Character Photo Selection for Mobile Platform

Chuchu Xu(✉), Xinguo Yu, and Chao Sun

National Engineering Research Center for E-Learning, 
Central China Normal University, Wuhan, China
ccxu@mails.ccnu.edu.cn, xgyu@mail.ccnu.edu.cn, csun@ccnu.edu.cn

Abstract. As the smartphones are widely used in daily life, people are habituated to take photos in anywhere at anytime. The character photos, which contain both people and landscapes, are the most extensive ones in all kinds of photos. However, it is usually time-consuming and laborious for people to select and manage desired character photos manually. Therefore, automatic photo selection became extremely significant. Most of the existing methods select best character photos by assigning a absolute score or binary label, regardless of the photo’s content. The selected character photos are hence not satisfied. In this paper, we propose an effective automatic framework to select best character photos, which first automatically eliminates photos with unattractive character, then ranks photos and assists users to select the desired ones. Moreover, our framework is especially useful for mobile camera in practical application scenarios. To reduce the burden of photo selection and improve the efficiency of automatic selection from a large number of photos on mobile platform, we divide the photo selection into two stages: eliminating of unattractive photos based on designed efficient and effective features related to faces and human postures, and ranking the remaining photos using a two-stream convolution neural network. To ensure that informative features are selected and less useful ones are suppressed, we design and utilize the attention mechanism module in network. Experiments demonstrate the effectiveness of our method on the automatic selection of the satisfied photos from a large number of photos.

Keywords: Photo selection · Mobile camera · Convolution neural network

1 Introduction

To record the good life or other events, people often manually take a series of character photos in daily life. The widely use of mobile cameras enables people to easily take a series of character photos about the same object or scene from different angles in different positions [22]. However, such kind of character photos are usually too many to be easily selected and managed. When people want to select the most visually appealing character photos from the collection to share
on the social media or preserve in albums, it is usually time-consuming and laborious for people to browse the entire set of photos and then select desired ones, delete unwanted ones. Therefore, helping users to automatically select the desired character photos became extremely significant.

Automatically selecting a pleasing photo from a series of character photos is always a challenging problem, as the selection of the best photo is a subjective process. Previous researches on photo selection usually assign a absolute score (e.g., from zero to ten) or a label (e.g., high or low) to each photo by evaluating the quality of the photo based on the feature representation of the characters in the photo [10]. Such methods are independent evaluations for a series of photos, rather than comparative evaluations, which are easy to generate similar scores for similar photos, and difficult to select satisfactory character photos well. Inspired by the knowledge of praxiology, we notice that eliminating the unattractive photos is quite easier than selecting the satisfied ones, as the drawbacks are always conspicuous. Hence, we propose an automatic character photo selection method, which first automatically eliminates photos with unattractive character, then ranks photos and assists users to select the desired ones. Figure 1 displays some illustrations of photo selection results. Specifically, we divide our photo selection approach into two stages. At the first stage, we design a method to eliminate photos, which extracts efficient and effective features related to faces and human postures to gain character’s quality scores in the photo. Based on the evaluation, we then eliminate photos with low-scoring or containing unattractive characters, so as to reduce the number of photos in the subsequent selection process. At the second stage, we utilize the Siamese network architecture [5] based on Resnet50 network [13] to gain a relative ranking of remaining photos. In addition, we notice that people can quickly choose high-value information from a large amount of information when reviewing a photo. The attention mechanism [28] in deep learning is essentially similar to this human visual attention mechanism. Therefore, we introduce the attention module combining the channel attention module and spatial attention module in the Siamese network, to improve the efficiency and accuracy of photo information process.

The main contributions are summarized as follows:

1. This paper proposes an effective automatic framework for selecting best photos from a large set of character photos, which first automatically eliminates photos with unattractive character and then lists higher-ranking photos. The method can assist users to select the desired photos without manual operations.
2. This paper designs a set of efficient features related to faces and human postures for photo elimination stage. The features are proved to be effective in experiments.
3. This paper redesigns a two-stream convolution neural network based on Resnet50 network using the photo pairs as its input. In order to obtain high-value information of salient areas, we introduce the attention
mechanism combining the channel attention module and spatial attention module in our method.

The paper is organized as follows. We introduce related works on feature extraction for person and photo selection in Sect. 2. The details of method are shown in Sect. 3. The details of experiments and analysis are introduced in Sect. 4. Finally concluding remarks are given in Sect. 5.

![Fig. 1. Illustrations of photo selection results. The green boxes represent the best character photo in this series of photos. The numbers in the upper right indicate the ranking of this photo in this series of photos. Marking ‘eliminated’ in the photo means it is eliminated out during the elimination stage.](image)

2 Related Work

2.1 Feature Extraction for Person

In character photos with people as the main subject, people regions are the most attractive regions in photos. Therefore, the information involving people, such as face, body posture, is very important for selecting character photo [3,8,15]. Previous researches mainly extract face-related features to represent the photo quality and then select best photos. Zhu et al. [29] used facial expressions to determine if a portrait photo is attractive, and described a method for providing feedback on portrait to select the most attractive portraits from large video/photo collections. Li et al. [19] focused on the characteristics related the face regions, such as facial blurring, face composition, face closeness, to evaluate photo quality. Newbury et al. [22] extracted nine descriptive features about face orientation, facial emotions, face exposure and blurriness by the face detection function in Google Cloud Vision for distinguishing good and bad quality face photos. Wang et al. [27] designed a set of high-level features about human face, and combine generic aesthetic features to predict the distinction of multiple group photos of diverse human states under the same scene. These methods only
consider the factors of face quality in the photos and ignore the important information of the human postures. Actually, people’s preference are often guided by more complex factors, in addition to face states, there are also people’s posture states, such as the completeness of the human body area and the orientation of the human body [18]. Instead of selecting attractive photos based on quality of people, this paper designed both face and human posture features to evaluate the quality of people in photos, and eliminate photos with unattractive character.

2.2 Photo Selection

Photo selection has drawn the attention of both researchers and developers in recent years [12,16,21]. For example, Datta et al. [9] designed specific hand-crafted features and then made use of Support Vector Machine [24] and Decision Tree [23] to binary classification, and assign good or bad label to each photo. Tong et al. [26] adopt boosting to combine global low-level simple features (blurriness, contrast, colorfulness, and saliency) in order to classify professional photograph and ordinary snapshots. With the improvement of deep learning, convolution neural networks have shown superior performance in image understanding in recent years [11]. The RAPID model by Lu et al. [20] toke the entire photo and some parts of photo as input, and use an AlexNet-like architecture where the last fully-connected layer is set to output 2-dim probability for aesthetic binary classification. Talebi et al. [25] improved the loss function, and predicted the distribution of human opinion scores (from one to ten) using a convolution neural network. Several methods have recently been proposed to prove that selecting the best photo from a set of photos is a comparison-based process. Kong et al. [17] thought it is difficult to learn a common scoring mechanism for various photos, so they proposed to predict relative aesthetic rankings among images with similar visual content based on Siamese network. Chang et al. [7] collected the first large public dataset composed of series photos and established a relative ranking for “better” or “worse” photos from among a series of similar photos.

3 Method

3.1 Proposed Framework

The framework of the proposed automatic photo selection approach is shown in Fig. 2. A large number of series of character photos are used as input. In the photo elimination stage, the area of people in the photo is detected by extracting efficient and effective features related to human face and posture, and the quality of the characters in the photo is estimated. We then eliminate those photos with low quality of character. To gain a relative ranking of photos with high quality of character, we design a two-stream convolution neural network based on Resnet50 network which adopts the attention mechanism and combines the channel attention module and spatial attention module. In the following sections, we will elaborate the proposed method.
3.2 Elimination Stage

**Face Features Extraction.** In character photos with people as the main subject, the human face is the most attractive area [6]. The quality of the face has a dominant impact on the quality of the entire photo. Face quality can be affected by the illumination, blur, occlusion, and completeness of the face area. When people filter photos, they will not only consider the above factors, but also the states of the face in the photo, such as face orientation, closed eyes. If someone closes eyes, someone’s face is occluded or someone is not facing the camera in a photo, we will naturally exclude this photo from the candidates of satisfied character photos.

Therefore, we first detect whether there is a human face in each photo or not, then eliminate those photos without any face in. Subsequently, we extract face-related information of each face per photo from three aspects (Eyes states, Face orientation, Quality metric) respectively.

(1) **Eyes states.** If someone’s eyes are closed in a photo with people, the beauty of that will be greatly decreased. So we consider the left and right eye states (opening or closing) of each person in the photo. We obtain the confidence of left and right eye states of each person, marked as Cl and Cr respectively. Assuming N faces are detected in a character photo, we designed the features f1 and f2 of each face as follows:

\[
 f_1(i) = \begin{cases} 
 0 & \text{if } C_{l_i} < 0.3 \\
 1 & \text{otherwise} 
\end{cases} \quad i \in \{1, \ldots, N\},
\]

\[
 f_2(i) = \begin{cases} 
 0 & \text{if } C_{r_i} < 0.3 \\
 1 & \text{otherwise} 
\end{cases} \quad i \in \{1, 2, \ldots, N\},
\]

where \( C_{l_i}(C_{r_i}) \) indicates that the degree of opening or closing about the left (right) eye of the \( i^{th} \) person in the photo. We empirically set the threshold value...
to 0.3. When the confidence value of left(right) eye state is less than 0.3, it means the left(right) eye is closing, otherwise opening.

(2) Face orientation. If someone in a photo looks at the camera, but the head tilted, the photo is not a high-quality photo and is easily abandoned by people. Therefore, we get the roll, pitch, and yaw angle of each face as $R$, $P$, $Y$, where $R \in [-180, 180]$, $P \in [-90, 90]$, $Y \in [-90, 90]$. The recommended value range of $R$, $P$, $Y$ is from $-45$ to 45 degree. It is considered that people is facing the camera. So we can eliminate only those face which is beyond of this range. The definitions of features $f_3$, $f_4$, $f_5$ of each face are as follows:

$$f_3(i) = \begin{cases} 0 & \text{if } R_i \not\in [-45, 45] \\ 1 & \text{otherwise} \end{cases} \quad i \in \{1, 2, ..., N\} \tag{3}$$

$$f_4(i) = \begin{cases} 0 & \text{if } P_i \not\in [-45, 45] \\ 1 & \text{otherwise} \end{cases} \quad i \in \{1, 2, ..., N\} \tag{4}$$

$$f_5(i) = \begin{cases} 0 & \text{if } Y_i \not\in [-45, 45] \\ 1 & \text{otherwise} \end{cases} \quad i \in \{1, 2, ..., N\} \tag{5}$$

(3) Face Quality metrics. In character photos, the most basic requirements are that the face area in the photo should be well illuminated, complete, clear and unoccluded [1]. Through simple quality metrics, We can easily eliminate photos that do not follow the basic requirements. Hence, we get four features attributes about illumination, occlusion, completeness, blur as $I$, $O$, $C$, $B$. $I$ indicates the illumination level of the face area and its value ranges from 0 to 255. We empirically set the threshold value to 40. When the confidence value of illumination is less than 40, it means the illumination is poor. $O$ represents the degree of occlusion of each face, and the range is from 0 to 1, where 0 means no occlusion, and 1 indicates complete occlusion. The confidence value of occlusion which is greater than 0.8 means severe occlusion. $C$ represents the facial completeness in the photo, where the face area overflowing the image boundary is marked as ‘False’, otherwise ‘True’. $B$ indicates the degree of facial blur and its range is from 0 to 1, where 0 means the clearest face, 1 means the most blurry face. The threshold value is set to 0.7. The definitions of features $f_6$, $f_7$, $f_8$, $f_9$ of each face are as follows:

$$f_6(i) = \begin{cases} 0 & \text{if } I_i < 40 \\ 1 & \text{otherwise} \end{cases} \quad i \in \{1, 2, ..., N\} \tag{6}$$

$$f_7(i) = \begin{cases} 0 & \text{if } O_i > 0.8 \\ 1 & \text{otherwise} \end{cases} \quad i \in \{1, 2, ..., N\} \tag{7}$$

$$f_8(i) = \begin{cases} 0 & \text{if } C_i = 'False' \\ 1 & \text{otherwise} \end{cases} \quad i \in \{1, 2, ..., N\} \tag{8}$$

$$f_9(i) = \begin{cases} 0 & \text{if } B_i > 0.7 \\ 1 & \text{otherwise} \end{cases} \quad i \in \{1, 2, ..., N\} \tag{9}$$
Posture Features Extraction. Figure 3 shows the comparison of photos about different postures, illustrating the importance of extracting posture features. Therefore, we extract three important attributes related to the human body postures, and design features to eliminate photos with unattractive character based on these attributes information.

(1) Normal or abnormal human body. To tell whether the human body in the photo is normal is not to judge non-human creatures such as animals. We need to consider the completeness of people in character photo. A normal human body mainly refers to a human body in the photo with more than one-half of the body exposed, and generally the waist can be seen. An abnormal human body refers to a human body that has been severely cut off in the photo, such as only two heads or only two legs. We obtain ‘is_human’ attribute of each person in each photo. The ‘is_human’ attribute is marked as BH. If the human body in the photo is a abnormal human body, we mark BH as ‘False’, otherwise ‘True’. Assuming M bodies are detected in a character photo, we design the features $f_{10}$ as follows:

$$f_{10}(i) = \begin{cases} 0 & \text{if } BH^i = 'False' \\ 1 & \text{otherwise} \end{cases}, \quad i \in \{1, 2, ..., M\}$$  \hspace{1cm} (10)

(2) Body orientation. Similar to the face orientation attribute, the human body orientation is also a significant factor that we should consider when selecting character photos. For the front-facing and back-facing postures, people usually prefer the former. We obtain body orientation attribute of each person per photo.
and mark it as BO. BO includes three possible values: ‘front’, ‘side’, ‘back’. Therefore, feature f11 is formulated as follow:

\[ f_{11}(i) = \begin{cases} 
0 & \text{if } BO^i = 'back' \\
0.5 & \text{if } BO^i = 'side' \\
1 & \text{if } BO^i = 'front' 
\end{cases} \quad i \in \{1, 2, ..., M\} \quad (11) \]

(3) Body occlusion. When a large human body area appears in the character photo, we certainly hope that this body area is not occluded. We obtain body occlusion attribute of each person as BOC. BOC includes three possibilities of no occlusion, mild occlusion, and severe occlusion. Therefore, feature f12 is formulated as follow:

\[ f_{12}(i) = \begin{cases} 
0 & \text{if } BOC^i = 'severe' \\
0.5 & \text{if } BOC^i = 'mild' \\
1 & \text{if } BOC^i = 'no' 
\end{cases} \quad i \in \{1, 2, ..., M\} \quad (12) \]

Person Quality Estimation. Based on designed features above, they can be formalization as:

\[ F_{k1} = \frac{1}{N} \sum_{i=1}^{N} f_{k1}(i) \quad k1 \in \{1, 2, ..., 9\}, \quad i \in \{1, 2, ..., N\} \quad (13) \]

\[ F_{k2} = \frac{1}{M} \sum_{i=1}^{M} f_{k2}(i) \quad k2 \in \{10, 11, 12\}, \quad i \in \{1, 2, ..., M\} \quad (14) \]

where \( F_{k1}(k1 \in 1, 2, ..., 9) \) represents the eight attributes of all faces in each photo and \( F_{k2}(k2 \in 10, 11, 12) \) represents the three attributes of all bodies in each photo. We concatenate \( F_{k1} \) and \( F_{k2} \) as the 12 features of the person quality in this photo. Then we train a Artificial Neural Network to estimate the quality of detected people per photo. There are 5 hidden fully-connected layers in the network. Every layer uses ReLU activation. The input of the network is the 12 features, and the output layer is a scalar, reflecting the quality of person. Through the quality, we can first eliminate photos with unattractive character.

3.3 Selection Stage

In the previous elimination stage, we only consider the attributes of the person area in the photo. In this section, we will select best character photo according to the high-level semantic information of the entire photo. We explore using deep learning method to achieve automatic character photo selection.
Siamese Network. Automatic selecting the best photo from a set of character photos is a comparison-based process to a large extent [18]. For example, the quality of a particular photo seem low when viewed in an isolated manner, while in a group of photos, the same photo may be the best choice compared to other photos. In contrast to the previous methods that directly assign an absolute score for each photo, we consider to use Siamese network architecture to learn the difference of photo pairs and gain a relative ranking. In Siamese network architecture, two inputs are sent into two identical sub-networks independently, which share the same network parameters in both training and prediction phases. Inspired by [7], we utilized a Siamese network architecture which contains two identical streams of fine-tuned ResNet50 network with shared weight for extracting features. Photo pairs are taken as input.

Attention Module. We have noticed, when people observe a photo, they often firstly notice some important areas of the photo. These areas contain the information that most arouses the user’s interest and best expresses the abundant content of the photo. Humans can use limited attention to quickly select high-value information from a large amount of information. The attention mechanism in deep learning is essentially similar to the human visual attention mechanism. Inspired by [28], we introduce attention mechanism to increase feature representation in Siamese network, which can effectively select important information and suppress less valuable information. The extracted features from the last layer of ResNet50 are respectively fed into a channel attention module and a spatial channel attention module.

1) Channel attention module. The channel attention module is mainly to explore the relationship of feature maps between different channels. The channel attention module simultaneously uses average-pooling and max-pooling functions along the spatial dimension to calculate the context information of each channel, obtaining two channel attention vectors $C_{avg}$ and $C_{max}$ respectively. Then two vectors are fed into a two-layer fully connected network. Finally, the output obtained is the channel attention feature vector, denoted as $F_c$. In brief, the channel attention module is formulated as:

$$F_c = \sigma(W_{fc2}\delta(W_{fc1}(C_{avg} + C_{max})))$$

where $\delta$ denotes the ReLU activation and $\sigma$ denotes the sigmoid function. $W_{fc1}, W_{fc2}$ are the weight parameters of fully connected network.

2) Spatial attention module. The spatial attention module focuses on the spatial relationship between different feature maps, to make the network model focus on the feature spatial position of the feature maps. Similar to the channel attention module, the spatial attention module first uses average-pooling and max-pooling functions along the channel dimension to aggregate the context of each spatial position, yielding two spatial attention maps $S_{avg}$ and $S_{max}$ respectively. Then, two maps are concatenated and input into a 7*7 convolutional layer. Finally, the
output obtained is the spatial attention feature map, denoted as $F_s$. In brief, the spatial attention module is formulated as:

$$F_s = \sigma(Conv[S_{avg}, S_{max}])$$  \hspace{1cm} (16)

The features processing combining channel attention module and spatial attention module is shown in Fig. 4. We concatenate the features of the outputs from the two-stream networks respectively. Then, we calculate their distance and pass the distance to two hidden layers to classify the features of the photo pairs, each of which consists of a linear fully connected layer and a tanh activation layer. In addition, we use cross-entropy loss as the cost function. The final output of the network indicates which of the two input photos is better.

![Diagram of the features processing combining channel attention module and spatial attention module.](image)

**Fig. 4.** Diagram of the features processing combining channel attention module and spatial attention module.

## 4 Experiments and Analysis

### 4.1 Dataset

The data used in this paper is from the photo series collected in [7] for series photo selection. The dataset includes 15,545 photos (5,953 series). The pairwise preference scores on two photos have been manually labeled. If the preference score in a photo pairs is greater than 0.5, it means that the first photo is better than the second one. Out of all 15,143 photo pairs, we sampled 4639 pairs photos with people for training, 585 pairs for validation, and the remaining 500 pairs for testing. In addition, in order to explain the stability and accuracy of our model better, we divided the sample dataset into single-person photos (2759 pairs for training, 361 pairs for validation and 308 pairs for testing) and group photos (1880 pairs for training, 224 pairs for validation and 192 pairs for testing) to evaluate our model. Figure 5 shows some illustrations of the dataset.
4.2 Features Importance Ranking

In elimination stage, we utilize the face detection method in BaiduAI [2], a recognition tool, to detect the presence of faces in each photo, and then combine the face attribute analysis method and human body attribute analysis method from BaiduAI to extract twelve person-related attributes in each photo. We empirically set different threshold values for these attributes and design efficient and effective features related to person.

We used the Random Forest [4] to rank the importance of features. The importance ranking of the top twelve features is shown in Fig. 6. It is observed that the importances of pitch feature and orientation of human body feature are much higher than other features, which indicates that the orientation of face and human body is the dominant feature when eliminating photos with unattractive characters in. The importance of roll, yaw, left and right eyes states, normal human body and facial illumination is also exceeded the average value, which means that these features also play an important role in the photo elimination stage. The reason for the lower ranking of facial blur is that the facial blur features in character photos are related to the resolution of the photo. If the photo is of low resolution, the face is also easily treated as blurred. Based on the above analysis of the features importance, we decide to choose these twelve human-related features in elimination stage.

4.3 Experimental Comparison and Analysis

Accuracy Comparison with Different Methods. We experimentally verified that our method enables achieve better character photo selection. We compared our method with baseline method proposed in [7] for character photo selection. ResNet50 [14] were also introduced as baselines. The performance of different methods were evaluated in terms of classification accuracy on photo
Fig. 6. The importance ranking of the top twelve features pairs. Meanwhile, we perform experiments on single-person dataset and group dataset. Table 1 summarizes the comparison results. From the comparative analysis of Table 1, it is observed that our approach has a higher accuracy rate than the method presented in [7] and [14] on three datasets. In addition, our method with attention mechanism, is significantly better than [7] and [14] without attention mechanism, revealing the importance of applying the attention mechanism to our method.

Table 1. Accuracy comparison results of different methods on three datasets.

| Method        | Accuracy (sample dataset) | Accuracy (single-person) | Accuracy (group) |
|---------------|---------------------------|--------------------------|------------------|
| [7] (Alexnet) | 61.61                     | 62.94                    | 59.71            |
| [14] (Resnet50)| 62.56                     | 64.03                    | 59.44            |
| Ours          | **66.03**                 | **66.67**                | **62.24**        |

Accuracy Comparison with Different Attention Modules. We further perform experiments to compare three different ways of arranging attention modules: Siamese-Channel attention, Siamese-Spatial attention, and Siamese-Channel & Spatial attention (our method). The comparison results are shown in Table 2. On the sample dataset, it can be further observed that the accuracy rate of our method is higher than that of Siamese-channel attention and Siamese-spatial attention. This indicates that the performance of our method is better than other methods with single attention module, which verifies the necessity of effective combination of spatial attention module and channel attention module.
Table 2. Accuracy comparison of using different attention modules on three datasets.

| Method                     | Accuracy (sample dataset) | Accuracy (single-person) | Accuracy (group) |
|----------------------------|----------------------------|----------------------------|------------------|
| Siamese-Channel attention  | 61.16                      | 64.91                      | 58.46            |
| Siamese-Spatial attention  | 63.84                      | 65.79                      | 60.13            |
| Ours                       | **66.03**                  | **66.67**                  | **62.24**        |

Stability Comparison with Different Models. We calculated F1-score to compare the stability of our model and models proposed in [7,14] on the sample data set. F1-score measures comprehensive performance as a harmonic mean value of precision and recall. Figure 7 shows the comparison of the F1-score curves of our method and other baselines on the validation set during the training process. It can be intuitively observed that our method leads to an relative improvement in the performance of the training process. The experimental results further verify the stability and effectiveness of our model.

![F1-score Curve](image)

Fig. 7. F1-score curves on the validation set during training.

5 Conclusion

We have presented a novel method of automatic photo selection in this paper, which is able to automatically list higher-ranking photos and help users to select
the desired ones. Our method first eliminates photos with unattractive character in elimination stage. Then it takes the remaining photos as input and utilizes the Siamese network architecture based on the Resnet50 network to gain the relative ranking of these photos. To improve the efficiency and accuracy of information processing, we have introduced attention mechanism in Siamese network. Experiments show that our method lead to promising results. Moreover, our method is especially useful for the mobile platform in practical application scenarios. In the future, we will further explore more discriminative features to capture the subtle differences between character photos.

Acknowledgement. We thank the anonymous reviewers for helpful suggestions. This work is supported by National Natural Science Foundation of China under Grant 61802142.

References

1. Abaza, A., Harrison, M.A., Bourlai, T., Ross, A.: Design and evaluation of photometric image quality measures for effective face recognition. IET Biom. 3(4), 314–324 (2014)
2. BaiduAI: BaiduAI open platform. https://ai.baidu.com. Accessed 17 Oct 2020
3. Best-Rowden, L., Jain, A.K.: Learning face image quality from human assessments. IEEE Trans. Inf. Forensics Secur. 13(12), 3064–3077 (2018)
4. Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J.: Classification and regression trees (cart). Biometrics 40(3), 358 (1984)
5. Bromley, J., et al.: Signature verification using a “Siamese” time delay neural network. Int. J. Pattern Recogn. Artif. Intell. 7(04), 669–688 (1993)
6. Cavalcanti, C.S.V.C., Gomes, H.M., De Queiroz, J.E.R.: A survey on automatic techniques for enhancement and analysis of digital photography. J. Braz. Comput. Soc. 19(3), 341–359 (2013). https://doi.org/10.1007/s13173-013-0102-1
7. Chang, H., Yu, F., Wang, J., Ashley, D., Finkelstein, A.: Automatic triage for a photo series. ACM Trans. Graph. (TOG) 35(4), 1–10 (2016)
8. Choi, J., Lee, B.J., Zhang, B.T.: Human body orientation estimation using convolutional neural network. arXiv preprint arXiv:1609.01984 (2016)
9. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Studying aesthetics in photographic images using a computational approach. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3953, pp. 288–301. Springer, Heidelberg (2006). https://doi.org/10.1007/11744078_23
10. Deng, Y., Loy, C.C., Tang, X.: Image aesthetic assessment: an experimental survey. IEEE Signal Process. Mag. 34(4), 80–106 (2017)
11. Dong, Z., Shen, X., Li, H., Tian, X.: Photo quality assessment with DCNN that understands image well. In: He, X., Luo, S., Tao, D., Xu, C., Yang, J., Hasan, M.A. (eds.) MMM 2015. LNCS, vol. 8936, pp. 524–535. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-14442-9_57
12. Dong, Z., Tian, X.: Multi-level photo quality assessment with multi-view features. Neurocomputing 168, 308–319 (2015)
13. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
14. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision & Pattern Recognition (2016)
15. Hernandez-Ortega, J., Galbally, J., Fierrez, J., Haraksim, R., Beslay, L.: FaceQnet: quality assessment for face recognition based on deep learning. In: 2019 International Conference on Biometrics (ICB), pp. 1–8. IEEE (2019)
16. Hosu, V., Goldlucce, B., Saupe, D.: Effective aesthetics prediction with multi-level spatially pooled features. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9375–9383 (2019)
17. Kong, S., Shen, X., Lin, Z., Mech, R., Fowlkes, C.: Photo aesthetics ranking network with attributes and content adaptation. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9905, pp. 662–679. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46448-0_40
18. Kuzovkin, D., Pouli, T., Cozot, R., Le Meur, O., Kervec, J., Bouatouch, K.: Image selection in photo albums. In: Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval, pp. 397–404 (2018)
19. Li, C., Gallagher, A., Loui, A.C., Chen, T.: Aesthetic quality assessment of consumer photos with faces. In: 2010 IEEE International Conference on Image Processing, pp. 3221–3224. IEEE (2010)
20. Lu, X., Lin, Z., Jin, H., Yang, J., Wang, J.Z.: Rapid: rating pictorial aesthetics using deep learning. In: Proceedings of the 22nd ACM International Conference on Multimedia, pp. 457–466 (2014)
21. Mohan, J., Philumon, J.: Automatic selection of better image from a series of same scene. Int. J. Adv. Res. Comput. Sci. 8(5) (2017)
22. Newbury, R., Cosgun, A., Koseoglu, M., Drummond, T.: Learning to take good pictures of people with a robot photographer. arXiv preprint arXiv:1904.05688 (2019)
23. Quinlan, J.R.: Induction on decision tree. Mach. Learn. 1 (1986)
24. Sain, S.R.: The nature of statistical learning theory. Technometrics 38(4), 409 (1997)
25. Talebi, H., Milanfar, P.: NIMA: neural image assessment. IEEE Trans. Image Process. 27(8), 3998–4011 (2018)
26. Tong, H., Li, M., Zhang, H.-J., He, J., Zhang, C.: Classification of digital photos taken by photographers or home users. In: Aizawa, K., Nakamura, Y., Satoh, S. (eds.) PCM 2004. LNCS, vol. 3331, pp. 198–205. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-30541-5_25
27. Wang, Y., Ke, Y., Wang, K., Zhang, C., Qin, F.: Aesthetic quality assessment for group photograph. arXiv preprint arXiv:2002.01096 (2020)
28. Woo, S., Park, J., Lee, J.Y., So Kweon, I.: CBAM: convolutional block attention module. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 3–19 (2018)
29. Zhu, J.Y., Agarwala, A., Efros, A.A., Shechtman, E., Wang, J.: Mirror mirror: crowdsourcing better portraits. ACM Trans. Graph. (TOG) 33(6), 1–12 (2014)