Synthesizing Personality-Dependent Body Postures Using Generative Adversarial Networks

Frederik Calsius and Stylianos Asteriadis

October 11, 2019

Synthesizing Personality-Dependent Body Postures Using Generative Adversarial Networks

Frederik Calsius[1] and Stylianos Asteriadis[1][0000−0002−4298−6870]

Department of Data Science and Knowledge Engineering, Maastricht University, Netherlands

Abstract. In Personality Computing, one of the major goals is to estimate the personality of an individual by making use of computational techniques. Among the existing models that classify personality traits, the Big-5 factor model is probably the most popular, due to the fact that it provides a compact and complete set of traits describing human personalities. This paper presents an adversarial method that allows the generation of body postures exhibiting the characteristics of a chosen personality trait. The Big-5 and a broader model are analyzed and, in particular, we propose and analyze a technique for generating silhouettes with different levels of extroversion, as well as the aspect of a broader model corresponding to how over (or under) constrained a person is. The proposed approach can be applied in domains such as automatic character animation, marketing, and the broader field of Affective Computing.

Keywords: Adversarial Autoencoder · Personality Computing · Character Animation.

1 Introduction

Personality Computing is addressing three main problems: automatic personality recognition, perception and synthesis [1]. In this paper, a novel method is introduced that allows the synthesizing of skeletons of a specific personality trait, with the help of Generative Adversarial Networks. In particular, we are analyzing the aspects of the Big-5 model corresponding to extroversion/introversion, but also the trait corresponding to over or undercontrained personalities, stemming from the broader model explained in [2]. The two models have a direct relation which has been examined in [3]. *Openness, creativity, agreeableness, neuroticism and conscientiousness* are the traits entailed by the Big-5 model. The broader model encapsulates all personality traits under the following 3 traits: *underconstrained, resilient* and *overconstrained* [2].

The adversarial method proposed creates a mapping of human silhouettes of a specific personality trait onto a predefined probability distribution. The

architecture that is used for this is an Adversarial Autoencoder [4], which is an aggregate of an Autoencoder and a Generative Adversarial Network (GAN) [5]. The data that was used to train the proposed models is coming from the SALSA dataset [6] which involves 18 individuals, annotated on the Big-5 model. This dataset has been recorded in a poster session meeting and looks at the body postures and expressivity of the participants. The postures, being obtained in *in-the-wild* conditions, constitute a challenging benchmark used in the research community for personality analysis using computer vision techniques [6].

In the proposed work, certain personality traits are analyzed, as expressivity has been shown to be higher for them. Future work is going to focus on other traits, which through initial findings have been found to be more challenging to model using the proposed approaches. The proposed models allow the generation of new skeletons that exhibit the *style* characteristics of the personality traits involved: extroversion, introversion and the traits of over and underconstrained personalities. In this work, we are also analyzing those characteristics that transfuse the impression of a specific personality trait. Lastly, the human rated accuracy of the synthesized postures for specific personality traits is researched and compared with that corresponding to actual data from real life.

The structure of the remainder of this paper is as follows: Section 2 provides our proposed method. Section 3 covers the experiments. In Section 4, the results of the experiments are discussed. Finally, Section 5 concludes the work, and possible future work is discussed.

2   Proposed Methodology

2.1   Personality traits

Big-5 The Big-5 model, also known as the five-factor model, is a taxonomy for personality traits. It classifies personality traits over five different dimensions: extroversion, agreeableness, conscientiousness, neuroticism and openness [7]. There are various questionnaires that researchers use for self-reporting on one's personality traits [8], while a brief interpretation behind their rationale is the following: People that score high on openness are amenable to accept radical new ideas or beliefs. Conscientious people are aware of what they do and the consequences of acts. Extrovert people are rather outgoing, generally have confident behavior and tend to be talkative. Agreeable people are generally considered very co-operative. People that score high on neuroticism tend to over-think things and worry a lot [2].

Broader model Various researchers have argued that focusing on personality traits in isolation, as done in the Big-5 model, can be a limiting factor. Attempts have thus been made to propose further models that specify personality traits as a combination of the Big-5 factors, such as the broader model proposed in [2] and also used in our work, which is one of the best accepted models satisfying this, where three mutually exclusive personality traits are considered. This broader

model assigns a person only one of the three following personality traits, without any scoring assignment. Namely, these traits are: overconstrained, underconstrained and resilient [9]. Overconstrained people score high on neuroticism and conscientiousness but score low on the extroversion scale. On the contrary, underconstrained people score high on the extroversion and neuroticism scale, but score low on agreeableness and conscientiousness. Resilient people have an average score in regard to neuroticism, yet have an above average score on the four other traits that are encapsulated by the Big-5 model.
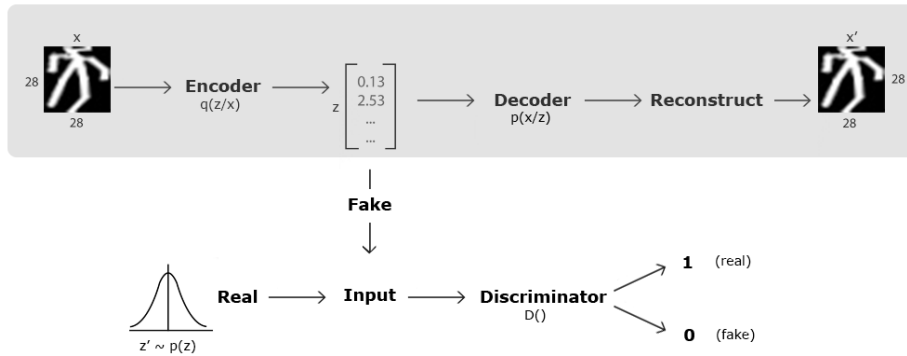

Fig. 1. Architecture of an Adversarial Autoencoder where x is the input, z the latent encoding, x' the output, z' the prior distribution and $D()$ the discriminator. The grey area shows the components of a standard autoencoder

2.2   Adversarial Autoencoder

Adversarial Autoencoders (AAEs) were introduced in 2016 by Makhazi et al. [4]. AAEs were introduced as a technique that matches a prior probability distribution to an aggregated posterior. By doing this, generating from any part of the prior distribution will satisfy the requirement to generate meaningful outputs [4]. The AAE architecture combines a typical Autoencoder with a Generative Adversarial Network (GAN). The architecture also states that a prior distribution is chosen (Figure 1). The goal is that, after training, the decoder of the AAE can generate real-looking fake samples. These fake samples are generated by passing a latent encoding, which is forced to follow the prior distribution, to the decoder.

Training of an AAE can be split into 2 phases: the reconstruction phase (Figure 2) and the regularization phase (Figure 3). During the former, only the autoencoder is considered. Training samples are fed to the encoder, which compresses them into an n-dimensional vector (n = size of latent space). This latent vector is fed to the decoder which is tasked with reconstructing the original

input as accurately as possible. To measure the quality of reconstruction, the Mean Squared Error (MSE) is used, where the decoder's output is pixel-wise compared with the encoder's input. MSE is also referred to as the reconstruction loss, or autoencoder loss.

Secondly, during the regularization phase, the encoder and the discriminator are considered. The discriminator is trained as a classifier for the encoder output and some random input. The random input is a vector that is sampled from the prior distribution. When a random input from the prior is passed to the discriminator, an output of 1 is expected. In case the input comes from the encoder output, a 0-value is expected. As a second stage in this training phase, the encoder output is connected to the discriminator. This means that the encoder output becomes the discriminator input. Additionally, the weights of the discriminator are fixed to their current values. The target of the discriminator is also fixed to 1. After this, the encoder is fed with input images. Due to the previous constraints that are in play (fixed discriminator value and fixed weights), the encoder is forced to output a latent encoding that follows the prior distribution. Backpropagation is used such that the encoder learns the correct weights regarding the prior distribution. By the end of the training process, the encoder can generate a latent code that is in par with the desired distribution. For more details, the work in [4] analyzes the training procedure analytically.

Fig. 2. First training phase: Reconstruction phase

3 Experiments

3.1 Dataset

SALSA To validate results and the use of Generative Autoencoders for personality-conditional body postures, we made use of a publicly available database, namely the SALSA dataset [6]. It consists of video-recorded data of 18 people in social settings. In this work, we made use of the sequences of the dataset corresponding to participants attending a poster session and discussing freely with each other. A typical instance of the dataset is shown in Fig. 4

The events were captured by four cameras. All of the people that were part of the event received a scoring of their personality traits. This is done according to the Big-5 model. The scores are obtained through the BFI-10 index, which is a
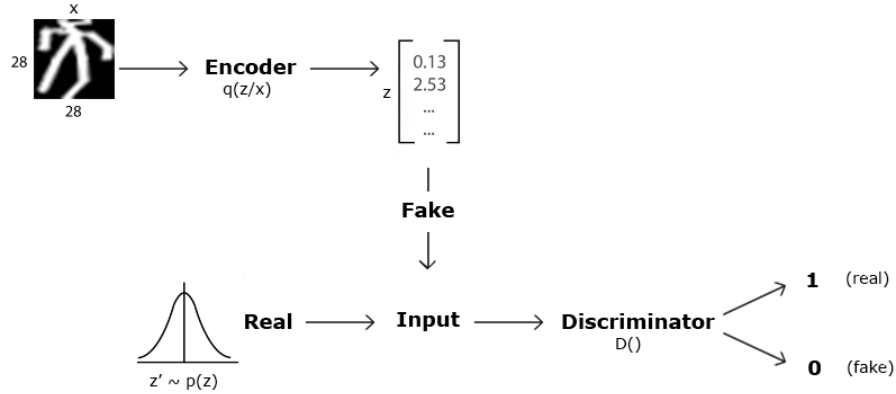
Fig. 3. Second training phase: Regularization phase

Fig. 4. Sample images of the SALSA dataset, one sample for each camera. Source: http://tev.fbk.eu/salsa

questionnaire that contains 10 different questions. The participants answer the questions with a score ranging from 1 to 7. The conversion of the questions to the scoring of the personality traits is done according to the Big-5 marker scales [10].

Cleaning and preprocessing Since this dataset is from a real situation, pre-processing was necessary. First, the video is processed such that each video frame is stored as an image. Then each person is marked with an ID and the data for each person is extracted. Next, OpenPose [11] is used to extract the joint positions for each person, per ID, in order to have a raw description of the person's body posture in each frame. Subsequently, the body joints are connected in order to complete the structural information accounting for natural-looking skeletons. Afterwards, the frame is transformed into a binary image with a black background and a white skeleton. This approach is very similar to the work done on personality recognition in nonsocial settings, proposed in [3]. Next, a bounding box, based on the coordinates extrema of the skeletons, is determined, and the

rest of the binary image is cropped out. Lastly, the image is resized to 28x28. This is a common dimension in image processing; the infamous MNIST dataset is the most characteristic example for this [12].

To account for faulty detections and tracking, a lot of image sequences where the OpenPose tracker did not deliver successful results were removed from the dataset (Fig. 5). Figure 6 shows typical examples of skeletons used in our experiments. These skeletons clearly show the body parts that we are interested in, i.e. legs, arms, head, shoulders.

As a last pre-processing step, skeletons underwent an in-plane rotation so they are all in an upright position. This was done in order to correct for the different angles of view of the cameras recording the participants. Not employing this step would result in the generation of skeletons with arbitrary and unnatural rotations.

Fig. 5. Examples of faulty skeletal images in the dataset

Fig. 6. Examples of skeletal images that are accepted

Different personality models Our approach constructs two models for each personality trait. One model is trained on all the positive examples for a trait, while the other model is trained on all the negative examples for that trait.

Regarding the extroversion and introversion traits, the median score for each personality trait is used as the threshold to determine the positive and negative examples, similar to many works proposed in the literature [1]. The final models use 900 nodes in the first layer, while having 255 nodes in the second one. This hyperparameter setup is similar to the setup in the work for personality classification in nonsocial settings described in [3]. The model was trained for 200 epochs with 500 batches per epoch and a batch-size of 100. The two models (positive and negative) that are generated from the extroversion and introversion traits, respectively, can be seen as one model for extroverts and one model

for introverts. Figures 7, 8, 9 show the loss graphs of the extroversion model. The graphs for the introversion model follow a similar trend. The autoencoder loss shows how accurately the decoder output is reconstructed compared to the original input. The discriminator loss is indicative of how often our discriminator is deceived. A high loss can indicate a discriminator that fails significantly in distinguishing between real and synthesized samples, i.e., that the generated samples are of high quality. Lastly, the generator loss is informative of how often the generated samples are identified as fake by the discriminator. Low generator loss can indicate that the generated samples are realistic.

Fig. 7. Autoencoder loss - extroversion: Shows how accurately the decoder reconstructs the original input

Fig. 8. Discriminator loss - extroversion: Shows how well the discriminator can separate real from fake examples

Fig. 9. Generator loss - extroversion: Shows how well the generator creates examples that the discriminator cannot catch

Since there exists a mapping between the Big-5 and the broader framework of [3], models were trained for personality traits of the broader model as well and, in particular, as mentioned before, the traits corresponding to over and underconstrained personalities. This broader framework does not use questionnaire scorings but, rather, it is a classifier for distinct categories [3]. Similarly to before, the models used were trained for 200 epochs with 500 batches per epoch and a batch-size of 100. As for the deep layers, both layers use 500 nodes. The size of the latent-space is 5. This setup was found by extensively trying out different hyperparameter settings and considering the results of the trained model. The most important factor taken into account was that each model, after training, was able to follow the prior distribution without having any 'gaps' into the trained distribution. In case of gaps in the trained distribution, it is possible that at generation, meaningless or blurry skeletons are generated. Secondly, the autoencoder loss was taken into account. The trained models with the aforementioned setup yielded the best results when considering the density of the distribution and the autoencoder loss after training. Figures 10-12 show the autoencoder loss, generator loss and discirminator loss for the *overconstrained* trait, respectively. Similar loss-values are obtained also for the *underconstrained* trait.
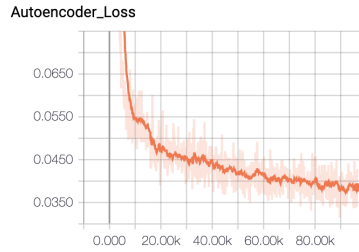

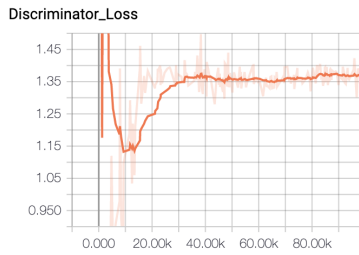**Fig. 10.** Autoencoder loss - trait for overconstrained personalities

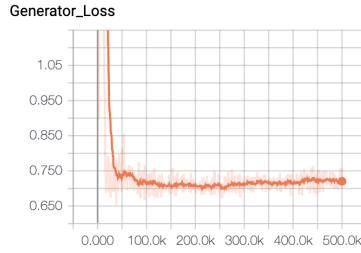**Fig. 11.** Discriminator loss - trait for overconstrained personalities

Fig. 12. Generator loss - trait for overconstrained personalities

4   Results

4.1   Human rated accuracy

The primary experiment of this paper is to determine the human rated accuracy on the generated body postures. In order to determine this, a questionnaire was administered to 33 human observers. The questionnaire investigated extroversion, introversion, overconstrained and underconstrained personality traits. Since each observer annotated 10 images per trait, the results presented in this section are based on 330 collected answers for each personality trait. To find out how well humans can recognize a personality trait from a sample coming from an actual image, human-rated accuracy was measured also for samples from the actual SALSA data set, shown on figure 13 [6], separately from synthesized body postures generated by our own models, shown on figure 14. This allowed us to assess whether mismatchings in human annotation and synthesized postures are due to the generating model proposed or pre-existing inherent difficulties in making such assessments.

Fig. 13. Randomly selected skeletons from the dataset, from left to right: extrovert, introvert, overconstrained, underconstrained

For each of the introversion and extroversion model, 330 distinct observations were collected for the synthesized postures and 330 observations for the examples from the dataset. Each picture was classified in a binary manner by the people who took the questionnaire. Binary classification is allowed here since both personality traits are mutually exclusive. Table 1 shows the human recognition rate for these traits for the samples from the dataset.

Fig. 14. Random examples of generated skeletons, from left to right: extrovert, introvert, overconstrained, underconstrained

Table 1. Confusion matrix: Extroversion trait on real, SALSA data

		Actual	
		Extrovert	Introvert
Predicted	Extrovert	209	121
	Introvert	113	217

When calculating the accuracy of the above confusion matrix, the extroversion trait is labeled correctly 63.3% of the time, whereas the introvert trait is labeled correctly 65.8% of the time.

Table 2 shows the human recognition rate on the synthesized samples. The accuracy for the synthesized examples comes out on 71.5% for the extroversion trait and 59.7% for the introversion trait. It becomes obvious that the human-rating accuracy does not differ significantly between real data and the ones our models generate.

Table 2. Confusion matrix: Extroversion trait on Synthesized data

		Actual	
		Extrovert	Introvert
Predicted	Extrovert	236	94
	Introvert	133	197

Another way of testing the human rated accuracy is by looking at the overall ratings of each image and applying a voting scheme on the annotations each sample received (i.e. what the majority of the observers per sample consider as the most probable trait for the specific sample). This approach yields a correct classification of 83.3% for the SALSA images on both extrovert and introvert, and a correct classification of 86.7% for extroverts and 70% for introverts for the synthesized examples

Similar testing was done for the overconstrained and underconstrained traits. Both methods described above are used. Again, 330 distinct annotations were

collected for the synthesized skeletons, and 330 for the samples of the SALSA dataset. Table 3 gives the confusion matrix with the results for images of the SALSA dataset.

Table 3. Confusion matrix: overconstrained / underconstrained trait on real, SALSA data

		Actual	
		Overconstrained	Underconstrained
Predicted	Overconstrained	169	161
	Underconstrained	160	170

Table 4 gives the confusion matrix for the overconstrained and underconstrained traits from the synthesized images.

Table 4. Confusion matrix: overconstrained / underconstrained trait on Synthesized data

		Actual	
		Overconstrained	Underconstrained
Predicted	Overconstrained	167	163
	Underconstrained	141	189

Lastly, the accuracy is checked when looking at the majority of the classifications. For the overconstrained trait, images of the dataset got classified correctly only for a 51.5% of the cases, similarly the synthesized samples were classified correctly for 50% of the time. The underconstrained examples from the SALSA dataset were classified correctly 51.2% of the time, and the synthesized samples 57.3% of the time. The above results show that, for human observers, there is no obvious body expressivity accounting for overconstrained or underconstrained personality types.

To make sure that there was no bias in the collected data coming from the human observers, the results of each person were individually reviewed. The amount of times they answered introvert or extrovert was collected, as well as the amount of times each person answered overconstrained or underconstrained. A paired t-test, for a confidence interval of 95%, was conducted on the results of extrovert and introvert, as well as on the results for overconstrained and underconstrained. The p-values of these tests were > 0.05, meaning that in both cases there was no statistically significant difference between the results and,

thus, there was no bias for annotating any specific trait with higher frequency than its counter-extreme.

4.2   Expressiveness of the traits

The part of the questionnaire that had annotators classify the synthesized skeletons also included a generic question of how expressive the posture is. The annotators were asked to rate the expressivity of the skeletons on a scale from 1 to 5. In this way, we were able to see if annotators perceive the expressivity between two different types of traits differently. The extroversion trait scored an average expressivity equal to 3.41/5.00 while introversion scored 2.92/5.00. For the Overconstrained trait, the average expressivity score equals 3.09/5.00 and the underconstrained trait has a scoring of 3.21/5.00. To see if there are significant differences between the expressivities of pairs of traits, a series of t-tests was conducted. The results of these controls are presented in table 5. The t-tests were performed for a confidence interval of 95%. A p-value < 0.05 means there is a significant difference between the expressiveness of the traits considered each time. This holds for all traits, except for extroversion and underconstrained, as well as for introversion and overconstrained, lastly there is also no noticeable difference between the overconstrained and underconstrained trait.

Table 5. Overview of the t-test results between the expressiveness of personality traits. The values represent the p-value and are significant when $p < .05$

	Extrovert	Introvert	Overconstrained	Underconstrained
Extrovert		0.001	0.003	0.059
Introvert	0.001		0.098	0.006
Overconstrained	0.003	0.098		0.274
Underconstrained	0.059	0.006	0.274	

4.3   Characteristics of a personality trait

In the questionnaire, a section was devoted to determining which specific part of a body posture exhibits the characteristics of a specific personality trait. People were asked to choose from a list of features which gave them the impression that a posture was of a certain personality trait. The options they had to choose from were the following: Head pose, Arm pose, Leg pose, Arm spatial extent

This series of questions allowed multiple answers. In total, there were 330 different provided annotations. The results can be seen in table 6 for the *extroversion-introversion* pair of traits. Table 7 shows the results for the broader model. The results are normalized to fit in the 0 - 1 interval. The value represents the percentage of how often a characteristic was selected per personality trait.

Table 6. Results of the characteristics that transfuse an impression of a personality trait

Trait Properties				
	Head	Arms	Legs	SE[1]
Extrovert	0.233	0.588	0.427	0.324
Introvert	0.361	0.476	0.333	0.17

Table 7. Results of the characteristics that transfuse an impression of a personality trait

Trait Properties				
	Head	Arms	Legs	SE
Underconstrained	0.279	0.576	0.464	0.27
Overconstrained	0.349	0.54	0.333	0.252

From the above tables, it becomes evident that the positioning of the arms has the biggest influence on how people perceive a personality trait. Secondly, the positioning of the legs, or more generally, a persons' stance seems to have a big impact as well.

5   Discussion

Based on the above results, it can be assumed that the synthesized extroversion and introversion traits are more easily recognizable by humans, while the human recognition rate for synthesized samples is not far from that of the samples corresponding to real data, showing that our proposed architecture can capture those features that account for different postures accounting for the extrema traits of extroversion and introversion. On the contrary, synthesized overconstrained and underconstrained personality traits do not perform well, while similar is the case also for the real samples corresponding to these traits. This shows that there is a high inherent challenge in associating human postures with the trait of over or underconstrained, contrary to the case of extro/introversion. The results of the classified skeletons were verified with a paired t-test to ensure there was no bias in the data.

---

[1] Spatial Extent of the arms

Secondly, the expressiveness of the synthesized skeletons was researched. A significant difference exists between all combinations of traits, except for the extroversion and underconstrained traits, the introversion and overconstrained traits, as well as between the overconstrained and underconstrained traits. According to the personality models, underconstrained people do score high on the extroversion scale, and this can be an explanation for not observing a significant difference between these two traits. Other than extroversion, there is no significant difference between the introverts and overconstrained people. This can also be explained by the relation between these two personality traits, where overconstrained people score high on the introversion scale. Lastly, there is no significant difference between the overconstrained and underconstrained postures, and an assumption here can be that both traits score similar on the neuroticism scale; however except for neuroticism, their scoring on other relatable traits are opposites. From the questionnaire it turned out that the positioning of the arms and legs have the biggest influence on how people perceive a personality type of an image they are presented with.

Finally, a current issue is that some of the generated skeletons tend to be noisy or blurry around the edges. Post-processing techniques might be sufficient for correcting this, but have not yet been tried. Another option for future research could be found in the recent development of "super resolution" [13]. This method reduces the noise in images drastically. It can also generate high resolution images from a low resolution image.

6   Conclusion and Future Works

This paper describes a method that maps personality-conditional body postures onto prior distributions. The proposed approach can successfully map and generate skeletons that exhibit the characteristics of introvert and extrovert personalities. Future work will focus on generative models for the rest of the traits of both the Big-5 framework and the broader model used in this paper. Secondly, the human rated accuracies on the analyzed personality traits of the Big-5 model were close to each other, for the samples from both the SALSA dataset and the synthesized skeletons, showing the potential of our model being employed for generating postures perceived as extrovert or introvert to the same extent as real ones. For the broader model, two of the three traits involved were synthesized in this work, however, the human rated accuracy for these body postures is very weak, also in the case of deducing one's personality from real postures.

The generator often synthesizes skeletons that look very similar. It could be an interesting future development to prevent the generator from synthesizing very similar skeletons. The goal of this would be to promote the generation of more distinct skeletons, potentially using combinations of the different characteristics of a specific personality trait. Moreover, future work should focus on a dedicated, large-scale analysis of human-perceived synthesis of personality traits, extending our current human ratings to much larger populations, while taking motion into account should also be considered.

References

1. Vinciarelli, A., Mohammadi, G. (2014). A survey of personality computing. IEEE Transactions on Affective Computing, 5(3), 273-291.
2. Sava, F. A., Popa, R. I. (2011). Personality types based on the Big Five model. A cluster analysis over the Romanian population. Cognitie, Creier, Comportament/Cognition, Brain, Behavior, 15(3).
3. Dotti, D., Popa, M., Asteriadis, S. (2018). Behavior and Personality Analysis in a nonsocial context Dataset. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 2354-2362).
4. Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I., Frey, B. (2015). Adversarial autoencoders. arXiv preprint arXiv:1511.05644.
5. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S. Bengio, Y. (2014). Generative adversarial nets. In Advances in neural information processing systems (pp. 2672-2680).
6. Alameda-Pineda, X., Staiano, J., Subramanian, R., Batrinca, L., Ricci, E., Lepri, B., Sebe, N. (2015). Salsa: A novel dataset for multimodal group behavior analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(8), 1707-1720.
7. McCrae, R. R., John, O. P. (1992). An introduction to the fivefactor model and its applications. Journal of personality, 60(2), 175-215.
8. Rammstedt, B., John, O. P. (2007). 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, 41(1), 203-212.
9. Asendorpf, J. B., Borkenau, P., Ostendorf, F., Van Aken, M. A. (2001). Carving personality description at its joints: Confirmation of three replicable personality prototypes for both children and adults. European Journal of Personality, 15(3), 169-198.
10. Perugini, M., Di Blas, L. (2002). The Big Five Marker Scales (BFMS) and the Italian AB5C taxonomy: Analyses from an emic-etic perspective. Hogrefe Huber Publishers.
11. Cao, Z., Hidalgo, G., Simon, T., Wei, S. E., Sheikh, Y. (2018). OpenPose: real-time multi-person 2D pose estimation using Part Affinity Fields. arXiv preprint arXiv:1812.08008.
12. Deng, L. (2012). The MNIST database of handwritten digit images for machine learning research [best of the web]. IEEE Signal Processing Magazine, 29(6), 141-142.
13. Ledig, C., Theis, L., Huszr, F., Caballero, J., Cunningham, A., Acosta, A., Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4681-4690).