

Photobook Creation Using Face Recognition and Machine Learning



N. Aishwarya, N. G. Praveena, B. S. Akash Arumugam, and J. Pramod

Abstract Nowadays, mobile photography plays a pivotal role in major parts of human lives. Every individual captures a lot of photographs in their personal device. However, management of these photographs by person becomes very crucial. Hence, photobook provides an attractive solution for personal storing and printing. In fact, today many commercial products are available to support the above point. But all of them require human annotation which is a time-consuming process and also leads to unreliable authentication. Hence, in this work, face recognition and deep learning techniques are utilized for photobook creation, wherein the user can create albums for a particular group of people or individuals and annotate them. Also, for every new photograph captured, the photobook will update itself automatically. Further, user interaction is introduced in the proposed work which makes it more reliable and results in good performance.

Keywords Face recognition · Convolutional neural networks (CNNs) · Rectifying linear unit (ReLU) · Computer vision (CV)

1 Introduction

With the impetuous advancement in the digital world, people capture zillions of photographs every single day, and managing and organizing photographs manually is a laborious task and time-consuming process [1–3]. Photobook is found to be an attractive solution to overcome this limitation, which can be automated by albuming. Today, many commercial products are readily available to serve this purpose. But they involve human annotations, which will be a herculean process that very few will prefer. Therefore, it is extremely recommendable to automate this cataloging and organizing process. Organizing photographs into albums is one way to keep similar photographs together and makes it easily accessible to the user. Automatic

N. Aishwarya (✉)

Amrita School of Engineering, Amrita Vishwa Vidyapeetam, Chennai, India

N. G. Praveena · B. S. Akash Arumugam · J. Pramod

R.M.K College of Engineering and Technology, Anna University, Chennai, India

annotation of faces in photographs can be done using face detection and face recognition algorithms. In the past decades, extensive studies have been done in the field of computer vision and machine learning for the above techniques [4–6]. With the advent of many vigorous face detection algorithms, the face annotation process does not appear to be an intricate process. However, the challenge still lies in the process of face recognition. With the advent of many robust face detection algorithms, the process of face annotation does not seem to be an intricate process. However, the challenge still lies in the process of face recognition. On the one hand, due to high variance in illumination, facial expressions, cues, poses, etc., in real-time pictures, it is often tough to extract precise facial features and develop an efficient face model. On the contrary, the existing face recognition algorithms fail to annotate all the faces in group photographs which greatly reduce their efficiency for real-time implementation [7]. Hence, deep learning technology using convolutional neural networks (CNNs) is proposed in this paper to address the above shortcomings.

CNN is a special type of neural network that transforms the given input images into corresponding feature maps using input layer, hidden layer which has convolutional layer, pooling layer, and output fully connected (FC) layer [8–10].

In this work, firstly, CNN is applied to an image to extract the set of unique complex facial features that include length and breadth of the face, width of the nose, eyes, lips, space between the eyes, skin tone, texture, etc. Then, these facial features are compared with the features obtained from the training data set. If the facial features are highly correlated with the training data set, then the image will be added to an existing album. Otherwise, a new album will be created for the annotated image.

The remaining paper is organized as follows. Related works are briefly described in Sect. 2. Section 3 explains the proposed system. Section 4 concludes the paper and future work.

2 Related Work

There is an extensive research on photobook creation using face recognition and machine learning. In this subsection, a brief description of the work done previously in the automatic photobook management was explained.

In [11, 12], two clustering methods were proposed in which the features of the representative images are taken from the data set. Platt [13] adopted a photograph clustering approach in which the temporal information was used for creating auto album. Platt et al. [14] extended their work proposed in [13] in which new interface was developed for assisting the users to get the photographs more efficiently.

In [15], clustering-based approach was proposed for the photograph selection from the set of photographs for creating an album. Papadopoulos et al. [16] developed a clustering approach for photograph representation related to landmarks and events in a particular place. In [17, 18], the face tagging was widely explored in photograph clustering technology based on the features extracted from the face and landmarks.

In this, photographs of same faces were grouped and formed a cluster and hence tagging cost was reduced. Suh et al. [19] grouped the faces into clusters and then labeled them which results in lack of supervision.

Chen et al. [20] proposed a semiautomatic family photo album framework which could categorize the faces in different angles of the same person into a small number of clusters very accurately and effectively. However, for better predication and retrieval of the photographs, time information for face annotation is not carried out. Yuansheng et al. [21] proposed a face album management system composed of two pools: certain and uncertain for organizing the photographs by identity. When more faces have to be recognized, both the pools work together for album organization. If new face is detected, it is initially stored in uncertain pool and later it is recognized with the new faces and then stored in the pool. However, recognizing such faces of conventional identities and discovering new identities in album is difficult.

Yan et al. [22] introduced CNN for face recognition in which Caffe was used for feature extraction during the training and validation process. The facial features were classified automatically on the nine-layered network. In [23], a subset of images was selected from an initially labeled data set and added to the training data set. Synthetic images were generated by feeding noise to every training sample. Once the training data set is created, these were given to the CNN to generate the trained model. The testing set is then recognized with the above trained network. This type of learning comes under supervised learning which involves human-annotated data and demands huge memory capacity and additional computational power due to deep learning models.

Coskun et al. [24] proposed face recognition process consisting of three stages: Preprocessing stage, then color space conversion and resizing of images were developed which extracts the facial features of the images using trained data set. Further, Softmax classifier is used to develop the classified resultant stage. In [25], Google Photos has built-in facial recognition similar to that of Facebook and Apple Photos. This software organizes photographs based on the faces by using an algorithm that identifies and groups photographs of people to help users easily find them. Google Photos provides unlimited online storage of digital images and videos. The original images are compressed after uploading, but the difference is hardly visible. It offers online backup as the images and videos are available in all the devices owned by the user. But, facial recognition feature is prohibited in few countries due to certain privacy laws. It also moves all the images and videos from device to the cloud.

In [26], iPhone has search option to find the pictures of a person, places, object, or event. Recognized faces can be labeled with a name. All the photographs of a person can be found under his/her label. iPhone groups the library photographs into Moments and Collections arranged by time and location. Album feature in iPhone consists of Created albums, Shared albums, and automatic collections of various types of photographs and videos like panoramas, selfies, slow motion, and time lapse. Photographs can also be manually added to the albums by the user. iPhone has made it easier to store the photographs online and access them from any device. The simulation results of the proposed method are found to have better results, but high computation time and supervised learning are essential to train the data set with respect to input images.

3 Proposed System

In this section, the proposed method is presented in detail that shows how CNN helps to extract the crucial features of an image to create a photobook.

3.1 Overview of Convolutional Neural Network

In the recent past, deployment of convolutional neural network (CNN) has led to prodigious success in many pattern recognition tasks. This is mainly due to the ability of CNN to work in a visual system similar to that of a human brain. One of the most exciting applications of pattern recognition which is the focal point of the proposed work is face recognition. A convolutional neural network basically has an input and output layer, along with multiple hidden layers. The hidden layers of the CNN consist of convolutional layers, ReLU layers, pooling layers, fully connected layers, and normalization layers. They are driving major growth in the field of computer vision (CV), which has noticeable applications for self-driving cars, drones, robotics, medical diagnoses, security, and treatments for the visually impaired persons.

3.2 Feature Extraction

The main motive of this work is to create a photobook which groups the images into appropriate album based on human facial features. Figure 1 shows the overview of the proposed method.

Initially, the training data set is formed randomly from a collection of images captured by multiple persons. The image is captured using digital camera. Then, the face detection technique using Haar cascade classifier [27] is used to detect the presence of a human face or a group of human faces. Then, the detected faces are resized to lower dimension and convoluted with different filters like horizontal edge, vertical edge, blur filter, sharpening filter to extract the features of the face. This forms the convolutional layer. Figure 2 shows the convolutional layer for an image segment. For example, a segment of input image of size 7×7 is taken and four convolutional filters are applied to it. The resultant convoluted matrix and the convoluted image segment are shown in Fig. 2.

After obtaining the convoluted layers, they are given to rectifying linear unit (ReLU) to increase the nonlinearity. The convolution operation and the activation function of ReLU in CNNs are given in Eq. 1.

$$y^i = \max\left(0, \sum_i k^{ij} \otimes x^i + b^j\right) \quad (1)$$

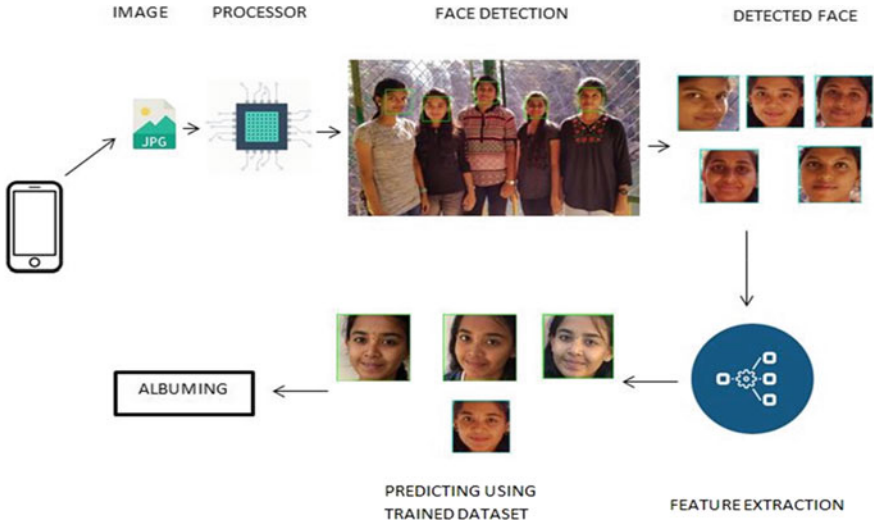


Fig. 1 Overview of the proposed system

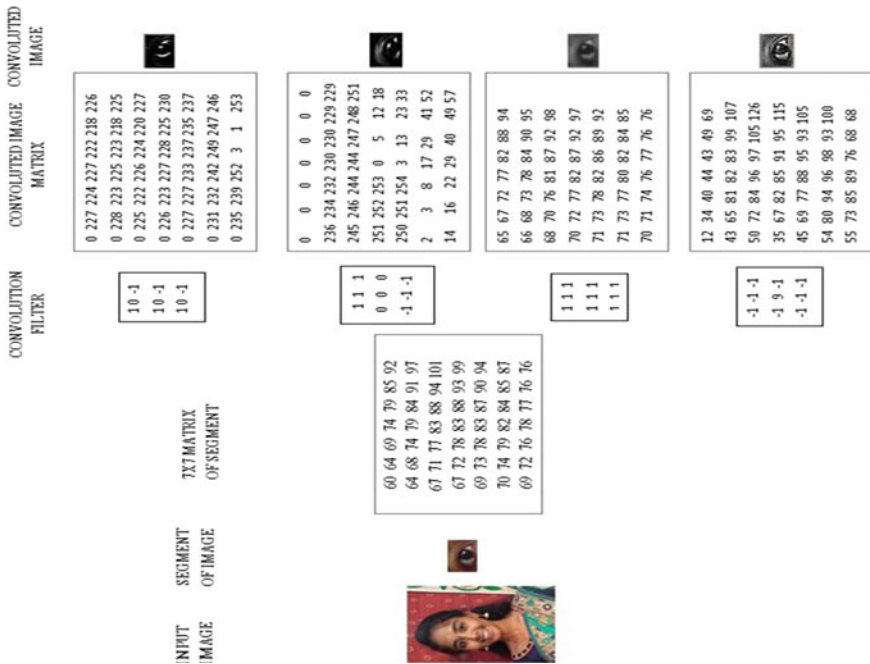


Fig. 2 Example of convolutional layer

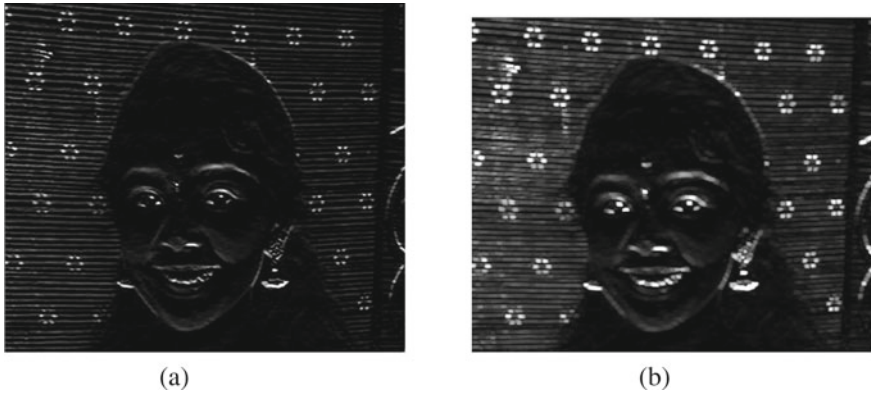


Fig. 3 CNN outputs **a** Convoluted image and **b** max-pooled image

As a next step, max pooling is done to extract only the important features of the image. To accomplish this task, the network acquires a property known as spatial variance. This property is capable to detect the object in the image of the network without being confused by the differences in the image's textures, the distances from where they are shot, their angles, etc. Figure 3 shows the convoluted image and max-pooled image. It is observable from Fig. 3 that the max-pooled image preserves most of the important features of the given input image.

Once the max-pooled image is obtained, flattening is done to convert the max-pooled output into vector. This vector is fed as input to the neural network. It is well known that neural networks are built of perceptrons. In a neural network, the first layer of perceptron makes decisions by weighing the input evidence. The face features that are extracted from the convolutional layer of CNN are fed as input to first layer of perceptron. Each of the perceptrons in the second layer makes a decision by weighing the outputs of the first layer. For example, the eye extracted from the first layer is examined for iris, eyebrows, eye lashes, etc. In this way, the perceptron in the second layer will make a complex decision at more abstract level than the perceptrons in the first layer. And even more intricate decisions can be made by perceptrons in the third layer. Thus, a more sophisticated decision-making can be done. The weights of perceptron can be changed for better performance by backpropagation [28]. The above multiple layers form the hidden layer, which contains classes of training set. The processed input image vector is compared with all the classes, and the probability of matching is obtained. The input image is annotated by using the class with highest probability.

Figure 4 shows the formation of fully connected layer for the given input image. In Fig. 4, the input image is passed through input layer, hidden layer, and output layer of fully connected layer. The input layer forms a processed image vector of the input image. The fully connected layer contains many hidden layers, and these layers multiply the flattened vector with random weights and bias. The output layer

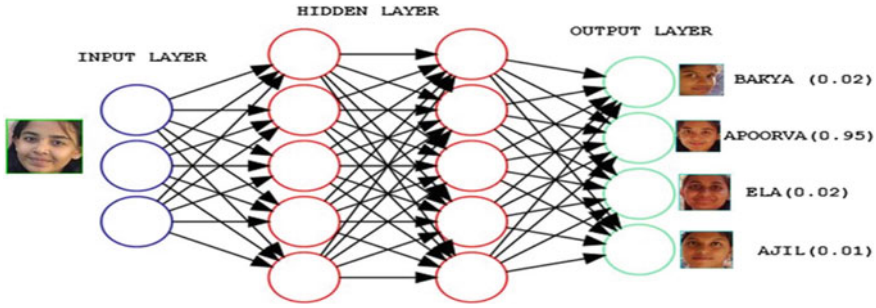


Fig. 4 Formation of fully connected layer

displays the array of face with their respective probabilities. Algorithm 1 explains the detailed process of fully connected layer.

Algorithm 1: Fully Connected Layer

Step 1: Let I be the processed image vector.

Step 2: Form hidden layer h_1 by multiplying the image vector I with weight matrix w and add bias b .

$$h_1 = I * w + b$$

Step 3: Compare h_1 with the pre-trained classes in the hidden layer.

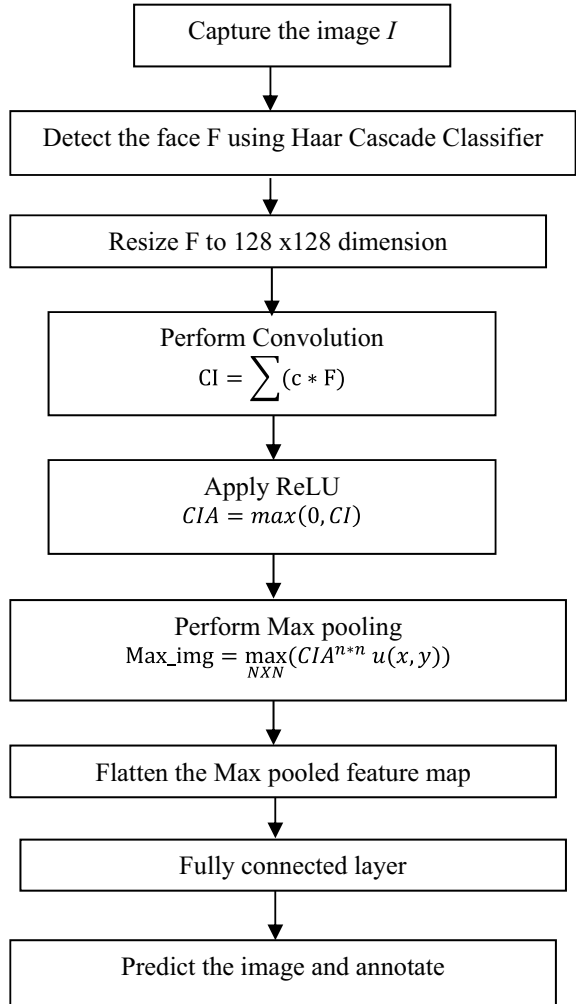
Step 4: The class with highest probability is the matched class and the image is annotated with corresponding label

Step 5: If the prediction deviates from actual output, revise the weights using back propagation and go to step 2.

3.3 Photo Albuming

Once the human facial features are obtained, the features are compared with the trained data set. If there is only one person and the person already has an album, then the image is stored in their corresponding album. If the person does not have an album, then a dialog box will appear to indicate that the user has to create a new album to store the image. If a group of human faces are detected and the image does not has an existing album, then a dialog box will appear with two options “friends” and “family,” and the user has to choose the appropriate album. Whenever an image is taken with the same group of people or some of the people in that group, the image is automatically stored in the appropriate album. Figure 5 shows the work flow of the proposed method.

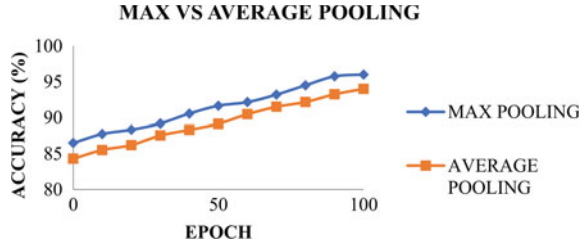
Fig. 5 Work flow of the proposed method



3.4 Results and Discussion

In general, the pooling layer in the neural network reduces the variance and computation effort of the algorithm. There are two types of pooling. One is max pooling which selects the maximum pixel value in the group, and the other one is average pooling which finds the mean of all the pixel values in the specified group. In this paper, max pooling operation is chosen as the accuracy results are better for this operation. To support this statement, Fig. 6 shows the comparison graph between max pooling and average pooling in terms of accuracy for various epochs. From Fig. 6, it is apparent that the proposed technique achieves better results for max pooling operation which

Fig. 6 Comparison between max pooling and average pooling










selects the brighter pixel values of an image and hence the sharp features of the image can be clearly distinguished.

We evaluate our technique with 100 images downloaded from Kaggle, and for each image, augmentation (shifting, rotation, resizing, cropping, etc.) has been done to create 10,000 samples for training the neural network. To show the efficiency of the proposed algorithm, accuracy results of two group of test images are given in Table 1. From , it can be verified that the proposed technique recognizes the faces correctly achieving greater accuracy values. Also, for these test images, only two albums are created in the database which further verifies the effectiveness of the algorithm. To validate the chosen classification model, two parameters precision and recall [20] are considered. Precision indicates the measure of the relevant data points, whereas recall indicates how the model correctly identifies the relevant data points. Figure 7 shows the plot between precision and recall. It is clearly visible that from Fig. 7 the algorithm achieves better classification rate with lesser false positives.

4 Conclusion

In this paper, we present the creation of photograph management system named “photobook” which groups photographs automatically into albums based on human facial features. Initially, once the image is captured, face detection technique using CNN is utilized to detect the presence of a human face or a group of human faces. Then, to enhance the face recognition accuracy, both the low-level and high-level features of the human face are extracted. The extracted features are compared with the training data set. Finally, the albums are created based on the compared facial features. The proposed algorithm provides a faster and easier browsing of photographs of a particular person or a group of people. As a practical system, user interaction is also introduced which makes it more reliable and results in good performance. In the future, the proposed work will be extended to categorize images with non-human faces and objects as well. Furthermore, it will be useful to group videos by recognizing the human faces in them.

Table 1 Accuracy scores for two test images with the application of image augmentation

Test image		Accuracy	Album
	Left shifted	0.9753	Folder 1
	Right shifted	0.9801	
	Shifted down	0.9868	
	Shifted up	0.9789	
	Mirror image	0.9413	
	Left shifted	0.9814	
	Right shifted	0.9884	

(continued)

Table 1 (continued)

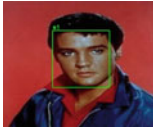


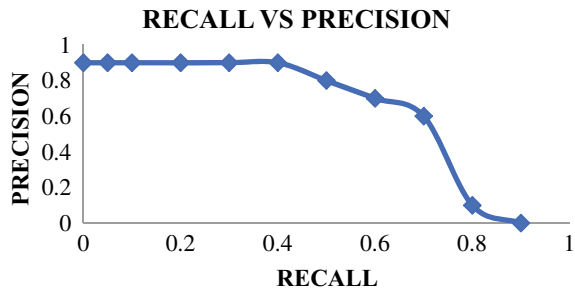
Test image		Accuracy	Album
	Shifted down	0.9868	
	Shifted up	0.9801	
	Mirror image	0.9445	

Fig. 7 Precision–recall curve



References

1. Tian Y, Liu W, Xiao R, Wen F, Tang X (2007) A face annotation framework with partial clustering and interactive labeling. In: IEEE conference on computer vision and pattern recognition, 2007. CVPR'07. IEEE, 2007, pp 1–8
2. Jheng H-W, Chen B-C, Chen Y-Y, Hsu W (2014) Automatic facial image annotation and retrieval by integrating voice label and visual appearance. In: Proceedings of the ACM international conference on multimedia. ACM, 2014, pp 1001–1004
3. Cui J, Wen F, Xiao R, Tian Y, Tang X (2007) Easyalbum: an interactive photo annotation system based on face clustering and re-ranking. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, 2007, pp 367–376
4. Chellappa R, Wilson CL, Sirohey S (1995) Human and machine recognition of faces: a survey. Proc IEEE 83(5):705–740
5. Zhao W, Chellappa R, Phillips PJ, Rosenfeld A (2002) Face recognition: a literature survey, Technical Report CAR-TR-948, Center for Automation Research, University of Maryland
6. Canedo D, António JR (2019) Facial expression recognition using computer vision: a systematic review. Appl Sci 1–31
7. Wang M, Deng W (2019) Deep face recognition: a survey. arXiv, 1–26

8. LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. *Proc IEEE* 86(11):2278–324
9. [Online]. Available: https://en.wikipedia.org/wiki/Deep_learning (accessed 2 August 2018)
10. [Online]. Available: <http://cs231n.github.io/convolutional-networks/> (accessed 2 August 2018)
11. Kennedy L, Naaman M (2008) Generating diverse and representative image search results for landmarks. In: International world wide web conference
12. Simon I, Snavely N, Seitz, SM (2007) Scene summarization for online image collections. In: International conference on computer vision
13. Platt JC (2000) AutoAlbum: clustering digital photographs using probabilistic model merging. In: Proceedings of IEEE workshop on content-based access of image and video libraries, pp 96–100
14. Platt JC, Czerwinski M, Field B (2003) PhotoTOC: automatic clustering for browsing personal photographs. In: Fourth IEEE Pacific Rim conference on multimedia
15. Sinha P, Pirsivash H, Jain R (2009) Personal photo album summarization. In: Proceedings of ACM multimedia
16. Papadopoulos S, Zigkolis C, Kaporis S, Kompatsiaris Y, Vakali A (2010) ClustTour: City exploration by use of hybrid photo clustering. In: Proceedings of ACM multimedia
17. Cooper M, Foote J, Girgensohn A, Wilcox L (2003) Temporal event clustering for digital photo collections. In: Proceedings of ACM multimedia
18. Mei T, Wang B, Hua X-S, Zhou H-Q, Li S (2006) Probabilistic multimodality fusion for event based home photo clustering. In: Proceedings of international conference on multimedia and expo
19. Suh B, Bederson BB (2004) Semi-automatic image annotation using event and torso identification. Technical report, HCIL-2004-15, Computer Science Department, University of Maryland, MD
20. Chen L, Baogang Hu, Zhang L, Li M, Zhang HongJiang (2003) Face annotation for family photo album management. *Int J Image Graph* 3(01):81–94
21. Xu Y, Peng F, Yuan Y, Wang Y Face album: towards automatic photo management based on person identity on mobile phones. In: International conference on acoustics, speech and signal processing. IEEE, 2017
22. Yan K, Huang S, Song Y, Liu W, Fan N (2017) Face recognition based on convolution neural network. Published in IEEE conference 2017. Chinese control conference
23. Aiman U, Vishwakarma VP (2017) Face recognition using modified deep learning neural network. In: 2017 8th international conference on computing, communication and networking technologies (ICCCNT), Delhi, 2017, pp 1–5
24. Coşkun M, Uçar A, Yildirim Ö, Demir Y (2017) Face recognition based on convolutional neural network. In: 2017 International conference on modern electrical and energy systems (MEES), Kremenchuk, 2017, pp 376–379
25. Google Photos—<https://www.google.com/photos/about/>
26. Apple, macOS—Photos, <http://www.apple.com/mac/iphoto/>
27. Mustafa R, Min Y, Zhu D (2014) Obscenity detection using haar-like features and gentle Adaboost classifier. *The Sci World J*, 2014:1–6. Article ID 753860
28. Alsmadi M, Omar K, Noah M, Azman S (2009) Back propagation algorithm: the best algorithm among the multi-layer perceptron algorithm. *Int J Comput Sci Netw Secur* 9:378–383