Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Human-computer interaction based on face feature localization ☆

Yan Shi [a], Zijun Zhang [a,*], Kaining Huang [a], Wudi Ma [a], Shanshan Tu [b]

[a] BengBu University, Bengbu City, Anhui Province 233000, China
[b] Faculty of Information Technology, Beijing University of Technology, 100124 Beijing, China

ARTICLE INFO

Article history:
Received 2 November 2019
Revised 10 December 2019
Accepted 10 December 2019
Available online 16 December 2019

Keywords:
Human interaction
Face point detection
Expression recognition
Sight tracking

ABSTRACT

Human-computer interaction is the way in which humans and machines communicate information. With the rapid development of deep learning technology, the technology of human-computer interaction has also made a corresponding breakthrough. In the past, the way human-computer interaction was mostly relied on hardware devices. Through the coordinated work of multiple sensors, people and machines can realize information interaction. However, as theoretical technology continues to mature, algorithms for human-computer interaction are also being enriched. The popularity of convolutional neural networks has made image processing problems easier to solve. Therefore, real-time human-computer interaction can be performed by using image processing, and intelligent of human-computer interaction can be realized. The main idea of this paper is to use the real-time capture of face images and video information to image the face image information. We perform feature point positioning based on the feature points of the face image. We perform expression recognition based on the feature points that are located. At the same time, we perform ray tracing for the identified human eye area. The feature points of the face and the corresponding expressions and implementation movements represent the user's use appeal. Therefore, we can analyze the user's use appeal by locating the face feature area. We define the corresponding action information for specific user face features. We extract the user's corresponding information according to the user's face features, and perform human-computer interaction according to the user's information.

1. Introduction

Human-computer interaction is the process of information exchange between people and systems. There are many different types of systems, ranging from various machines to computer systems and software. With the development of related technologies in ubiquitous computing, the way of human-computer interaction is also constantly enriched. The earliest human-computer interaction method is realized by manual operation of input machine language instructions. The medium of interaction is computer language. It is worth mentioning that the earliest command language for interaction is machine language. With the development of the graphical interface, the medium of human-computer interaction has also been transformed into a graphical interface. The graphical interface has a better interactive experience, making the user's interaction more convenient. Computer feedback can be made more clearer to the user. With the further development of information communication technology, the way of human-computer interaction has also been greatly improved. The variety of human-computer interaction has become more abundant. The medium used for interaction has also become more diverse. The main ways of human-computer interaction are speech recognition, gesture recognition, tracking, and so on [1–3]. The new form of human-computer interaction technology makes the interaction process simpler, faster, and easier to understand. In this new form of interaction, the amount of information that can be passed is greatly increased.

Nowadays, with the maturity of face analysis technology, face features are used for various applications. In this paper, the face feature area is located, and the interaction process is directly performed through the face. Face detection and recognition is one of the important applications in the field of deep learning. Face recognition refers to the positioning of facial key points on a face after detecting a face. After locating the face feature area, we preprocess the data and use the recognition algorithm to extract features and complete the task of face recognition [4], as shown in Fig. 1. As an important field of pattern recognition, face recognition has always been one of the research hotspots. With the rapid development of

☆ This paper has been recommended for acceptance by Maofu Liu.
* Corresponding author.
E-mail address: bbxyzzj@163.com (Z. Zhang).

https://doi.org/10.1016/j.jvcir.2019.102740
1047-3203/© 2020 Elsevier Inc. All rights reserved.

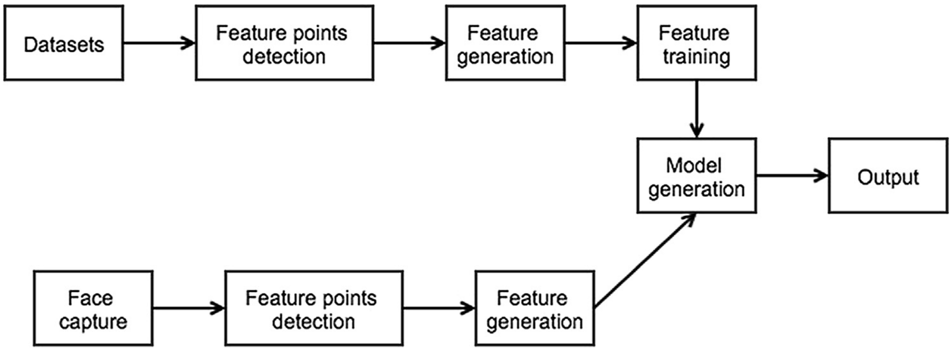

Fig. 1. Human-computer interaction process based on face detection.

deep learning, the technology of face recognition has become more mature. The image is collected by the relevant device, and the content in the image is analyzed and analyzed. The designed algorithm is used to confirm the position of the facial features in the captured video image, and the angle and posture of the face are analyzed accordingly [21–23,34]. By locating the feature area of the face, we identify and analyze the facial expression. Expression recognition is an important direction for computers to understand human emotions and an important aspect of human-computer interaction. By analyzing expressions, you can capture user information and make decisions. Analyze emotional commonly used convolutional neural networks (CNN). By analyzing multiple convolution and collection layers in the CNN, higher and multi-level features of the entire face or local region can be extracted and have good classification performance of facial expression image features. Experience has shown that CNN is superior to other types of neural networks in image recognition. Through the role of CNN, our work can achieve better effect expression recognition. Through facial expression analysis, you can judge the emotions and psychology of the immediate user [7,8,14].

Another important aspect in the process of human-computer interaction is sight tracking. Through the line of sight tracking, the user's focus can be easily observed, which is more conducive to analyzing the user's area of interest and analyzing the user's choices and preferences. We use the human eye as a source of input for computers. By tracking the user's sight, we determine the sight range of the human eye and complete the corresponding human-computer interaction. In this work, we estimate the user's position on the screen by performing human eye positioning on the captured image. In this way, the user can complete the human-computer interaction work without wearing any hardware equipment.

The work of this paper is mainly aimed at human-computer interaction. Human-computer interaction is a very complex concept. This article mainly interacts with face analysis. In this paper, by capturing the face and analyzing the face area, the facial features of the face are located. After locating the relevant features, we send the data into the convolutional neural network for feature analysis to resolve the facial features of the captured faces. At the same time, for the eye area of the face, we further carry out the tracking of the eye, and always judge the scope of the human eye. By combining the overall facial features, expression features, and sight tracking, we analyze a three-dimensional user usage information and bind behavior information according to the user's specific facial actions to achieve human-computer interaction [15,17,24,25].

The main contributions of this article are:

- Realize human-computer interaction by identifying facial features.
- The facial features related to the user information are divided into an overall facial feature, an expression feature, and sight tracking.
- Binding specific actions in the facial features of the user into behavior information, and realizing human-computer interaction through facial feature information.

The remainder of this paper is organized as follows. In Section 2, we introduce related work on facial feature point localization, facial expression recognition and related methods of sight tracking. The implementation of the proposed method is described in detail in Section 3. In Section 4, we introduced the model training method and the corresponding datasets. Finally, we conclude this paper in Section 5.

2. Related work

In the work of this paper, we interpret the face feature area for human-computer interaction. In this paper, the way to parse the face feature area mainly includes positioning the face feature point, performing face expression analysis, and face line of sight tracking. These three aspects of study work are important research areas in computer vision. In these three aspects of work, relevant researchers have made research progress. Next, we will introduce related work in turn for these three aspects.

In the work of face feature point positioning, face key point detection is also called face key point positioning or face alignment, which refers to a given face image, and locates the key area of the face, including eyebrows, Eyes, nose, mouth, facial contours, etc. The methods of face key point detection are mainly divided into three types: model-based ASM (Active Shape Model) and AAM (Active Appearnce Model), based on Cascaded pose regression (CPR) and based on the methods of deep learning. This article mainly uses the method of deep learning. The method proposed by Sun et al. proposes a new method for estimating facial key points through a 3-level convolutional neural network [5]. At each level, the output of the network is robust and accurate. The deep structure of the convolutional network can extract advanced features from all face regions in the initial stage, which facilitates the accurate positioning of key points [6]. There are two main advantages to this approach: 1. The context information for the entire face is used. 2. The geometric constraints of the key points have been implied. The disadvantages of the local optimization method are avoided. The last two stages of the network are trained to locally optimize the initial prediction values. Their research presents a method for cascading regressions. A three-level convolutional network is used to detect facial key points. Unlike the existing method, which roughly estimates the initial position of the face key points, the convolutional network makes an accurate estimate at the first level. This effectively avoids the local minimum problem. The convolutional network takes the entire face as input, best utilizes the context information, and extracts global high-level features at the top level of the deep framework, which can effectively predict the key position even if the local low-level features become unreliable. At the same time, because multiple points are predicted at the same time, the constraints of key points are also implied. The method proposed by Zhou et al. is used to locate the documents of multiple facial feature points, and achieves high-precision positioning of 68 facial feature points. Zhou's work is mainly based on Sun's work. The method of locating the feature points of the face 68 is also based on the idea of positioning from coarse to fine, and the network belongs to DCNN. In the input aspect of the network, instead of using the face area image detected by the face detector as the input of the network, the CNN is used to predict the bounding box of the face. This improvement improves the initial level positioning accuracy much [11].

In the field of facial expression recognition, many scholars now use the method of convolutional neural network in deep learning to achieve. Kuo et al. proposed an innovative model in their work. Their article balances a simplified FER model between accuracy and model size, providing a cost-effective model reference solution for embedded devices. At the same time, the method validates the proposed method on two standard data sets is better than the current best method. Their research collected data sets from three different scenarios to verify the performance of the model in multiple scenarios. Kuo et al. proposed a lighting enhancement strategy that mitigates the over-fitting problem of training on data combined with different data sets [27]. The traditional FER method uses manual features such as LBP, BoW, HoG, and SIFT, and has achieved good results on some data sets. The sequence-based approach models the expression changes by features that are manually extracted from the video. Because of the variety of lighting and posture in a real environment, this presents a challenge for traditional methods. Their article uses a more appropriate CNN architecture to solve this problem [13]. Zhang et al. proposed an end-to-end deep learning model that combines different poses and expressions to achieve parallel face image synthesis and position-invariant facial expression recognition. Their work explicitly separates identity representation learning from expression and posture changes through expression and pose encoding, enabling the model to automatically generate face images of arbitrary expressions in any pose. The model implements the most advanced facial expression recognition performance on the Multi-PIE [18], BU-3DFE [10] and SFEW [12] data sets [16].

In the related work of line-of-sight tracking, the simple principle is that the LED emits infrared light, captures the reflected light through the built-in camera, and then undergoes image processing to know the angle and direction of the user's eye movement, so that it can be controlled by the eyeball, such as a computer. The goal of Timm et al.'s project proposes an accurate and robust method of center positioning of the eye by using image gradients. Their method yields a simple objective function that contains only dot products. The maximum value of this function corresponds to the position where most gradient vectors intersect, that is, corresponding to the center of the eye. Their methods are invariant to changes in proportion, posture, contrast and illumination. Their method was evaluated in the BioID database for eye center and iris localization, demonstrating a significant improvement in both accuracy and robustness [19,20].

This paper will combine the three aspects of face key location, face expression recognition and line-of-sight tracking, and will jointly derive user face features for human-computer interaction. The specific work is introduced in Section 3.

3. Our proposed approach

In the work of this paper, we focus on the human-computer interaction problem in the direction of face feature location. In this paper, by combining face key point location, face expression recognition, and sight tracking to construct a combination feature, we bind specific combination features into behavior information, and then complete the human-computer interaction work according to the combination features.

3.1. Framework

The framework of this paper is shown in Fig. 2. The work of key point location is mainly carried out through a three-layer convolutional neural network. Facial expression recognition mainly adopts an end-to-end deep learning model, which uses different gestures and expressions to realize face image synthesis and facial expression recognition. The work of sight tracking is mainly through the use of image gradients for eye center positioning. After obtaining these three characteristics, we combine the three features into the three-layer neural network for training, in order to make the machine respond reasonably according to the combination characteristics.

3.2. Face point detection

This paper takes the structure of a three-layer convolutional neural network. The primary network is a deep convolutional network with four convolution levels, absolute value correction and local shared weights. The networks of the second and third layers share a common shallow structure. Since they are designed to extract local features, deep structures and local sharing weights are not required. We take multiple levels of regression to combine the three convolutional neural networks. Due to the large posture change and the instability of the face detector, the relative position of the face point and the bounding box may vary over a wide range. The input area of the first level network should be large to cover many possible predictions. The output of the first level network provides a strong a priori condition for subsequent detection, so the second level of detection can be done in a smaller area and the process repeated. We predict the location of each point together with multiple networks at each level. The input areas of these networks are different. The final predicted position of the face can be officially expressed as

$$x = \frac{x_1^{(1)} + \cdots + x_{l_1}^{(1)}}{l_1} + \sum_{i=2}^{n} \frac{\Delta x_1^{(i)} + \cdots + \Delta x_{l_i}^{(i)}}{l_i}$$

for an n-level cascade with l_i predictions at level i.

We need to consider the design of a three-layer convolutional neural network. On the first level, we use three deep convolutional networks, F1, EN1 and NM1, whose input areas cover the entire face (F1), eyes and nose (EN1), nose and mouth (NM1). Each network simultaneously predicts multiple face points. For each facial point, the predictions for multiple networks are averaged to reduce the variance. F1 contains four convolutional layers and the largest merged pool, as well as two fully connected layers. EN1 and NM1 have the same deep structure, but each layer is different in size because of their different input area sizes. The second and third level networks will take as input the local patch centered on the predicted position of the previous level of face points and only allow small changes to the previous prediction. The size of the patch and the search range continue to decrease along the cascading. The last two levels of prediction are strictly limited because the local appearance is sometimes ambiguous and unreliable. The

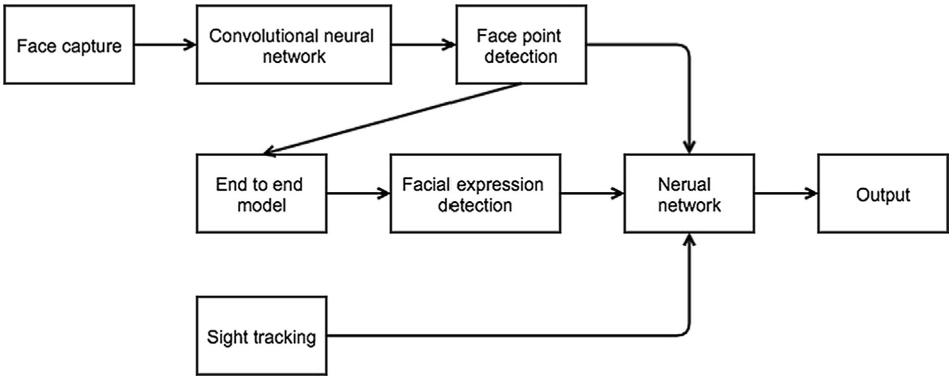

Fig. 2. The framework of our model.

predicted position of each point on the last two levels is given by the average of two networks with different patch sizes. Although the first layer of the network is designed to robustly estimate the location of key points with little or no error, the last two layers of the network are designed to achieve high accuracy. All of the last two levels of the network share a shallow public structure because their tasks are low-level.

3.3. Facial expression recognition

We propose an end-to-end learning model that simultaneously performs facial image synthesis and gesture-invariant facial expression recognition by using different gestures and expressions together. The architecture of our model contains a generator, two discriminators and a classifier. Before passing the image into the model, we first use the lib face detection algorithm with 68 landmarks for face detection [26]. After pre-processing, we input the facial image into the encoder-decoder structured generator *G* to learn the identity representation. Specifically, Genc learns the mapping from the input image to the identity representation $f(x)$. The representation is then concatenated with the expression and the gesture codes e and p to feed to Gdec for facial changes. Through the minimax two-player game between generator *G* and discriminator *D*, we can obtain new tag face images with different poses and emoticons by adding the corresponding tags to the input of the decoder. Here we use two discriminator structures including *Datt* and *Di*. *Datt* is used to learn the entanglement representation, and another *Di* is used to improve the quality of the generated image. After the facial image is synthesized, then the classifier *Cexp* is used to perform our FER task. We use a deep modeling approach to classifiers that ensures that at each level these functions become more and more constant for annoying factors while retaining discriminative information about facial expression recognition tasks.

3.4. Sight tracking

We analyzed the vector field of the image gradient. We mathematically describe the relationship between the possible center and the direction of all image gradients. Set a possible center and provide a gradient vector at position $\boldsymbol{x}_i$. Then, the normalized displacement vector should have the same direction as the gradient $\boldsymbol{g}_i$. If we use the vector field of the image gradient, we can calculate the dot product between the normalized displacement vector related to the fixed center and the gradient vector $\boldsymbol{g}_i$. The dot product is used to take advantage of this vector field. The best center $\boldsymbol{c}^*$ of the circular object in the image of pixels $\boldsymbol{x}_i, i \in \{1, \cdots, N\}$ is given by:

$$\boldsymbol{c}^* = \arg\max\left\{\frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{d}_i^T\boldsymbol{g}_i\right)^2\right\}$$

$$\boldsymbol{d}_i = \frac{\boldsymbol{x}_i - c}{\|\boldsymbol{x}_i - c\|}$$

The displacement vector is scaled to unit length to get equal weight for all pixel locations. To improve robustness to linear variations in light and contrast, gradient vectors should also be scaled to unit length. The objective function produces a maximum at the center of the pupil. By considering only the gradient vector with a significant amplitude, the computational complexity can be reduced. In order to obtain the image gradient, we calculate the partial derivative $\boldsymbol{g}_i$.

4. Experimental results and analysis

In the experimental part of the work of this paper, we conducted three parts of experiments based on three characteristics. For the feature point localization experiments of human faces, we mainly carried out experiments LFPW [8]. For facial expression recognition, we performed data experiments on the Multi-PIE [18], BU-3DFE [10], and SFEW [12] data sets. For eye tracking, we mainly do this at BioID [21]. We begin with the dataset introduction.

4.1. Dataset introduction

BioID: The dataset consists of 1521 gray-scale face images collected from 23 subjects, each of which is a 384 × 286 resolution image. All images were labeled with the position of eyes.

LFPW: The dataset is divided into training samples and testing samples. The training samples consists of 811 images while the testing samples consists of 224 images. All images are labeled with 68 face feature points.

Multi-PIE: The dataset consists of more than 750,000 images collected from 337 subjects. The dataset contains 15 camera view points, 19 luminance variations and a large range of facial expression.

BU-3DFE: The dataset is released by Binghamton University, which is used for facial expression analysis. The dataset consists of 2500 facial expression models collected from 100 subjects, 56% of whom were female and 44% male. In addition to the natural expression, there are six other expressions (happiness, angry, fear, disgust, surprise and sadness), each of which contains four levels of intensity. Thus, each subject is collected 25 images and there are 2500 images in total.

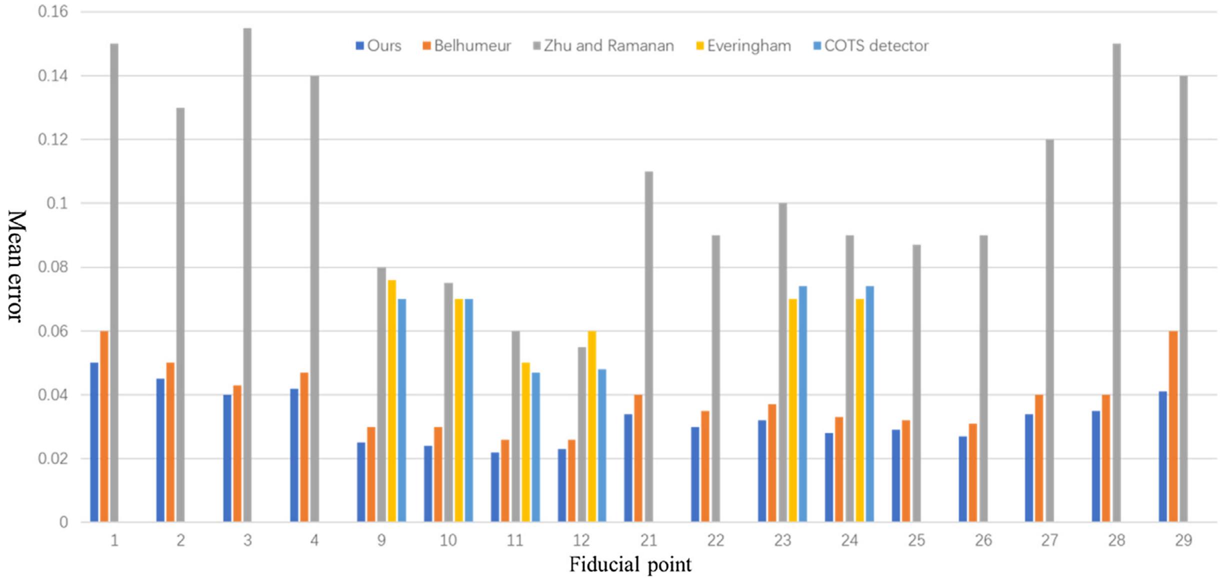

Fig. 3. The comparison performance tested on LFPW dataset.

SFEW: The dataset is collected from movies; thus, it is very close to real-world environment. There are seven expressions including a natural expression and six prototypic expressions.

4.2. Comparative study

4.2.1. Face feature localization evaluation

For evaluating performance on LFPW dataset, we compare our method with several well-known algorithms including algorithms proposed by Zhu and Ramanan [28], Everingham et al. [29], Belhumeur et al. [9], and COTS detector. In our experiment, we evaluate the performance of human face feature point localization algorithms based on the distance between the localization generated by each algorithm and MTurk workers [30]. The comparison result is shown in Fig. 3. The experimental result shows that out proposed method achieves the least mean error compared with other four methods. Thus, our method achieves the best performance.

4.2.2. Facial expression recognition

We conduct facial expression recognition algorithms on three datasets: Multi-PIE, BU-3DFE, and SFEW. We compare our algorithm with a Bayesian classifier proposed by Guo et al. [32], SVM with kernel erbf, SVM with kernel rbf, N-Nearest with $N = 9$, and N-Nearest with $N = 5$. The comparison results tested on the three datasets are shown in Table 1, where the boldfaces denote the best performance. As we can see that our method achieves the best performance among six algorithms. Furthermore, we evaluate the performance of our method on SFEW dataset based on the seven expressions. The recognition accuracy is shown in Table 2. The experimental results show that our proposed method achieves satisfactory performance.

4.2.3. Eye tracking evaluation

In this subsection, we evaluate the performance of our sight tracking algorithms. The experiment is conducted on BioID dataset. We compare our method with several well-known algorithms including eye tracker proposed by Krafka [31], gaze estimation proposed by Baltrusaitis [32], visual-context boosting method for eye tracking [33]. In our implementation, we leverage a hardware named EyeLinkII to capture human eyes in real-time, which can be used for ground-truth. Table 3 reports the comparison result tested on different algorithms. Our proposed method achieves the best performance in tracking human sight.

Table 1
The accuracy of face expression recognition tested on different dataset and algorithms.

Algorithms	Multi-PIE	BU-3DFE	SFEW	Average
Guo et al.	0.6531	0.7304	0.8137	0.7324
SVM with kernel erbf	0.6329	0.6967	0.7694	0.6997
SVM with kernel rbf	0.6032	0.6875	0.7590	0.6832
N-Nearest with $N = 9$	0.6219	0.6791	0.7306	0.6772
N-Nearest with $N = 5$	0.5926	0.6128	0.6783	0.6279
Ours	**0.7342**	**0.8547**	**0.9103**	**0.8331**

Table 2
The face expression recognition accuracy tested on the SFEW dataset.

Emotion	Neutral	Happy	Sad	Fear	Disgust	Angry	Surprise
Precision	0.8994	0.8860	0.9140	0.9380	0.9213	0.9272	0.8927
Recall	0.8106	0.8381	0.8635	0.8528	0.8629	0.8330	0.8435

Table 3
The accuracy of eye tracking tested on different algorithms.

Algorithms	Accuracy
Krafka et al.	0.8796
Baltrusaitis et al.	0.8327
Song et al.	0.8573
Ours	**0.9168**

5. Conclusions

In this paper, the face image is processed by real-time capture of the face image and convolutional neural network. In this paper, face feature processing is performed by face image processing technology of three-layer convolutional neural network. We input the extracted feature points into an end-to-end deep learning model for expression analysis and extraction. We analyze the image field of the image gradient and define the center of the eye where the gradient vectors intersect to perform line-of-sight tracking. We fuse the characteristics of these three aspects that are defined. For combined features, we input features into the neural network to extract useful information. We enter the information into the computer. We analyze the feedback that the computer should have based on the combined features to perform human-computer interaction. The experiment proves that our work is very competitive.

CRediT authorship contribution statement

Yan Shi: Conceptualization. **Zijun Zhang:** Writing - original draft, Writing - review & editing. **Kaining Huang:** Formal analysis. **Wudi Ma:** Resources, Software. **Shanshan Tu:** Project administration.

Declaration of Competing Interest

The authors declared that there is no conflict of interest.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (No. 61801008), National Key R&D Program of China (No. 2018YFB0803600), Scientific Research Common Program of Beijing Municipal Commission of Education (No. KM201910005025) and Beijing National Natural Science Foundation (No. L172049).

References

[1] Q. Wang, P. Lu, Research on application of artificial intelligence in computer network technology, Int. J. Patt. Recognit Artif Intell. 33 (5) (2019) 1959015.
[2] Biao Han, Xiangrui Yang, Zhigang Sun, Jinfeng Huang, Su. Jinshu, OverWatch: A cross-plane DDoS attack defense framework with collaborative intelligence in SDN, Secu. Commun. Netw. 2018 (2018).
[3] Sang-Bing Tsai, You-Zhi Xue, Po-Yu Huang, et al., Establishing a criteria system for green production, Proc. Instit. Mech. Eng. Part B: J. Eng. Manuf. 229 (8) (2014) 1395–1406.
[4] X. Zhou, X. Liang, X. Du, J. Zhao, Structure based user identification across social networks, IEEE Trans. Knowl. Data Eng. 30 (6) (2018) 1178–1191.
[5] Y. Sun, X. Wang, X. Tang, Deep convolutional network cascade for facial point detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3476–3483.
[6] J. Wang, Z. Xu, New study on neural networks: the essential order of approximation, Neural Netw. 23 (5) (2010) 618–624.
[7] Xu Mingliang, Su Hao, Yafei Li, Xi Li, Jing Liao, Jianwei Niu, Pei Lv, Bing Zhou, Stylized aesthetic QR code, IEEE Trans. Multim. 21 (8) (2019) 1960–1970.
[8] J. Li, X. Liang, Y. Wei, T. Xu, J. Feng, S. Yan, Perceptual generative adversarial networks for small object detection, CVPR, 2017.
[9] P.N. Belhumeur, D.W. Jacobs, D.J. Kriegman, N. Kumar, Localizing parts of faces using a consensus of exemplars, The 24th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011, IEEE, 2011.
[10] L. Yin, X. Wei, Y. Sun, J. Wang, M.J. Rosato, A 3d facial expression database for facial behavior research, in: FG, IEEE, 2006, pp. 211–216.
[11] E. Zhou, H. Fan, Z. Cao, et al., Extensive facial landmark localization with coarse-to-fine convolutional network cascade, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2013, pp. 386–391.
[12] A. Dhall, R. Goecke, S. Lucey, T. Gedeon, Static facial expression analysis in tough conditions: data, evaluation protocol and benchmark, in: ICCV, IEEE, 2011, pp. 2106–2112.
[13] C.M. Kuo, S.H. Lai, M. Sarkis, A compact deep learning model for robust facial expression recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 2121–2129.
[14] Xu Mingliang, Xiaozheng Xie, Pei Lv, Jianwei Niu, Hua Wang, Chaochao Li, Ruijie Zhu, Zhigang Deng, Bing Zhou, Crowd behavior simulation with emotional contagion in unexpected multi-hazard situations, IEEE Trans. Syst. Man Cybern. Syst. (2019), https://doi.org/10.1109/TSMC.2019.2899047.
[15] Y. Yang, J. Song, Z. Huang, et al., Multi-feature fusion via hierarchical regression for multimedia analysis, IEEE Trans. Multim. 15 (3) (2012) 572–581.
[16] F. Zhang, T. Zhang, Q. Mao, et al., Joint pose and expression modeling for facial expression recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3359–3368.
[17] Xu Mingliang, Pei Lv, Mingyuan Li, Hao Fang, Hongling Zhao, Bing Zhou, Yusong Lin, Liwei Zhou, Medical image denoising by parallel non-local means, Neurocomputing 195 (2016) 117–122.
[18] R. Gross, I. Matthews, J. Cohn, T. Kanade, S. Baker, Multipie, Image Vis. Comput. 28 (5) (2010) 807–813.
[19] Peng Ye et al., Unsupervised feature learning framework for no-reference image quality assessment, 2012 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2012.
[20] O. Jesorsky, K.J. Kirchberg, R. Frischholz, Robust face detection using the hausdorff distance, Proc. AVB- PA, 2001.
[21] Xu. Mingliang, Li Chunxu, Lv* Pei, Lin Nie, Rui Hou, Bing Zhou, An efficient method of crowd aggregation computation in public areas, IEEE Trans. Circ. Syst. Video Technol. 28 (10) (2018) 2814–2825.
[22] C.R. Qi, H. Su, K. Mo, L.J. Guibas, Pointnet: Deep learning on point sets for 3d classification and segmentation, arXiv preprint: 1612.00593, 2016.
[23] A. Liu, H. Liu, Y. Xiao, S.B. Tsai, H. Lu, An empirical study on design partner selection in green product collaboration design, Sustainability 10 (1) (2018) 133.
[24] Pei Lv, Zhujin Zhang, Chaochao Li, Yibo Guo, Bing Zhou, Xu Mingliang, Crowd behavior evolution with emotional contagion in political rallies, IEEE Trans. Comput. Soc. Syst. 6 (2) (2018) 377–386.
[25] X. Zhu, P. Milanfar, A no-reference sharpness metric sensitive to blur and noise, in: Quality Multim. Exp. Int. Workshop, 2009-Jul, 2009, pp. 64–69.
[26] S. Yu, J. Wu, S. Wu, D. Xu, Lib Face Detection, 2016. https://github.com/ShiqiYu/libfacedetection/.
[27] Zongsheng Huang, Jiajia Nie, Sang-Bing Tsai, Dynamic collection strategy and coordination of a remanufacturing closed-loop supply chain under uncertainty, Sustainability 9 (2017) 683.
[28] X. Zhu, D. Ramanan, Face detection, pose estimation, and landmark localization in the wild, 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2012.
[29] M. Everingham, J. Sivic, A. Zisserman, Hello! My name is... Buffy" – automatic naming of characters in TV video, BMVC, 2006.
[30] Xu Mingliang, Mingyuan Li, Xu Weiwei, Zhigang Deng, Yin Yang, Kun Zhou, Interactive mechanism modeling from multi-view images, ACM Trans. Graph. 35 (6) (2016) 236, 1–236:13.
[31] K. Krafka, A. Khosla, P. Kellnhofer, H. Kannan, S. Bhandarkar, W. Matusik, et al., Eye tracking for everyone, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, https://gazecapture.csail.mit.edu/.
[32] T. Baltrusaitis, P. Robinson, L.P. Morency, OpenFace: An open source facial behavior analysis toolkit, 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2016.
[33] M. Song, D. Tao, Z. Sun, X. Li, Visual-context boosting for eye detection, IEEE Trans. Syst. Man Cybern. Part B (Cybernetics) 40 (6) (2010).
[34] Junxiao Xue, Hui Yin, Pei Lv, Xu Mingliang, Yafei Li, Crowd queuing simulation with an improved emotional contagion model, Sci. China Inf. Sci. 62 (4) (2019), 44101:1–44101: 3.