Convolutional Neural Network Model for Facial Feature Extraction in Personality Traits Recognition

Nurrul Akma Mahamad Amin[1], Nilam Nur Amir Sjarif[2], Siti Sophiayati Yuhaniz[3]
[1,2,3]Advanced Informatics Department, Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia
[1]nurrulakma@graduate.utm.my, [2]nilamnur@utm.my, [3]sophia@utm.my

Article history

Received:
27 Oct 2023

Received in revised form:
10 Nov 2023

Accepted:
16 Nov 2023

Published online:
18 Dec 2023

*Corresponding author
nurrulakma@graduate.utm.my

Abstract

Personality computing lies on the effort of designing computational models to recognize personality traits from different sources of input such as visual (image), verbal (text) and vocal (audio). To recognize personality traits from facial images, the recognition model needs to extract facial features such as eyes, mouth, nose, etc. Facial features are relevant for personality judgements where they are providing information about human expression and behaviours. For example, by detecting eye's movement, dimension and size, machines can describe people's personality. Basically, people with higher score of conscientiousness have greater fluctuations in their pupil size while people who blink faster are more neurotic. The potential of each facial feature to automatically describe human personality has prompted research in personality recognition using CNN-based techniques. On the other hand, convolutional neural networks (CNN) have proven to be greatly success in the field of image processing including face recognition and detection. Since face detection is a key task in personality recognition using facial features, adoption of CNN model is a reliable choice. This study aimed to explore CNN-based models, which have been used and modified in the personality computing field. In this paper, we review several CNN-based models which have been employed for personality traits recognition, with an emphasis on extraction of facial features. Based on the finding, there are four models (VGGNet, ResNet, FaceNet and OpenFace) which are widely used and frequently adopted by previous studies in personality trait recognition. The adoption of CNN-based models in facial features extraction for personality traits recognition helps in achieving a good accuracy result. Thus, there is a huge potential to enhance and modify CNN-based model to generate more efficient model for facial features extraction tasks. Initially, this paper briefly explains personality computing and facial features extraction for personality recognition task. Next, this paper discusses the characteristics of network layer in CNN-based models for facial features extraction and personality recognition. Finally, this paper concludes that each of these models has its own features strength that are potentially useful for facial features extraction in development of personality traits recognition model.

Keywords: *facial feature extraction, personality computing, personality traits recognition, CNN, Big Five model*

* *nurrulakma@graduate.utm.my*

1. Introduction

Personality Computing (PC) is a thriving research topic that combines knowledge in psychology and computer science. This kind of research topic aims to extract information related to the individual's personality from machine sensory data. It focuses on the development of computational algorithms to automatically detect human personality. Availability of machine sensory information such as videos, user behaviours in social media platforms, smartphone usage, online shopping patterns, and written text in digital media spur on the development of personality recognition system. Personality recognition system make use all of that information as an input data and then classifying it into several different personality model attributes such as Big Five model which consists of five attributes dimension includes openness, conscientiousness, extraversion, agreeableness, and neuroticism [1][2]. In addition, instead of the Big Five dimensions, personality recognition system also uses other metrics such as confident, intelligent, unhappy, weird, and neutral [3].

Personality computing can also be seen as an extension or addition to affective computing where both are two related fields in human-machine interaction. The tasks in affective computing usually involved emotion or facial expressions recognition, pain detection, mental health, and depression recognition while personality computing concerns on recognizing human personality traits or person's characteristics and self-quality [4]. Emotion and personality recognition often employ face images as the primary input data for the recognition process. Prior to 2015, research efforts in affective computing focused more on sentiment polarity and emotion recognition than on personality traits recognition. This is because of the difficulty in obtaining adequate datasets coupled with the complexity of personality recognition processes itself [5]. However, in 2016, the organization of ChaLearn Looking at People Challenge 2016 (ECCV2016) became an ideal beginning for researchers to investigate the problem of personality analysis. This challenge aimed to recognize the Big Five traits in the videos of people speaking in front of the camera. The challenging aspect of personality traits recognition system development is associated with numerous circumstances that may have an impact on the final outcomes, including individual and cultural differences, random noise present in the video input, distinctive facial features, and various articulation styles [6][7]. Automatic human personality judgements have been used to enhance the recognition system in many applications or practical solutions in social computing. As an example, personality recognition is helpful for predicting sentiment polarity and subjectivity in Twitter [8]. Afterward, a virtual robot called Zara was developed by Fung *et al.* [9], who used emotion and sentiment recognition to interact with user. Zara is also able to give a personality analysis based on users' pronouncement of speech. Personality traits recognition have also been exploited in user profiling, product personalization and business marketing. As an example, Micu *et al.* [10] proposed an AI platform which automatically gathers customer data during their stores visit, for managers to create novel solutions that are customized to specific client demands and desires.

This study aimed to explore CNN-based models, which have been used and modified in the personality computing field. In depth, this study intended to discover the structure of CNN-based models such as VGGNet, Resnet, FaceNet and

OpenFace which have been used for personality traits recognition. This study was conducted according to the guidelines of Kitchenham and Charters guidelines of systematic literature review (SLR) method [11]. This study selected previous studies published between 2016 and 2023 through academic databases including IEEEXplore, Web of Science, Scopus, SpringerLink, ScienceDirect, and Google Scholar, using the following keywords: "facial feature extraction", "personality computing", "personality traits recognition", "CNN" and "Big Five model". We identified several key papers that were strongly related to the use of Convolutional Neural Network Model for Facial Feature Extraction in Personality Traits Recognition. The remainder of this paper is organized as follows: In Section 2, we discuss the background of facial features extraction for personality traits recognition. Section 3 explains about CNN-based model for facial features extraction including VGGNet, ResNet, FaceNet and OpenFace model. Finally, Section 4 briefly discusses findings and potential future research direction.

2. Facial Features Extraction for Personality Traits Recognition

Generally, there are several main processes in the development of the personality traits recognition model which are carried out consecutively, namely data pre-processing, features extraction and selection, classification modelling and final prediction. During the data pre-processing stage, information from the raw data source (video) will be extracted separately, consisting visual, audio, and text. Pre-processing also requires the process of noise reduction, data normalization and compression of stored information to produce clean input data. Next, feature extraction step aimed to extract as many as related features from the input data while feature selection concerned to choose the most relevant features for better classification and final prediction score [12].

Different feature modalities such as visual images, audios and texts have been used to automatically classify a person's personality [3]. Based on existing research, combination of multimodal features was a good choice for personality prediction [1][2]. The most frequently used features are visual features which involve facial features from face images [13][14]. Detecting faces and their features such as the mouth, nose, eyes, gazes, and head nods was a difficult challenge in the past. In contrast, nowadays deep learning algorithms are able to solve this task. By using a simple convolutional neural network (CNN), it can easily help to detect key points on parts of faces images. Previous study has demonstrated that there is a correlation between facial key points and the Big Five personality model attributes. Cai & Liu [15] discovered that the points from the right jawline to the chin contour showed a significant negative correlation with agreeableness in their study. According to Kachur *et al.* [16], artificial neural networks algorithm was able to reveal multidimensional personality profiles based on static morphological facial features. Morphological facial features involve the shape and structure of the front of the head from the chin to the top of the forehead where the mouth, eyes, nose and other features are located. Another study found that personality traits can be reliably predicted from facial images using deep learning (CNN and Resnet) and showed an accuracy results more than 70%. Based on the study experiments, it showed that accuracy of neuroticism and extroversion was the most accurate with accuracy exceeded 90% [4]. In the next section, we discussed details of neural network design in CNN-based model for facial features extraction.

3. Convolutional Neural Network (CNN) Model

3.1. VGGNet

VGGNet is one of CNN-based models that supports up to 19 layers and is primarily concerned with the effect of convolutional neural network depth on its accuracy. VGGNet was introduced by Simonyan & Zisserman from the Visual Geometry Group Lab of Oxford University [17]. VGGNet was driven by the desire to reduce the number of parameters in the CONV layer and enhance training time. The VGGNet model has several versions, such as VGG11, VGG16, VGG19, and etc, with the sole difference between each model is the total number of layers in the network. For example, VGG19 has 19 layers with 144 million parameters while VGG16 has 16 layers with 138 million parameters. VGGNet has been used in personality traits recognition to extract facial images from the series of videos input [7][18]. Generally, VGGNet model is like traditional CNN model where it consists of four layers including input layer, convolutional layers, fully connected layers, and hidden layers. The convolutional layers in VGGNet are used in a smaller filter (3x3 or 1x1 filter) and the convolution step is fixed. VGGNet also has three fully connected (FC) layers where the first two of FC layers with 4096 channels and followed by FC layers with 1000 channel to predict 1000 labels (ImageNet dataset). In addition, the pooling layer in VGGNet is not followed each convolutional layer but the pooling layers distributed under different conv layers. The convolutional layers of VGGNet model as illustrated in Figure 1.

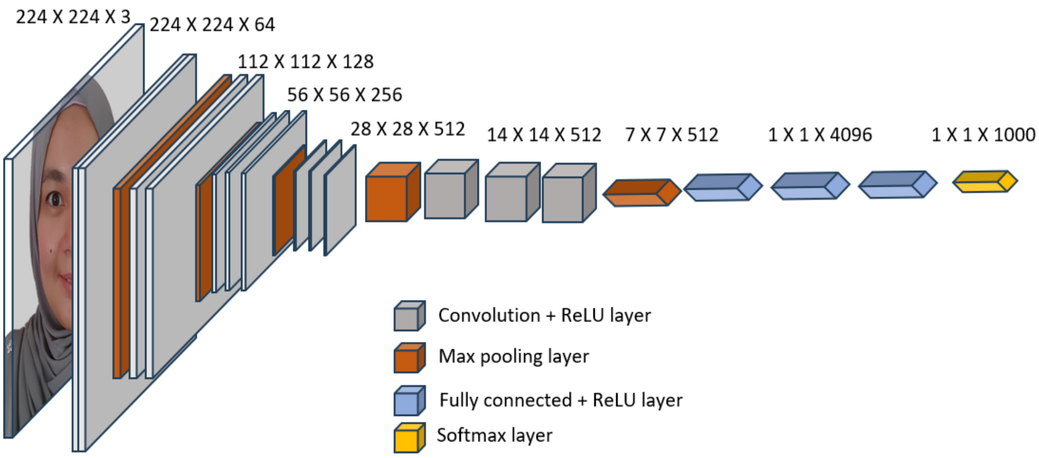

Figure 1. Illustration of VGGNet Model

3.2. ResNet

Residual Network (ResNet) is another type of CNN-based model that was introduced with the basic idea of "skip connections" are built into the network [19]. The skip connections layers in ResNet model aim to overcome the vanishing gradient problem by allowing gradients to flow across these levels. The ResNet model has several models where each differs in the number of layers such as ResNet18, ResNet50 and ResNet101. For example, ResNet18 consists of 18 layers with approximately 11 million parameters while Resnet50 has 50 layers with around 23 million parameters. The more numbers of layers stacked on the network model the more parameters produced by model. Table 1 shows number of layers and parameters for several ResNet models on ImageNet dataset [19]. Previous studies on personality traits recognition have utilized ResNet101 model to extract face features for personality prediction and yielded prediction accuracy more than 91% [1][2].

Table 1. ResNet Model (ImageNet)

	Number of Layers	Number of Parameters
ResNet18	18	~11 M
ResNet34	34	~21 M
ResNet50	50	~23 M
ResNet101	101	~42 M
ResNet152	152	~ 58 M

3.3. FaceNet and OpenFace

FaceNet was developed to perform the task of face recognition, verification and clustering using deep convolutional. FaceNet model achieved state-of-the-art face recognition performance using only 128-bytes per face on Labeled Faces in the Wild (LFW) dataset [20]. FaceNet differs from other models where it learns the mapping from the input images and builds embeddings without relying on any bottleneck layer for recognition or verification tasks. Another significant point of FaceNet is its loss function where it uses triplet loss function namely anchor, positive and negative. The model should ensure that anchor image distances are closer to positive images as compared to negative images. FaceNet model consists of 22 convolutional layers with approximately 140 million parameters.

On the other hand, Open Face was developed with capability to performed facial landmark, head pose, and eye-gaze detection [21]. The OpenFace model was inspired by Facenet, yet the OpenFace is more lightweigh. OpenFace has a weight of about 14MB, compared to FaceNet and VGG-Face which have weights of 90MB and 566MB respectively. While implementing openFace to analyze an image, the face detection library will initially create a bounding box around the face and then send each face to the neural network separately. Next, the model computes 128 dimensions face embeddings to quantify a face and then trains a Support Vector Machine (SVM) on top of the embeddings. Finally, the model successfully recognizes faces in images. OpenFace model was highly recommended for face recognition system using mobile device or real-time application using simple webcam [22]. Besides that, OpenFace is also employed in personality traits recognition to handle visual modality features extraction especially facial features detection [5][23].

Table 2 shows the summary of previous studies which adopted CNN-based model for facial features extraction in personality traits recognition. The summary table consists of information regarding the type of CNN-based model, extracted features and model accuracy. Based on Table 2, it shows that CNN-based models have been frequently used for local face cues extraction while OpenFace models have also been used for extracting eye gazes and head positions.

Table 2. Model Accuracy on Previous Study

CNN Model	Authors [Ref]	Features	Model Accuracy
VGGNet	Zhang *et al.* [7]	Local face cues	0.9130
	Kaya *et al.* [24]	Local face cues	0.9173
	Zhao *et al.* [18]	Local face cues	0.9167
ResNet	Yunan Li *et al.* [25]	Local face cues	0.9188
	Aslan *et al.* [1]	Local face cues	0.9180
	Suman *et al.* [2]	Local face cues	0.9143
OpenFace	Subramaniam *et al.* [26]	Local face cues	0.9120
	Hemamou *et al.* [23]	Local face cues and head positions	0.6450
	Sun *et al.* [5]	Local face cues and eyes gaze	0.9207
FaceNet	Williams *et al.* [27]	Local face cues	0.9060

4. Conclusion

As discussed in this paper, several CNN-based models have been developed in previous studies with various enhancements. For each model discussed above, they are greatly performed in their own way. Each of these models also has its own advantages and disadvantages. For example, ResNet-152 model requires a lot of computing resources in terms of training time and computational power. In contrast, OpenFace has the advantage of being more lightweight compared to Resnet and FaceNet models. Meanwhile, VGGNet was able to solve image recognition tasks using deep network layers and small convolutional filters with highly accurate prediction result. Thus, there is huge potential to enhance and modify neural network layer in CNN model to generate more efficient model for facial features extraction task. Currently CNN-based model increases the number depth layer to extract more relevant features and resulted the more parameters in the end layer. Whenever the network depth is deeper, the fitting ability of CNN is stronger. However, simply increasing the depth of the network layer alone will not guarantee increased accuracy [28]. On the other hand, when the network depth exceeds a certain range, this leads to phenomenon of features disappearing [29] overfitting [30] and efficiency reduction [31]. Thus, the modification on CNN-based model is not only regarding the increasing of depth layer but also on how to utilize each layer to improve the quality of extracted features [32]. There are also some challenges associated with face detection and facial features extraction that potentially affect the model performance such as facial expression, illumination, pose variations and occlusion [33]. There are several techniques such as transfer learning, data augmentation, regularization, and hyperparameter tuning that are believed to be able to improve quality of extracted features, as well as improve the accuracy performance of CNN model. In future, it is interesting to explore more on modification of CNN-based model for facial features extraction to improve personality recognition performance.

Acknowledgments

We would like to thank our fellow researchers and friends for their general support and guidance. We would also thank Librarians at UTM for their assistance in supplying the relevant literature. Our sincere appreciation also extends to all our colleagues and others who have helped on various occasions.

References

[1] Aslan, S., Güdükbay, U., & Dibeklioğlu, H. (2021). Multimodal assessment of apparent personality using feature attention and error consistency constraint. Image and Vision Computing, 110. https://doi.org/10.1016/j.imavis.2021.104163.

[2] Suman, C., Saha, S., Gupta, A., Pandey, S. K., & Bhattacharyya, P. (2022). A multi-modal personality prediction system. Knowledge-Based Systems, 236. https://doi.org/10.1016/j.knosys.2021.107715

[3] Mehta, Y., Majumder, N., Gelbukh, A., & Cambria, E. (2019). Recent trends in deep learning based personality detection. Artificial Intelligence Review. https://doi.org/10.1007/s10462-019-09770-z.

[4] Li, J., Waleed, A., & Salam, H. (2023). A Survey on Personalized Affective Computing in Human-Machine Interaction. 13(9), 1–16. http://arxiv.org/abs/2304.00377.

[5] Sun, X., Huang, J., Zheng, S., Rao, X., & Wang, M. (2022). Personality Assessment Based on Multimodal Attention Network Learning with Category-Based Mean Square Error. IEEE Transactions on Image Processing, 31. https://doi.org/10.1109/TIP.2022.3152049.

[6] Vukojičić, M., & Veinović, M. (2021). Apparent Personality Analysis Based on Aggregation Model. https://doi.org/10.15308/sinteza-2021-220-225.

[7] Zhang, C. L., Zhang, H., Wei, X. S., & Wu, J. (2016). Deep bimodal regression for apparent personality analysis. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9915 LNCS. https://doi.org/10.1007/978-3-319-49409-8_25.

[8] Celli, F., & Zaga, C. (2013). Be conscientious, express your sentiment! CEUR Workshop Proceedings, 1096.

[9] Fung, P., Dey, A., Bin Siddique, F., Lin, R., Yang, Y., Bertero, D., Yan, W., Yin, R. C. H., & Wu, C. S. (2016). Zara: A virtual interactive dialogue system incorporating emotion, sentiment and personality recognition. COLING 2016 - 26th International Conference on Computational Linguistics, Proceedings of COLING 2016: System Demonstrations.

[10] Micu, A., Capatina, A., Cristea, D. S., Munteanu, D., Micu, A. E., & Sarpe, D. A. (2022). Assessing an on-site customer profiling and hyper-personalization system prototype based on a deep learning approach. Technological Forecasting and Social Change, 174. https://doi.org/10.1016/j.techfore.2021.121289.

[11] B.A Kitchenham & Charters. (2007). Guidelines for performing systematic literature reviews in software engineering. Technical Report, Ver. 2.3 EBSE Technical Report. EBSE, 1.

[12] Saleem, F., Khan, M. A., Alhaisoni, M., Tariq, U., Armghan, A., Alenezi, F., Choi, J. I., & Kadry, S. (2021). Human gait recognition: A single stream optimal deep learning features fusion. Sensors, 21(22). https://doi.org/10.3390/s21227584.

[13] Fu, J., & Zhang, H. (2021). Personality trait detection based on ASM localization and deep learning. Scientific Programming, 2021. https://doi.org/10.1155/2021/5675917.

[14] Song, S., Jaiswal, S., Sanchez, E., Tzimiropoulos, G., Shen, L., & Valstar, M. (2021). Self-supervised Learning of Person-specific Facial Dynamics for Automatic Personality Recognition. IEEE Transactions on Affective Computing. https://doi.org/10.1109/TAFFC.2021.3064601.

[15] Cai, L., & Liu, X. (2022). Identifying Big Five personality traits based on facial behavior analysis. Frontiers in Public Health, 10. https://doi.org/10.3389/fpubh.2022.1001828.

[16] Kachur, A., Osin, E., Davydov, D., Shutilov, K., & Novokshonov, A. (2020). Assessing the Big Five personality traits using real-life static facial images. Scientific Reports, 10(1). https://doi.org/10.1038/s41598-020-65358-6.

[17] Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings.

[18] Zhao, X., Tang, Z., & Zhang, S. (2022). Deep Personality Trait Recognition: A Survey. Frontiers in Psychology, 13, 2390. https://doi.org/10.3389/FPSYG.2022.839619/BIBTEX.

[19] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016-December. https://doi.org/10.1109/CVPR.2016.90.

[20] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 07-12-June-2015. https://doi.org/10.1109/CVPR.2015.7298682.

[21] Amos, B., Ludwiczuk, B., & Satyanarayanan, M. (2016). OpenFace: A general-purpose face recognition library with mobile applications. CMU School of ..., 2016 R., 16(October 2015).

[22] Baltrusaitis, T., Robinson, P., & Morency, L. P. (2016). OpenFace: An open-source facial behavior analysis toolkit. 2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016. https://doi.org/10.1109/WACV.2016.7477553.

[23] Hemamou, L., Felhi, G., Vandenbussche, V., Martin, J. C., & Clavel, C. (2019). HireNet: A hierarchical attention model for the automatic analysis of asynchronous video job interviews. 33rd AAAI Conference on Artificial Intelligence, AAAI 2019, 31st Innovative Applications of Artificial Intelligence Conference, IAAI 2019 and the 9th AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019. https://doi.org/10.1609/aaai.v33i01.3301573.

[24] Kaya, H., Gurpinar, F., & Salah, A. A. (2017). Multi-modal Score Fusion and Decision Trees for Explainable Automatic Job Candidate Screening from Video CVs. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2017-July. https://doi.org/10.1109/CVPRW.2017.210.

[25] Li, Y., Wan, J., Miao, Q., Escalera, S., Fang, H., Chen, H., Qi, X., & Guo, G. (2020). CR-Net: A Deep Classification-Regression Network for Multimodal Apparent Personality Analysis. International Journal of Computer Vision, 128(12). https://doi.org/10.1007/s11263-020-01309-y.

[26] Subramaniam, A., Patel, V., Mishra, A., Balasubramanian, P., & Mittal, A. (2016). Bi-modal first impressions recognition using temporally ordered deep audio and stochastic visual features. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9915 LNCS. https://doi.org/10.1007/978-3-319-49409-8_27.

[27] Williams, J., Kleinegesse, S., Comanescu, R., & Radu, O. (2018). Recognizing Emotions in Video Using Multimodal DNN Feature Fusion. 11–19. https://doi.org/10.18653/v1/w18-3302.

[28] Wesselink, E. O., Elliott, J. M., Coppieters, M. W., Hancock, M. J., Cronin, B., Pool-Goudzwaard, A., & Weber, K. A. (2022). Convolutional neural networks for the automatic segmentation of lumbar paraspinal muscles in people with low back pain. Scientific Reports, 12(1). https://doi.org/10.1038/s41598-022-16710-5

[29] Xie, X., Ma, Y., Liu, B., He, J., Li, S., & Wang, H. (2020). A Deep-Learning-Based Real-Time Detector for Grape Leaf Diseases Using Improved Convolutional Neural Networks. Frontiers in Plant Science, 11. https://doi.org/10.3389/fpls.2020.00751

[30] Zhai, Y., Cao, H., Deng, W., Gan, J., Piuri, V., & Zeng, J. (2019). BeautyNet: Joint multiscale CNN and transfer learning method for unconstrained facial beauty prediction. Computational Intelligence and Neuroscience, 2019. https://doi.org/10.1155/2019/1910624

[31] Shepley, A. J. (2019). Deep Learning for Face Recognition: A Critical Analysis. https://arxiv.org/abs/1907.12739v1

[32] Luo, J., & Hu, D. (2023). An Image Classification Method Based on Adaptive Attention Mechanism and Feature Extraction Network. Computational Intelligence and Neuroscience, 2023. https://doi.org/10.1155/2023/4305594.

[33] Borude, P. R., Gandhe, S. T., Dhulekar, P. A., & Phade, G. M. (2015). Identification and tracking of facial features. Procedia Computer Science, 49(1). https://doi.org/10.1016/j.procs.2015.04.220.