, pp. 1499-1503, 2016
19. W. AbdAlmageed, Y. Wu, S. Rawls, S. Harel, T. Hassner, I. Masi, et al., "Face Recognition using Deep Multi-Pose Representations," in *Proceedings of 2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1-9, 2016
20. V. Franc and V. Hlavác, "Multi-Class Support Vector Machine," in *Proceedings of 2002 16th International Conference on Pattern Recognition*, Vol. 2, pp. 236-239, IEEE, 2002

**Qiang Ma** is a Master's student in the School of Information and Control at Nanjing University of Information Science and Technology. His research interests including machine learning and face recognition.

**Qingshan Liu** received his M.S. degree from Southeast University in 2000 and his Ph.D. from the Chinese Academy of Sciences in 2003. From 2010 to 2011, he was an assistant research professor in the Department of Computer Science at the Computational Biomedicine Imaging and Modeling Center at Rutgers University. Prior to that, he was an associate professor in the National Laboratory of Pattern Recognition at the Chinese Academy of Sciences. From 2004 to 2005, he was an associate researcher in the Multimedia Laboratory at the Chinese University of Hong Kong. He is currently a professor in the School of Information and Control at Nanjing University of Information Science and Technology. His research interests include image and vision analysis.