WILEY | Hindawi

Research Article

2.5D Facial Personality Prediction Based on Deep Learning

Jia Xu [iD],[1,2] Weijian Tian,[1] Guoyun Lv [iD],[1] Shiya Liu [iD],[3] and Yangyu Fan [iD][1]

[1]*School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China*
[2]*North China University of Science and Technology, Tangshan, China*
[3]*Content Production Center of Virtual Reality, Beijing, China*

Correspondence should be addressed to Guoyun Lv; lvguoyun101@nwpu.edu.cn

The assessment of personality traits is now a key part of many important social activities, such as job hunting, accident prevention in transportation, disease treatment, policing, and interpersonal interactions. In a previous study, we predicted personality based on positive images of college students. Although this method achieved a high accuracy, the reliance on positive images alone results in the loss of much personality-related information. Our new findings show that using real-life 2.5D static facial contour images, it is possible to make statistically significant predictions about a wider range of personality traits for both men and women. We address the objective of comprehensive understanding of a person's personality traits by developing a multiperspective 2.5D hybrid personality-computing model to evaluate the potential correlation between static facial contour images and personality characteristics. Our experimental results show that the deep neural network trained by large labeled datasets can reliably predict people's multidimensional personality characteristics through 2.5D static facial contour images, and the prediction accuracy is better than the previous method using 2D images.

1. Introduction

There has been a long history of attempts to assess personality based on facial morphological features [1], a practice known as physiognomy. Of course, it is not only the East, but people of all ages all over the world have studied this field, and their attitudes are mixed. Aristotle once proposed that facial features can reflect personality characteristics to some extent. Personality is a kind of psychological structure in which a few stable and measurable individual characteristics are used to explain people's different behaviors [2]. Today, there is increasing research interest on the relationship between facial images and personality prediction. Studies by Todorov et al. [3–5] and others have shown that experts can reliably infer a person's personality traits from his facial appearance and use it to track crimes, campaigns, and medical care.

To date, there have been enough researches on character recognition by machine learning worldwide. In [6], the authors studied and proposed the key facial features that have an import impact on people's first impression. We can draw at least four valid inferences from other people's facial features [7]. Reference [8] examined the relationship between self-reported personality traits and first impressions. To investigate whether a computer can learn to assess human traits, the authors of [9] used a machine learning method to construct automatic feature predictors based on facial structure and appearance descriptors and found that all personality traits analyzed were accurate predictors. Recent studies on the static features of human face suggested that certain areas have evolved to play an important role in social communication [10] and that individuals with higher facial attractiveness personality traits have higher mating success [11].

Previous studies, either based on the five-factor model of personality or based on the Big Five (BF) model, have found there is a certain relationship between facial images and general personality characteristics. However, a cursory assessment of the predictions and results of these studies will reveal much controversy, with research results seemingly inconsistent and difficult to replicate [10] (see Table 1 for existing research results). These inconsistencies may be due

to the use of small numbers of stimulation schemes or to very large differences in the corresponding methods. Existing research datasets are insufficient, and most current research on face feature extraction has prioritized 2D faces, especially from front views [23]. However, a reliance only on 2D positive images causes much valuable information related to personality to be lost [24]. Prominent facial areas, such as prominent landmarks of the forehead, nose, and chin, are related to a person's personality [25]. In fact, positive and lateral facial expressions are naturally complementary. Therefore, multiple perspectives (front, side, and 2.5D) facial images are more likely to describe a person's personality comprehensively and accurately. Herein, we use the term "2.5D" to refer to combinations of front and side views.

The two main topics in existing face personality prediction research are the acquisition of datasets (face photos and personality data) and the design of computing networks.

1.1. Construction of Face Database.
The establishment of the face database plays a vital role in verifying our model and ensuring its ability of generalization. Ideally, the face database should contain face samples from people of different genders, races, and ages, displaying different personality traits. However, to date, there is no such database for automatic personality calculation. In fact, differences in research backgrounds often result in the creation of independent databases based on their specifics situations; consequently, the number, age, gender composition, expression, race, and posture of existing facial image samples are not identical. A summary of the facial image databases constructed in the existing face-based personality computing research can be seen in Table 2.

1.2. Selection of the Personality Evaluation Model.
Intuitively, evaluating a person's personality traits involves learning how to choose the adjectives from trait theory to describe it accurately. In all the literature on automatic personality prediction, two ways to evaluate a person's personality traits are described: (1) self-evaluation and (2) evaluation by others. Completion of the personality assessment scale in the first person, that is, a self-assessment, is traditionally considered to produce a person's real personality [22]. Completion of the questionnaire in the third person (e.g., substituting "this person tends to be sociable" for "I'm sociable") leads to attribution and results that are evaluated by others. Under the evaluation of others, each topic must be evaluated by several evaluators, and each evaluator must evaluate all the participants in the experiment. Statistical criteria such as the reliability in [27] allow the number of assessors to be set according to the agreement of both parties.

The theory of personality traits states that personality traits are an effective characteristic of individual behavior, an effective component of an individual, and a basic unit commonly used to evaluate personality. Common theories about traits include Allport's trait theory (common traits and personal traits), Cattell's theory of personality trait (in

which everyone has 16 traits), Eysenck's three-factor model (extroversion, psychoticism, and neuroticism), Tapperth's five-factor model (commonly known as the Big Five: extroversion, agreeableness, sense of responsibility, neuroticism, and openness), and Terrigen's seven-factor model (positive emotionality, negative potency, positive potency, negative emotionality, reliability, agreeableness, and heredity). In face-based personality computing, one important element is the selection of the appropriate trait theory model.

In existing research on personality prediction based on facial features, the methods used to evaluate personality are as shown in Table 3 below.

1.3. Selection of the Prediction Network.
In recent studies on facial personality, many methods have been adopted, such as the Parzen window [29], decision tree [30], naive Bayes [31], kNN, [31] and random forest [32]. Rojas et al. [14, 15] conducted a classification experiment using the most advanced classifier. Zeng [17] used a deep confidence network classification algorithm based on the backpropagation (BP) algorithm. Brahnam and Nanni [20] used principal component analysis (PCA) and random combinations of training and testing sets to train and test their models with 20 repetitions for each personality feature dimension. Kachur et al. [16] proposed a computer vision neural network (NNCV). Methods used for personality prediction in different studies are shown in Table 4.

The aim of this study is to investigate the association between facial image cues and self-reported Big Five personality traits by training a series of neural networks to predict personality traits in static face images. In view of the problems in previous studies, the contributions and innovations of this paper include the following: firstly, a large dataset composed of facial photos and personality characteristics is constructed. The dataset contains 13,347 pairs of data, 360 of which were collected from facial profile images, and a 2.5D dataset was constructed to obtain a more comprehensive understanding. Secondly, an improved deep learning algorithm was used to predict personality characteristics, which reduced the requirements from previous research on the quality of the face images; it was expected that a complex deep learning algorithm could be used to capture face images under uncontrolled conditions. Thirdly, the changing trend of facial characteristics of Asian college students with five personality dimensions ranked from high to low was predicted.

The experimental results of this paper show that, on the one hand, we can reliably predict some personality traits using static facial images; on the other hand, the performance of the facial feature extraction model in predicting personality based on 2.5D images is better than that of 2D images.

The rest of this paper is organized as follows. In Section 2, we describe the creation of our own dataset, including the face dataset and the Big Five personality assessment result dataset. In Section 3, we predict the personalities of positive faces based on an improved deep learning network and the changes in the average face from low to high. The method

TABLE 1: Summary of the literatures on the accuracy of personality prediction.

References	N photo	Assessment	Trait	Highest accuracy (%)
[12, 13]	186	Self-report	Conscientiousness (M)	81.56
			Conscientiousness (F)	82.22
			Skepticism (M)	72.64
			Skepticism (F)	82.22
			Dominance	91.23
[14, 15]	66	Other report	Threatening	90
			Extraversion	
[7]	244	Self-report	Openness, striving, and domination (F)	63
			Reliability, friendliness, and responsibility (M)	65
[16]	12,447	Self-report	Big Five (BF)	58
[17]	608	Self-report	Neuroticism	82.35
			Extraversion	84.31
			Rigorism	84.31
[18]	829	Other report	Neuroticism, openness, and extraversion	65
[19]	1856	Other report	Criminality	89.51
[15]	220	Other report	Warmth	76
			Reliability	81
[20]	480	Other report	Intelligence, maturity, sociality, dominance, warmth, and credibility	80
[19]	1856	Other report	Criminality	89.5
[21]	10,000	Other report	BF	89.11
[22]	10,000	Other report	BF	90.94

Note. The table reflects the prediction accuracies of existing studies on personality prediction with images of the whole face. *N* photo is the number of samples, assessment is the personality assessment method of the study, M is the result for images of males, and F is the result for images of females.

TABLE 2: Attributes of face datasets in existing studies.

References	Number of samples	Age	Gender	Race	Posture	Expression	Data source
[12, 19]	186	18–22	M-F	Asian	Front	Neutral	Students of Xiamen Institute of Technology (Arts and Science)
[13, 14]	66	20–30	M-F	White	Front	Neutral	Amateur actor face database from Karolinska [26]
[25]	244	18–37	M-F	White	Front	Neutral	Danish University of Technology Campus recruitment tester
[7]	608	18–22	M-F	Asian	Front	Neutral	Undergraduates of different disciplines and grades in a university in Jiangxi Province
[17]	829	20–39	M-F	Varied	Front	Neutral	Color-FERET
[24]	66	20–30	M-F	White	Front	Neutral	Amateur actor face database from Karolinska [26]
[13]	650	30–50	M-F	Varied	Front	Neutral	Images of real politicians
[18]	1856	18–55	M-F	Asian	Front	Neutral	Two subsets separately containing images of criminals and noncriminals
[15]	3998	18–25	M-F	Varied	Front	Neutral	Images downloaded from a social network
[19]	220	Varied	M-F	Varied	Front	Neutral	"FACES" software synthesis
[16]	12,447	Varied	M-F	White	Front	Neutral	Volunteers' self-photos
[20]	480	Varied	M-F	Varied	Front	Neutral	"FACES" software synthesis
[21]	5563	Varied	M-F	Varied	Blend	Neutral	Video collection: the ECCV ChaLearn LAP 2016 competition
[22]	10,000	Varied	M-F	Varied	Blend	Neutral	Video collection: the ECCV ChaLearn LAP 2016 competition

and experimental results of personality prediction with our model based on 2.5D face images are presented and discussed in Section 4. Finally, in Section 5, we analyze and give some examples of the applicability of the research results.

2. Dataset and Preprocessing

2.1. Samples and Procedure. The official language used in this study is Chinese. Participants were anonymous college student volunteers recruited by the research group through advertisements on the social network pages of colleges and universities. The data were based on a sample of 5,560 male and 8,547 female college students aged 18 to 25 (some face photos are shown in Figure 1). They were not paid financially but were given a free report on their Big Five personality traits. The data required for the experiment (face pictures and personality scores) were collected online through a dedicated personality research website and a mobile application. The participants signed and submitted an informed consent form, completed a five-person personality

TABLE 3: Personality evaluation models in personality prediction based on facial features.

References	Personality assessment	Personality trait theory	Feature expression
[12, 19]	Self-report	16 personality factors (16PF)	Score [0, 9]
[13, 14]	Other report	Artificially selected personality descriptors	14-dimensional personality score
[25]	Self-report	Big Five (BF)	Score [0, 9]
[7]	Self-report	BF	Score [1, 60]
[17]	Other report	BF	Score [0, 9]
[24]	Other report	Artificially selected personality descriptors	Score [0, 9]
[13]	Other report	Artificial selection: dominance, attraction, credibility and extraversion	Score
[18]	Other report	Artificial selection: dominance, warmth, sociality and credibility	Score [0, 9]
[15, 20]	Other report	Artificial selection: intelligence, maturity, warmth, sociality, dominance and credibility	Score [1, 3]
[16]	Self-report	BF	Score [0, 60]
[18]	Self-report	Eysenck's personality questionnaire—revised (EPQ-R)	Score [0, 120]
[28]	Other report	Eysenck's three-factor model	Score
[21]	Other report	BF	Score [0, 1]
[22]	Other report	BF	Score [0, 1]

TABLE 4: Personality prediction algorithms used in different studies.

Literature	Publication year	Algorithm
[12, 13]	2011	Parzen window, decision tree, naive Bayesian, kNN, and random forest
[10, 11]	2018	Deep learning network
[23]	2014	Deep confidence network based on the backpropagation algorithm
[24]	2018	Support vector machine (SVM)
[13]	2010	SVM (RankSVM)
[14]	2010	Logistic regression, kNN, SVM, and CNN
[20]	2016	SVM
[14, 33]	2018	SVM and CNN
[16]	2020	Computer vision neural network (NNCV)
[29]	2018	Parzen window
[30]	2018	Decision tree
[31]	2001	k-nearest neighbour and naïve Bayes
[32]	2001	Random forest
[21]	2016	Deep bimodal regression (DBR)
[22]	2016	DCC, UCAS, and BU-NKU

questionnaire, filled in their age, gender, and major, and uploaded frontal photos that showed a neutral, unsmiling expression and to avoid thick facial makeup and other decorations, such as hats. To study the contribution of a person's profile to personality prediction, we also collected pictures of the profiles of an additional 360 students.

2.2. Ethical Approval. Participants were required to agree in writing to participate in the study, and their data was collected only after obtaining their authorization. In addition, we anonymously collected self-reported personality assessment data by assigning a number to each participant. Furthermore, the face and personality data were only used for scientific research, and no personal data will be disclosed to the outside world.

2.3. Establishment of the Personality Dataset (Big Five Personality Traits). To study the contribution of the profile view to personality prediction, we collected profile views from an additional 360 students. We expended much effort to collect

personality trait data from the participants. Research and experimental results over the years have shown that the same behavioral characteristics appear in various environments and cultures with surprising regularity, indicating that they actually correspond to certain similar personality psychological phenomena [18]. Today, the Big Five is considered one of the most dominant and influential models of personality research [28]. This article uses the BF model. The Big Five are openness, conscientiousness, extraversion, agreeableness, and neuroticism. Each dimension is like a ruler, and the personality characteristics of each tester will fall in a certain position of each ruler. The closer this point lies to the end point of the ruler, the greater preference the user has toward the corresponding personality trait. A score of 0 to 60 is set for each dimension, such as agreeableness, where the higher the individual's score, the more easygoing and pleasant the personality is [21, 22].

Questionnaires based on a Likert scale are the most commonly used tools for scoring the BF dimensions [27]. The most popular items include the revised NEO-PI-R (240 items) [34], the NEO-FFI (60 items) [35], and BFI (44 items)

FIGURE 1: Front view images of some Asian college students.

[36] (see [2] for an extensive investigation). By retaining only the items that are most relevant to the results of the whole document [26, 37], a shorter questionnaire (60 items) can be established that can be filled much faster (the 60 questions are given in Supplementary Materials).

First-person questionnaires such as those in Annex 1 lead to self-assessment, which is traditionally considered to produce a person's real personality [14]. For self-assessment, the biggest limitation is that the subject may tend to bias her score toward the characteristics of social expectations, especially when the assessment may have negative results, such as failing an interview. As a result, statements such as "I tend to be lazy" may be rated as disagreeable because the respondent will attempt to convey positive impressions and hide negative features. However, a large number of experiments have shown that the self-assessment results are highly correlated with the evaluations of others provided by familiar observers (spouses, family members, etc.) [33]. This proved to be an important step in accepting the questionnaire as a method of personality evaluation. Therefore, we also used numerous self-assessments in the experiment.

5560 men and 8547 women completed the personality assessment questionnaire and uploaded 14021 photos. After final verification was performed and the face and personality data were merged, the dataset included 13,347 valid questionnaires and 13,347 related photos (see below). Participants ranged in age from 18 to 25, with an average age of 21.4 years for females, accounting for 62.1% of the total, and 20.7 years for males, accounting for 37.8%. We randomly divided the dataset into training dataset, test dataset, and verification dataset, accounting for 90%, 5%, and 5% of the total dataset, respectively. In addition, we randomly collected side face photos of another 360 participants to study the contribution of 2.5D faces to personality prediction.

2.4. Screening and Analysis of Image and Personality Data. Each participant was given scores on five personality trait dimensions based on the Big Five personality test, each of which was scored as a discrete number between 1 and 60. We use the tripartite method to divide the personality scores of different dimensions into "low, medium, and high," such as low neuroticism, medium neuroticism, and high neuroticism. The statistical results are shown in Table 5. For the sake of simplicity, the letters O, C, E, A, and N are used to denote openness, conscientiousness, extraversion, agreeableness, and neuroticism, respectively.

This result is basically in line with the personality characteristics of Asians, who are considered to be relatively conservative, kind-hearted, and introverted. Therefore, in our dataset, people with high openness and high neuroticism accounted for a relatively small proportion. To facilitate calculation and analysis, we classified the personality characteristics of the population into two additional categories: "not obvious" and "obvious" according to the data collected after the survey. Although the data were classified into "high, medium, and low" at first, because the proportions of participants with high neuroticism, high extraversion, and high openness were very low almost to the point of negligibility, we divided these small numbers of participants into the next closest categories (the final classification is shown in Table 6).

We applied the functions of face and eye detection, alignment, resizing, and clipping provided by Dlib library (dlib.net website) to process the face images and obtained a group of normalized images with the pixel size of 112×112.

By matching the answers to the questionnaire with the face photos one by one, we were able to obtain a valid set of Big Five questionnaires and images, totaling 13,347 pairs.

3. Neural Network for Personality Prediction Based on 2D Images

Previous studies on personality prediction were conducted by means of artificial feature collection, which could result in the loss of personality-related features [12, 14, 20, 38–41]. We predict that personality characteristics will be reflected in the person's entire facial image (including the profile) rather than in a certain number of isolated facial features. Consequently, we employed a deep learning method to extract high-level features from face images for personality prediction. We used MobileNetV2 and residual network version 50 (ResNet50), two deep learning networks that are popular in academia, to classify personality traits. Then, an improved personality prediction network—Soft Threshold-Based Neural Network for Personality Prediction (S-NNPP)—was proposed. To verify the experimental results, 5-fold cross-validation method was used. The data were randomly scrambled and divided into five pieces, and for each fold, one piece of data was further divided into equally sized test and validation sets, and the remaining four pieces as the training set. Take the average of the verification results from the five folds as the final result. In the training process, focus loss, data enhancement, upsampling, and cost sensitivity were introduced to solve the problem of sample imbalance. All training in this section was fine-tuned based on the ImageNet pretrained model with stochastic gradient descent as our optimization strategy. At the beginning of training, we set the learning rate to 0.001 and adopted the

TABLE 5: Three-category personality data classification.

Category	Trait (number of people)				
	O	C	E	A	N
Low	6328	5043	18	7934	29
Medium	7019	8226	10,081	5413	12,861
High	0	78	3248	0	457
Total	13,347	13,347	13,347	13,347	13,347

TABLE 6: Two-category personality data classification.

Category	Trait (numbers)				
	Neuroticism	Extraversion	Openness	Agreeableness	Conscientiousness
Not obvious	12,861	10,081	6328	7934	5043
Obvious	486	3266	7019	5413	8304
Total	13,347	13,347	13,347	13,347	13,347

ReduceLROnPlateau, which can adjust the learning rate dynamically according to the loss, as the learning rate optimization strategy.

3.1. Soft Threshold-Based Neural Network for Personality Prediction (S-NNPP). Recent studies have shown that networks based on attention mechanisms can achieve good performance in classification tasks. Therefore, an increasing number of networks add various attention operations in ResNet. The ResNeSt proposed in 2020 can be regarded as an "integrated master" that incorporates the best of the previous versions of ResNet. Based on the in-depth analysis of GoogLeNet, the selective kernel network (SKNet), and the squeeze-and-excitation network (SENet), a deep neural network called S-NNPP was designed. The objective is to select a network architecture for multiscale image feature extraction and to achieve good image classification performance. We introduced the multipath mechanism of GoogLeNet and the feature map attention module of SKNet in ResNeSt. We also introduced channel attention by adaptively recalibrating the channel characteristic response, following the architecture of SENet. Due to the outstanding performance of ResNeSt in image classification, we employed its basic modules to make subsequent network improvements.

The network diagram in Figure 2 shows that in the ResNeSt block, the 3 ∗ 3 convolution in ResNet was replaced by grouping convolution through splitting, and attention was paid through multiple branches. Here, grouped convolution was used in every path. Finally, following the softmax operation, the convolution results of each group were merged.

Personality data are easily labeled as fuzzy, so in this study, we introduce soft thresholding [42] to improve the model's adaptability to noisy data. In many signal denoising methods, soft thresholding was the core step. It is used to set the feature whose absolute value is below a certain threshold value to 0 and adjusting other features accordingly—that is, to perform shrinkage. Here, the threshold is a parameter that must be set in advance, and its value has a direct impact on the noise reduction results. In terms of soft thresholding

operation, the input-output relationship is shown in Figure 3.

The soft-thresholding formula is as follows:

$$\text{soft}(x, T) = \begin{cases} x + T \, (x \leq -T), \\ 0 \, (|x| \leq T), \\ x - T \, (x \geq T), \end{cases} \quad (1)$$

where $|x|$ is the wavelet transform coefficient and T is the preselected threshold.

It can be seen from the formula and the figure that the soft threshold function removes features whose absolute value is less than the threshold T and shrinks the features whose absolute value is greater than the threshold toward 0. When applied to the network, it can compress and retain the important features and filter the unimportant features. Due to the influence of various factors, the redundant information contained in different samples tends to be different, so different thresholds need to be set for different samples. Therefore, when performing soft threshold segmentation on the feature maps, we added a subnetwork to the basic network to automatically learn a set of thresholds. In this way, a unique set of soft thresholds can be obtained for each sample to remove redundant information. The adjusted basic network module is shown in Figure 4.

Figure 5 shows the network architecture of the soft threshold block.

In general, the Soft_ResNeSt network consists of two paths, one taking the entire image as input, and the other taking only the face region, which was obtained by an opensource OpenCV face region extractor. The improved ResNeSt module described above was used in the basic module of the two paths, and then according to a weighted parameter α, the prediction results were fused. Figure 6 shows the overall structure of our network.

In this section, the traditional BP network, two kinds of deep learning networks, and the improved S-NNPP network are compared in terms of their performance in personality prediction. Among them, the classification results of the lightweight MobileNetV2 for the personality data are good for neuroticism and extroversion, while the effect of
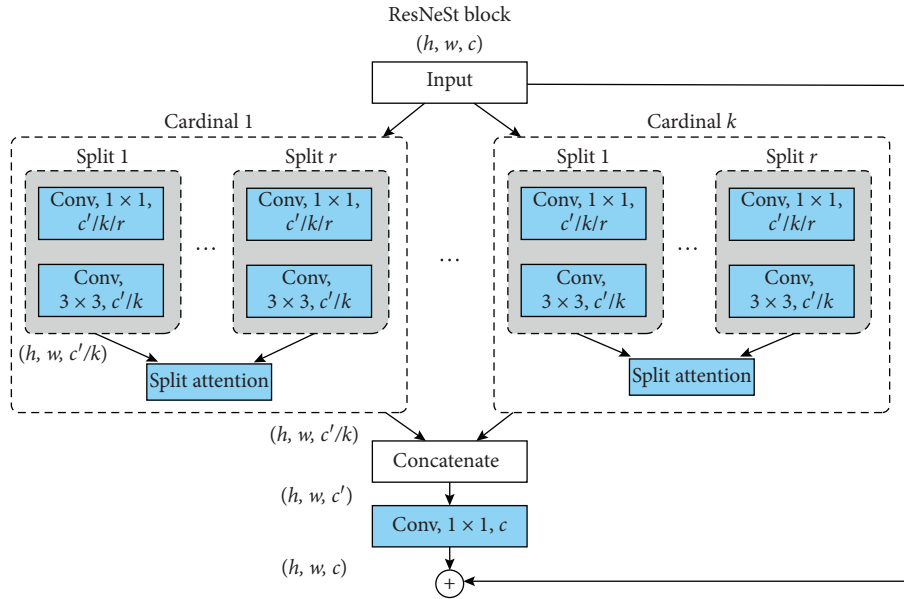
ResNeSt block
(h, w, c)

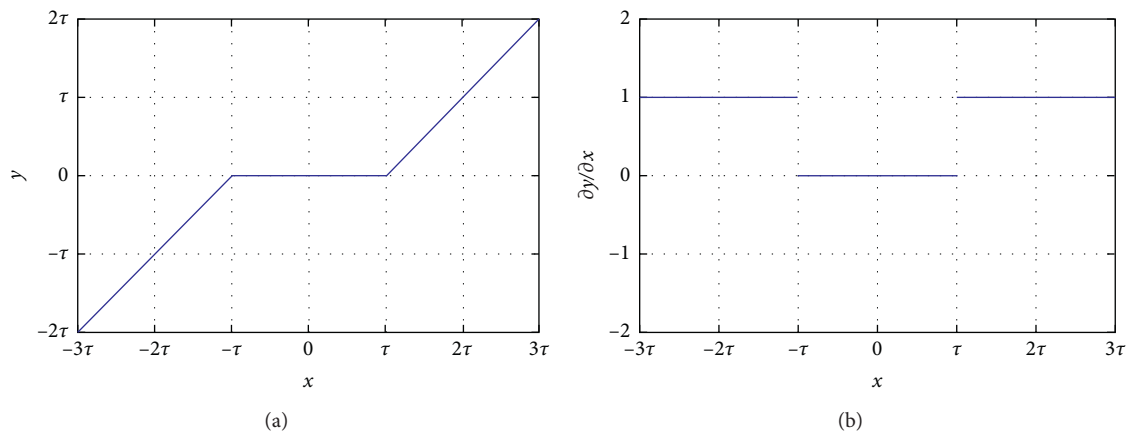FIGURE 2: Network structure of ResNeSt.

(a)

(b)

FIGURE 3: Soft-thresholding curves. The soft threshold function sets the input data with an absolute value lower than this threshold to zero, and the input data with absolute value greater than this threshold is also shrunk towards zero, and the relationship between input and output is shown in (a); the partial derivative of output y of the soft threshold function with respect to input x in (b).

openness, pleasantness, and responsibility is not obvious. Comparatively, the results of the complex network ResNeSt50 were slightly improved, indicating that the complicated network architecture can better extract depth characteristics related to personality. Finally, combining ResNeSt and soft-threshold technology, we generated S-NNPP, an improved personality prediction network, which is substantially better than MobileNetV2 and ResNeSt50 in predicting five personality dimensions.

3.2. Results and Discussion. In this study, the data were scrambled and randomly divided into five sections, one of the sections was further divided into equally sized test and validation sets, and the remaining four sections served as the training set. The verification data we used were from an independent verification dataset, which contained the predicted scores of 1335 facial images of 1335 volunteers. The final prediction result is the average of the verification results from using each of the five parts as the verification set.

We tested the accuracy of different neural networks in predicting five personality traits. The true and false positive rates and F1 scores of the three deep learning networks in predicting the five traits are shown in Table 7. Neuroticism and extroversion were significantly easier to identify than others, as indicated by a recognition rate of over 90% (see Figure 7: receiver operating characteristic (ROC) curves). The degree of recognition openness, agreeableness, and conscientiousness for the three networks is relatively weak but better than that by the line representing chance; this is different from the existing conclusions to some extent [43, 44]. There are several reasons why our research results may be different from other results. Firstly, all our volunteers were Asian, who, due to cultural differences, emphasize their
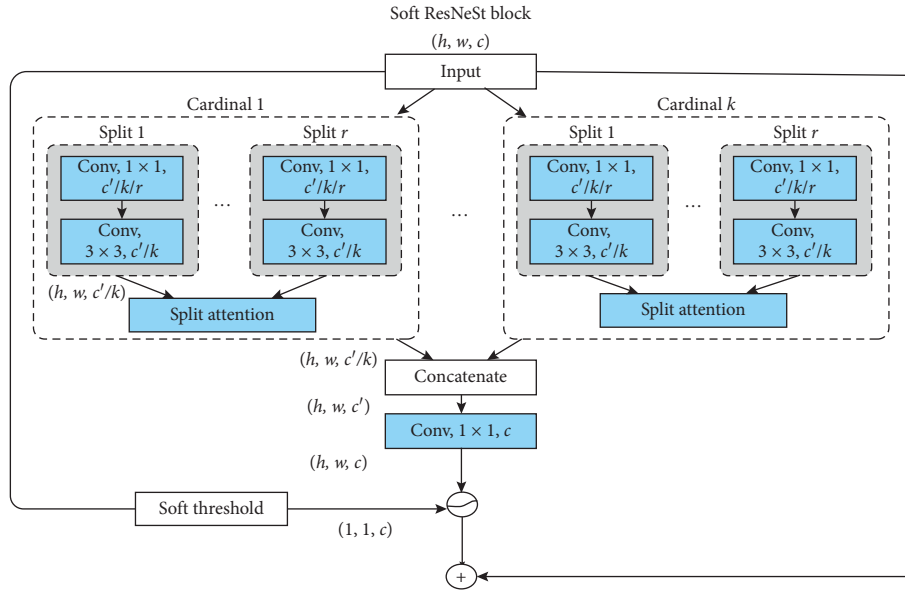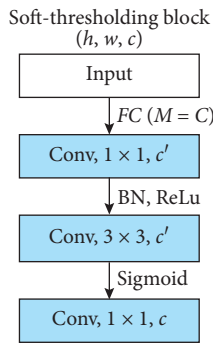
FIGURE 4: Basic network module with soft threshold.

The small network is used to obtain a set of thresholds.

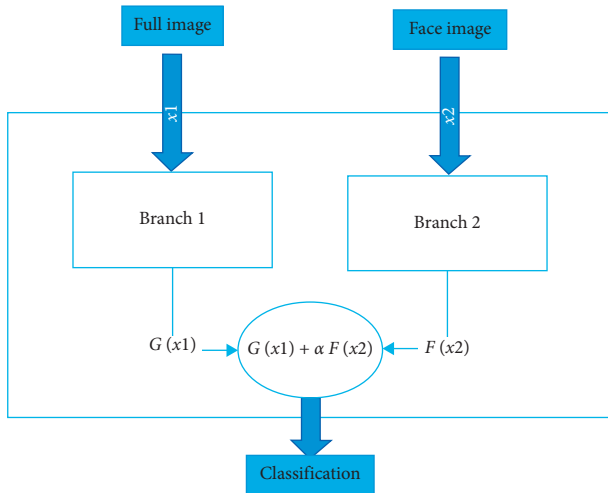FIGURE 5: Network architecture of the soft-thresholding block.

Asians place more emphasis on self-discipline and commitment, preciseness and meticulousness, resourcefulness and determination, and tenacity and steadiness [45]. Secondly, all our volunteers were college students, who tend to have relatively little contact with society and do not take much responsibility. Therefore, their understanding of self-consciousness and agreeableness may not be comprehensive, affecting the corresponding score on the self-esteem scale and further affecting the prediction performance of these two dimensions. Third, our research is based on facial images, in which there are obvious differences between the features of Chinese and Western people. For example, Westerners have obvious facial contours with high noses, while Asians have relatively flat facial contours and soft lines. Therefore, the prediction results of Chinese and foreigners, especially Westerners, personalities based on facial features are bound to be different. The above evidence illustrates the credibility of our results.

The ROC curves also showed that the model has good classification ability for neuroticism and extraversion but a slightly lower classification ability for openness, agreeableness, and conscientiousness.

Our research has shown that a person's personality has a certain relationship with his or her appearance. We estimated that machine learning (the deep learning network in our experiment) could reveal the multidimensional personality characteristics expressed based on the static shape of the face. We developed a neural network and trained it on a large dataset labeled with self-reported BF features without the participation of supervisory, third-party evaluators, avoiding the reliability limitations of human raters.

We further predicted that personality characteristics could be reflected in images of the entire face, not just in individual facial features. In order to verify our hypothesis, we developed S-NNPP, a deep neural network based on the attention mechanism, and added soft thresholding to achieve better prediction performance than existing

FIGURE 6: Overall network structure of S-NNPP.

"openness," "easygoing nature," and "sense of responsibility" less often than their Western counterparts. Instead,

networks. Specifically, we compared its performance with that of the BP network and two kinds of high-performance deep neural networks, MobileNetV2 and ResNeSt50, and found that our S-NNPP network effectively had the best prediction accuracy (see Table 7). We identified three reasons for the improvement in the model accuracy. Firstly, we collected 13,347 pairs of data (including self-reported facial image and personality data), larger than any other dataset yet reported worldwide (the previous record was 12,447 pairs of data in [16]). Secondly, in terms of algorithm improvement, the excellent performance of ResNeSt in ImageNet image classification advantages of network in classification. We used the base module of ResNeSt to develop our improved network. Furthermore, although personality data are easily labeled as fuzzy, we improved the model's ability to process data containing different noises by introducing soft threshold techniques. Thirdly, in our dataset, the face images had relatively consistent backgrounds, distances to the camera, angles, and lighting, making later data processing more convenient.

4. Network Neural Network for Personality Prediction Based on 2.5D Images

In personality prediction, we found that some facial regions, such as the forehead, nose, cheekbones, and chin, whose features cannot be well located in the frontal face image; instead, profile images tend to be required for more accurate detection [23]. In fact, the facial information contained in the positive and lateral perspectives is naturally complementary. As a result, the combination of the two perspectives (i.e., 2.5D) is expected to reflect the relationship between facial images and personality more comprehensively than 2D images alone.

4.1. Experimental Setup. In this section, 360 students (180 males and 180 females) were selected from the previous 2D database for collecting additional facial images, including front and side images, as well as BF personality self-evaluation scores. We again used the 50% cross-validation method described previously for effect analysis; the data were scrambled and randomly divided into five parts, of which one part was further divided into equally sized test and validation sets; the other four parts served as the training set. Specifically, 288 images and the corresponding scores were used as the training dataset, and the remaining 72 images and corresponding scores were used as the test dataset.

It must be noted that the 2.5D images in the database included one front face image and two facial contour images. Experiments have shown that the geometric features of the left and right profiles are highly correlated, and most of the differences between the two sides can be described in the front face images; therefore, the side features were only extracted from the left profile images.

4.2. Results and Discussion. We tested the accuracy in predicting the five personality traits with different networks. Table 8 shows the F1 scores of the three deep learning networks for the five personality traits. Obviously, the frontal and lateral faces emphasize complementary facial regions;

therefore, integration of the two kinds of image should result in more accurate personality measurements, motivating the use of 2.5D modeling used in this study. Following deep neural network training, the 2.5D prediction model achieved better personality prediction than the 2D model; particularly, the F1 score for extraversion increased to 93.02%, and the prediction performance for openness increased to 65.03%. This suggests that these two personality traits are more correlated with the information provided by facial contour images. There was no change, obvious or otherwise, in the prediction of other characteristics, which was directly related to the small size of the experimental 2.5D dataset.

Although we do not know how to train deep neural networks to learn human facial features, we know that the facial features extracted from the front and side images are completely different. Some facial features can be well described by front face images, while others can be accurately expressed in facial profile images. This indicates that the combination of the two perspectives provides more information about personality, so it is possible to further improve the performance of personality prediction.

5. Application

The personality characteristics of a person identified from facial images in real life can be used in many scenes. In daily social affairs, this technique is very useful for identifying personality types. Our method is a further development of traditional personality assessment methods. Facial-feature-based personality matching is expected to become a popular feature of all kinds of job hunting, social networking, and other similar websites, which can quickly recommend faces to users according to their preferences [45]. In addition, facial-feature-based personality prediction research can also be used in crime tracking [23], security inspections in transportation, driver evaluation, and target employee selection based on the images of faces in public security departments. We believe that considering the speed and low cost of this technology, its application potential is vast.

We recruited college students as the research objects in this study. In terms of application scenarios, our findings could help college students find jobs that suit their personalities in the form of "person-post matching"; the technology can provide a quick personality analysis to help employers interview employees when hiring. These models can also be applied to auxiliary functions such as student online consultation. However, we do not advocate solely relying on artificial intelligence to "identify people."

We conducted experimental analysis with an additional dataset to ensure the validity of our experimental results. This dataset consists of two groups of face images, each of which contains faceless photos of the corresponding sample that we initially collected. The two groups of pictures include images of students who qualified for postgraduate study without exams and images of students who were about to drop out due to failing a large number of courses or for violating school guidelines. Each group contains 8 face images, as shown in Figure 8.

TABLE 7: Prediction performance comparison.

| Trait	Traditional BP (%)			MobileNetV2 (%)			ResNeSt50 (%)			S-NNPP (%)		
	TPR	FPR	F1	TPR	FPR	F1	TPR	FPR	F1	TPR	FPR	F1
Neuroticism	80.00	3.28	87.55	85.93	3.03	90.98	87.41	1.52	92.55	92.91	1.43	96.55
Extraversion	72.48	6.78	81.51	75.52	6.45	83.40	76.51	1.69	86.04	84.44	1.52	91.84
Openness	47.24	40.00	49.38	49.59	38.36	50.63	54.14	32.84	57.83	58.54	30.56	62.25
Agreeableness	44.59	45.45	50.54	55.12	35.71	56.68	55.81	34.78	57.83	60.50	32.43	63.25
Conscientiousness	41.29	50.00	46.55	54.42	33.33	59.93	58.39	30.77	62.26	60.69	26.23	65.42
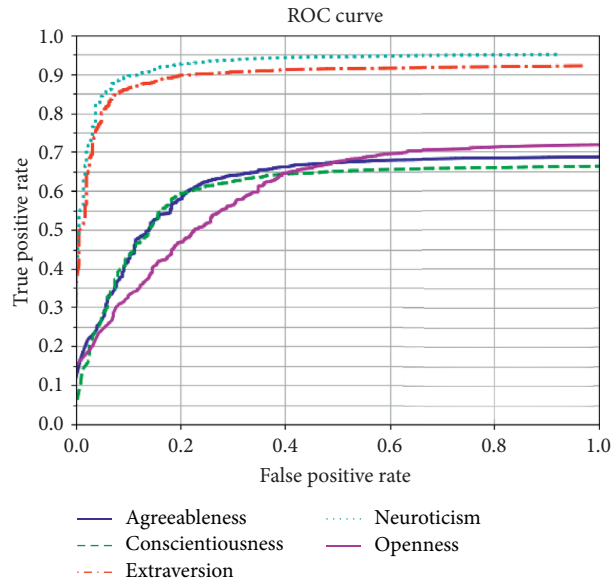
FIGURE 7: ROC curve.

TABLE 8: Prediction performance comparison for the 2.5D model.

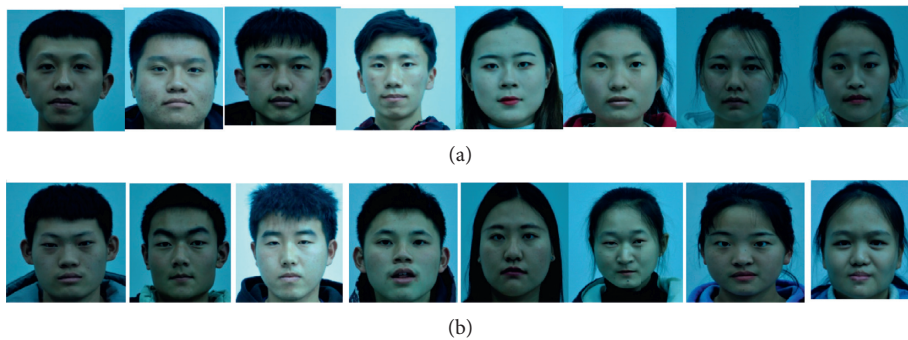| Trait	MobileNetV2 (%)		ResNeSt50 (%)		S-NNPP (%)	
	F1 (2D)	F1 (2.5D)	F1 (2D)	F1 (2.5D)	F1 (2D)	F1 (2.5D)
Neuroticism	90.98	89.89	92.55	90.6	96.55	95.89
Extraversion	83.40	85.04	86.04	88.75	91.84	93.02
Openness	50.63	51.57	57.83	58.06	62.25	65.03
Agreeableness	56.68	55.34	57.83	56.45	63.25	62.87
Conscientiousness	59.93	57.06	62.26	60.09	65.42	65.10

(a)

(b)

FIGURE 8: Frontal images of two different types of students. (a) Eight students who have obtained the qualification of master's degree without examination. (b) Eight dropouts who failed many courses.

For these two types of samples, some of the subjects had excellent grades and strong scientific research capabilities, and some had failed subjects or dropouts due to rule violations; however, there was a general understanding of certain aspects of the personality traits of all subjects. In addition, because samples in the same category share certain characteristics and behaviors in common, there may be a certain degree of similarity between the personality traits of these individuals; therefore, we used the S-NNPP network to verify these results. We found that there are certain similarities in personality traits among samples of the same category. For example, for the 8 postgraduate candidates, the model predicted that conscientiousness and pleasantness were the most significant traits, while some also showed strong openness; for the students at risk of dropping out, their neuroticism was generally high, while no obvious commonality was observed in the other dimensions.

Data Availability

Our research involves a large number of face images. Before the photo collection, we signed a confidentiality agreement, which guarantees that the face images will not be disclosed to the public, so we cannot provide such data.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Acknowledgments

Supplementary Materials

Supplementary material is the Big Five personality scale used in this paper, namely the NEO Personality Scale (the list of questions for self-personality assessment in this study). The 60-item scale measures the Big Five personality traits (Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism). (*Supplementary Materials*)

References

[1] E. L. Hicks, "On the characters of theophrastus." *The Journal of Hellenic Studies*, vol. 3, no. 11, pp. 128–143, 1882.

[2] D. C. Funder, "On the accuracy of personality judgment: a realistic approach," *Psychological Review*, vol. 102, no. 4, pp. 652–670, 1995.

[3] J. Luo and X. Dai, "Development of the Chinese adjectives scale of big-five factor personality III: based on the generalizability theory," *Chinese Journal of Clinical Psychology*, vol. 24, no. 1, pp. 88–94, 2016.

[4] D. J. Ozer and V. Benet-Martínez, "Personality and the prediction of consequential outcomes," *Annual Review of Psychology*, vol. 57, no. 1, pp. 401–421, 2006.

[5] N. N. Oosterhof and A. Todorov, "The functional basis of face evaluation," *Proceedings of the National Academy of Sciences*, vol. 105, no. 32, pp. 11087–11092, 2008.

[6] D. R. Carney, C. R. Colvin, and J. A. Hall, "A thin slice perspective on the accuracy of first impressions," *Journal of Research in Personality*, vol. 41, no. 5, pp. 1054–1072, 2007.

[7] K. Wolffhechel, J. Fagertun, U. P. Jacobsen et al., "Interpretation of appearance: the effect of facial features on first impressions and personality," *PLoS One*, vol. 9, no. 9, Article ID e107721, 2014.

[8] S. Alper, F. Bayrak, and O. Yilmaz, "All the dark triad and some of the big five traits are visible in the face," *Personality and Individual Differences*, vol. 168, Article ID 110350, 2021.

[9] M. Kosinski, "Facial width-to-height ratio does not predict self-reported behavioral tendencies," *Psychological Science*, vol. 28, no. 11, pp. 1675–1682, 2017.

[10] R. M. Godinho, P. Spikins, and P. O'Higgins, "Supraorbital morphology and social dynamics in human evolution," *Nature Ecology & Evolution*, vol. 2, no. 6, pp. 956–961, 2018.

[11] J. M. Carré and J. Archer, "Testosterone and human behavior: the role of individual and contextual variables," *Current Opinion in Psychology*, vol. 19, no. 2, pp. 149–153, 2018.

[12] R. Qin and W. Gao, "Modern physiognomy: an investigation o predicting personality traits and intelligence from the human face," *Science China Information Sciences*, vol. 61, Article ID 058105, 2018.

[13] R. Qin, *Personality Analysis Based on Facial Image*, University of Chinese Academy of Sciences, Beijing, China, 2016.

[14] E. Goeleven, R. De Raedt, L. Leyman, and B. Verschuere, "The karolinska directed emotional faces: a validation study," *Cognition & Emotion*, vol. 22, no. 6, pp. 1094–1118, 2008.

[15] P. V. Shebalin and W. Truszkowski, "Cooperation between intelligent information agents," in *Cooperative Information Agents V. CIA 2001*, M. Klusch and F. Zambonelli, Eds., vol. 2182, Springer, Berlin, Germany, 2001.

[16] A. Kachur, E. Osin, D. Davydov et al., "Assessing the big five personality traits using real-life static facial images," *Scientific Reports*, vol. 10, no. 1, 2020.

[17] Z. Zeng, *An Analysis of Students' Personality Traits Based on Face Features and Deep Learning*, Jiangxi Normal University, Nanchang, China, 2017.

[18] N. A. Moubayed, Y. Vazquez-Alvarez, A. McKay, and A. Vinciarelli, "Face-based automatic personality perception," in *Proceedings of the ACM Multimedia*, Orlando, FL, USA, November 2014.

[19] X. Wu and X. Zhang, "Automated inference on criminality using face images," 2017, http://arxiv.org/abs/1611.04135v2.

[20] S. Brahnam and L. Nanni, "Predicting trait impressions of faces using local face recognition techniques," *Expert Systems with Applications*, vol. 37, no. 7, pp. 5086–5093, 2010.

[21] F. Gurpinar, H. Kaya, and A. A. Salah, "Combining deep facial and ambient features for first impressionestimation," in *Proceedings of the European Conference on Computer Vision*, Springer, Amsterdam, The Netherlands, 2016.

[22] C. Zhang, H. Zhang, S. X. Wei, and J. Wu, "Deep bimodal regression personality analysis," in *Proceedings of the ECCV Workshop Proceedings*, Springer, Amsterdam, The Netherlands, 2016.

[23] A. Laurentini and A. Bottino, "Computer analysis of face beauty: a survey," *Computer Vision and Image Understanding*, vol. 125, no. 8, pp. 184–199, 2014.

[24] T. Zhang, R.-Z. Qin, Q.-L. Dong, W. Gao, H.-R. Xu, and Z.-Y. Hu, "Physiognomy: personality traits prediction by

learning," *International Journal of Automation and Computing*, vol. 14, no. 4, pp. 386–395, 2017.

[25] A. Laurencin and A. Bettino, "Computer analysis of face beauty: a survey," *Computer Vision and Image Understanding*, vol. 125, no. 8, pp. 184–199, 2014.

[26] B. Rammstedt and O. P. John, "Measuring personality in one minute or less: a 10-item short version of the Big Five Inventory in English and German," *Journal of Research in Personality*, vol. 41, no. 1, pp. 203–212, 2007.

[27] G. Boyle and E. Helmes, "Methods of personality assessment," in *The Cambridge Handbook of Personality Psychology*, pp. 110–126, Cambridge University Press, Cambridge, UK, 2009.

[28] J. I. Biel, L. T. Mosquera, and D. G. Perez, "Facetube: predicting personality from facial expressions of emotion in online conversational video," in *Proceedings of the 14th ACM International Conference on Multimodal Interaction*, Santa Monica, CA, USA, 2012.

[29] Z. Liu, Z. Jing, and W. Song, "From parzen window estimation to feature extraction: a new perspective," in *Proceedings of the Intelligent Data Engineering and Automated Learning–IDEAL 2018*, Madrid, Spain, November 2018.

[30] H. Esmaily, M. Tayefi, H. Doosti et al., "A comparison between decision tree and random forest in determining the risk factors associated with type 2 diabetes," *Journal of Research in Health Sciences*, vol. 18, no. 2, Article ID e00412, 2018.

[31] I. Rish, "An empirical study of the Naïve Bayes classifier," vol. 3, pp. 41–46, in *Proceedings of the IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*, vol. 3, pp. 41–46, IBM, New York, NY, USA, 2001.

[32] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[33] A. B. Khromov, *Te Five-Factor Questionnaire of Personality*, Kurgan State University, Kurgan, Russia, 2000.

[34] P. T. Costa and R. McCrae, "Domains and facets: hierarchical personality assessment using the revised NEO Personality Inventory," *Journal of Personality Assessment*, vol. 64, no. 1, pp. 21–50, 1995.

[35] R. R. McCrae and P. T. Costa, "A contemplated revision of the NEO five-factor inventory," *Personality and Individual Differences*, vol. 36, no. 3, pp. 587–596, 2004.

[36] O. John, E. Donahue, and R. Kentle, "The big five inventory-versions 4a and 54," Tech. Rep., Institute of Personality and Social Research, University of California, Berkeley, CA, USA, 1991.

[37] S. D. Gosling, P. J. Rentfrow, and W. B. Swann, "A very brief measure of the Big-Five personality domains," *Journal of Research in Personality*, vol. 37, no. 6, pp. 504–528, 2003.

[38] R. Q. Mario, M. David, T. Alexander et al., "Automatic prediction of facial trait judgments: appearance vs. structural models," *PLoS One*, vol. 6, no. 8, Article ID e23323, 2011.

[39] P. V. Shebalin, *Collective Learning and Cooperation between Intelligent Software Agents: A Study of Artificial Personality and Behavior in Autonomous Agents Playing the Infinitely Repeated Prisoner's Dilemma Game*, The George Washington University, Washington, DC, USA, 1997.

[40] R. Rosenthal, "Conducting judgment studies: some methodological issues," in *The New Handbook of Methods in Nonverbal Behavior Research*, Oxford University Press, Oxford, UK, 2005.

[41] X. Wu, X. Zhang, and C Liu, "Automated inference on sociopsychological impressions of attractive female faces," 2016, http://arxiv.org/abs/1611.04135v2.

[42] D. L. Donoho, "Denoising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 2002.

[43] S. Hu, J Xiong, P Fu et al., "Signatures of personality on dense 3D facial images," *Scientific Reports*, vol. 7, no. 1, p. 73, 2017.

[44] D. R. Valenzano, A. Mennucci, G. Tartarelli, and A. Cellerino, "Shape analysis of female facial attractiveness," *Vision Research*, vol. 46, no. 8-9, pp. 1282–1291, 2006.

[45] D. Wang and C. Hong, "Theoretical and empirical analysis on the differences between Chinese and western personality structure—taking Chinese personality scale (QZPS) and western five factor personality scale (NEO PI-R) as examples," *Acta Psychologica Sinica*, vol. 40, no. 3, pp. 327–338, 2008.