

# Explaining the Outputs of Convolutional Neural Network - Recurrent Neural Network (CNN-RNN) based Apparent Personality Detection Models using the Class Activation Maps

WMKS Ilmini<sup>1</sup>, TGI Fernando<sup>2</sup>

Faculty of Graduate Studies, University of Sri Jayewardenepura, Nugegoda, Sri Lanka<sup>1</sup>  
Intelligent Research Laboratory-Faculty of Computing, General Sir John Kotelawala Defence University,  
Rathmalana, Sri Lanka<sup>1</sup>

Department of Computer Science-Faculty of Applied Sciences,  
University of Sri Jayewardenepura, Nugegoda, Sri Lanka<sup>2</sup>

**Abstract**—This study aims to use the Class Activation Map (CAM) visualisation technique to understand the outputs of apparent personality detection models based on a combination of Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN). The ChaLearn Looking at People First Impression (CVPR'17) dataset is used for experimentation in this study. The dataset consists of short video clips labelled with the Big Five personality traits. Two deep learning models were designed to predict apparent personality with VGG19 and ResNet152 as base models. Then the models were trained using the raw frames extracted from the videos. The highest accurate models from each architecture were chosen for feature visualisation. The test dataset of the CVPR'17 dataset is used for feature visualisation. To identify the feature's contribution to the network's output, the CAM XAI technique was applied to the test dataset and calculated the heatmap. Next, the bitwise intersection between the heatmap and background removed frames was measured to identify how much features from the human body (including facial and non-facial data) affected the network output. The findings revealed that nearly 35%-40% of human data contributed to the output of both models. Additionally, after analysing the heatmap with high-intensity pixels, the ResNet152 model was found to identify more human-related data than the VGG19 model, achieving scores of 46%-51%. The two models have different behaviour in identifying the key features which influence the output of the models based on the input.

**Keywords**—Apparent personality detection (APD); convolutional neural network based recurrent neural network (CNN-RNN); class activation map (CAM); explainable AI (XAI)

## I. INTRODUCTION

Explainable AI (XAI) has gained attention in machine learning as it is crucial to comprehend the behaviour of these models, that is, how these models generate their outputs. These techniques can be used to gain a deeper understanding of the inner workings of a model and can help improve the trust and adoption of AI systems in various applications. Artificial neural networks and deep learning methods are often considered as black boxes, as the inner workings of these techniques and how they produce output based on input are not fully understood. Therefore, researchers tend to explore

techniques to make these into glass boxes, that is, to understand how input features contribute to the output. According to the literature [1], these techniques are divided into different categories.

### • Local Vs Global

- **Local:** Explain how a model makes a single prediction and evaluates its performance on a given set of examples.
- **Global:** Global techniques for explaining a model do not require a specific set of example data and instead provide an overall understanding of how the model works. These techniques can include analysing the model's architecture and studying the relationships between the model's parameters.

### • Model Specific Vs Model Agnostic

- **Model Specific:** These techniques only apply to a single model or a group of specific models.
- **Model Agnostic:** These techniques can be applied to any model to explain the model's predictions.

XAI techniques are grouped into different clusters based on the type of data, the purpose of interpretability, and the flow of interpretation signals, in addition to the two primary categories mentioned above.

The Saliency map is the oldest and most commonly used technique to explain convolutional neural network (CNN) predictions. The saliency map specifies the pixels that activate a particular layer in the network. The literature discloses three main approaches: Deconvolutional Network [2], Backpropagation [3], and Guided Backpropagation [4]. Table I summarises the most popular XAI techniques.

Researchers discovered various XAI methods to understand the deep learning model predictions in addition to the methods mentioned in Table I.

Apparent Personality Detection (APD) based on a person's appearance is a trending research topic in affective computing because apparent personality is helpful in various applications. A few of those applications are listed below:

- **Job Screening:** From the past [10] to the present [11], [12], psychological researchers have tended to find a relationship between personality and job performance. Barrick et al. [11] identified a relationship between Extraversion and Conscientiousness personality traits in the ratings of sales representatives. Inceoglu and Warr [13] conducted a study to reveal the relationships between job engagement and personality. They concluded that there is a relationship between Extraversion, Conscientiousness and Emotional Stability. Hence, the different personality traits contribute to job roles, performance, and satisfaction. Such as, a team leader should have a high level of Extraversion and Conscientiousness and a low level of Neuroticism.
- **Recommendation Systems:** Dhelim et al. [14] discuss the need for personality-aware recommendation systems. Hence, people with the same characteristics act in the same way. It is easy to recommend products or solutions if the customer's personality is known. The authors also mentioned that personality-aware recommendation systems are better when dealing with cold start and data sparsity issues than traditional recommendation techniques.
- **Social Robotics:** A study by Lee et al. [15] found that if the robot's personality is similar to the user's personality, users enjoyed dealing with the robot. Kirby et al. [16] highlight the importance of affective-social robots with emotions and apparent personalities. The robot can identify the user's state and act accordingly. It is essential to consider this when designing social robotics.
- **Personal Assistants:** There are many personal assistants available nowadays, including Apple Siri, Microsoft Cortana, Google Assistant, and Huawei Celia. These personal assistants can be enhanced by adding the automatic personality detection feature, which leads to higher user interaction with personal assistants.
- **Animation Movies:** Designing an animation-movie character is challenging since it should reflect the character's qualities, including personality [17]. Identification of the facial features which contribute to the different personality traits will be beneficial in this field to improve the outcomes.
- **Health Care and Counselling:** In psychology, researchers are researching the relationship between personality and mental health, personality and physical

health and personality and illness. Smith and MacKenzie [18] discuss how personality traits (such as neuroticism) affect a human's health. Hence psychology research proved that our mental and physical health is affected by personality. It will be beneficial to identify personality for early treatment processes and personalised counselling plans based on the personality.

- **Criminology:** Reid [19] explained the connection between personality and crime. Hence, with better personality prediction solutions, authorities can identify and prevent criminal activities.
- **Education and Personalised Learning:** Salazar et al. [20] highlight the importance of having an affective recommendations system in the education field. Moreover, they mentioned that it is vital to change the content based on the learning style, emotions and personality.

According to the review study conducted by [21], psychological studies, political forecasting, forensic, and word polarity detection can also be enhanced by automatic personality detection.

Thus, an individual's apparent personality can be used in different domains to improve performance and effectiveness. Researchers introduced deep learning solutions, including convolutional neural networks and recurrent neural network architectures, to measure apparent personality. After achieving higher accurate predictions by APD deep learning models, researchers tend to find how these models produce the output for given input features using XAI techniques. The other purpose of applying XAI in APD models is to identify prominent facial and non-facial features that affect the personality, which is more important to improve the trust and adoption of AI systems in the above mentioned applications. All works performed in this area used ChaLearn Looking At People First Impression V2 (CVPR'17) dataset [22]. This is the only dataset publicly available with labelled Big Five Personality traits [23].

Zhang and co-workers [24] applied a heatmap feature visualisation technique to visualise the features affecting the APD. They have used different deep learning architectures such as ResNet, DAN, and DAN+. Their study results convey that different models focused on different features of the face, including facial and non-facial data, including background data. Ventura and co-workers [25] conducted a quantitative study to identify prominent facial features and emotions that influence APD.

TABLE I. MOST POPULAR XAI TECHNIQUES APPLIED IN THE FIELD OF CNN

Technique	Year	Local/ Global	Model Specific/ Agnostic	Description
Deconvolutional Network [2]	2013	Local	Specific	Deconvolutional networks work as the inverse of convolution, pooling (unpooling), and inverse of ReLU. This technique recognises the features activated by the immediate layer for the given input. It reconstructs the input from the activations of the layer.
Backpropagation [3]	2014	Local	Specific	For a given input, calculate the gradients concerning the network parameters. This technique highlights the pixel space based on the gradients they receive, which implies the contribution of these pixels to the final output.
Guided Backpropagation [4]	2015	Local	Specific	Guided Backpropagation is a combination of a deconvolutional network and the backpropagation technique. This technique identifies the essential features based on the reconstruction signal's negative values (deconvolutional) and negative values of the input in the forward pass (backpropagation).
CAM [5]	2015	Local	Specific	Class Activation Map (CAM) detects different regions contributing to a given class score. The last fully connected layers are replaced by a global average pooling (GAP) layer, which averages the activations of feature maps. The GAP layer produces a vector, then calculate the weighted sum of the vector's components and sends it to the SoftMax layer. The calculated weighted values help identify the essential features that activate each convolutional feature map by projecting them back.
Deep LIFT [6]	2019	Local	Specific	This technique calculates the activation feature map by multiplying the input with the measured gradients for the given input with a class of interest.
Grad-CAM [7]	2020	Local	Specific	This is a more flexible version than CAM because this produces feature activation with fully connected layers. When the class of interest and input is produced to the network, the network calculates the gradient flow into the final convolutional layer.
Guided Grad-CAM [7]	2020	Local	Specific	Since the Grad-CAM cannot highlight fine-grained regions, the same authors suggest combining the Grad-CAM and Guided Backpropagation techniques to obtain the Guided Grad-CAM.
LIME [8]	2016	Local	Agnostic	Local Interpretable Model-Agnostic Explanations (LIME) manipulate the input data by creating a set of artificial data. These artificial data consist of part of the original input data. The artificial data is then introduced to the model and classified into different categories. Hence, the presence or absence of certain input parts can decide the contribution to the model's output.
SHAP [9]	2017	Local and Global	Agnostic	Shapley Additive Explanations (SHAP) is based on the Shapley values used in game theory. Shapley values are vastly applied in the cooperative game theory to find each player's contribution/ importance. The same theory is applied in the XAI to identify feature importance for the final output.

They applied CAM and Action Unit (AU) [26]. CAM is applied to find the discriminative regions in the scene data. CAM results convey that the facial regions, such as the eye, nose, and mouth areas, contribute to the final prediction. From the Action Coding System, 17 AU was applied to find the influence of emotions in APD. The results indicate that few AUs affected personality detection. They concluded these results with 50 images extracted with the highest personality scores.

Wei et al. [27] applied feature map visualisation to the models they trained to predict the apparent personality. Results show that ResNet identified the facial region as the primary contributor, while DAN and DAN+ activate background data rather than facial data. However, with plain background data, DAN and DAN+ identify facial data, while ResNet fails to identify facial data as primary contributing features. They summarised the model interpretability techniques results with 12 randomly selected images.

Yang and Glaser [28] used saliency map model interpretability techniques to interpret the APD models' outputs. They also concluded that ResNet pre-trained model-based APD architecture could identify facial features as primary contributors. Li et al. [29] calculated heatmap on scene data to identify the most contributing features using the Seaborn Python library [30]. Their findings revealed that critical facial features such as the eye, nose, and mouth

contribute to APD. However, non-facial features such as clothing and furnishing affect the APD model's output. They conducted a quantitative study by considering the face area and heatmap of contributing features and concluded that 73.96% of the highlighted points are face key points (eye, nose, and mouth). They used 32 frames for the experiment from each video from the test dataset of CVPR'17 [22].

A summary of the works conducted in this area used heatmap visualisation techniques such as saliency map techniques to interpret the prediction of APD models. Most of these works concluded that facial regions and non-facial data contribute to the output. A majority of these techniques tend to interpret the CNN architectures' output. These researchers used different pre-trained models in the development and various XAI techniques and concluded that different architectures tend to highlight different areas. Less attention has been paid to work focusing on describing the outputs of Convolutional Neural Networks based Recurrent Neural Network (CNN-RNN) architecture and work on conducting a quantitative study to prove the findings.

#### A. Contribution

Contributions of the work to the APD area are as follows:

- 1) Prior works mainly focused on explaining the CNN-based APD models. This work focused on CNN-RNN models.

2) A quantitative study is conducted to identify primary contributing features for the CNN-RNN-based APD model's output.

The primary aim of this work is to explain the CNN-RNN-based APD models using the CAM technique.

The rest of the paper is organised as follows: Section two discusses the Methodology, Section three contains the Results and Discussion, and Section four contains the Conclusion.

## II. METHODOLOGY

This section includes the methodology followed in this study to explain the predictions of the APD models. Fig. 1 shows the overall methodology followed to identify how the human data (facial and non-facial data excluding background) affected apparent personality.

According to Fig. 1, first, the dataset is pre-processed by dividing it into raw frames. Then the extracted frames were used to train, validate, and test the model. After completing the model development, the CAM visualisation technique was applied to the test dataset. The bitwise intersection between the heatmap and the background removed raw frames was calculated to clarify the facial and non-facial (non-background) features that contributed to the network's output.

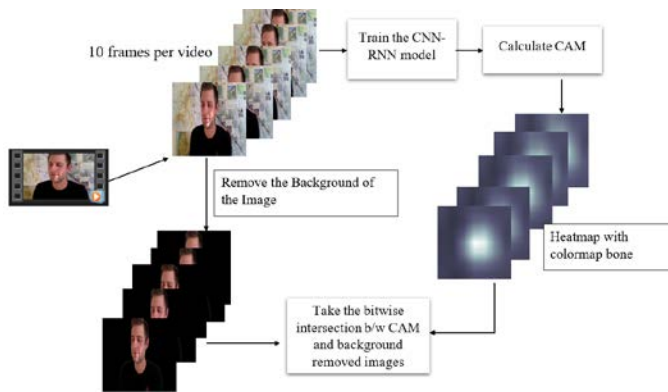


Fig. 1. Overall methodology

### A. Preparation of Data

The experiment used the CVPR'17 [22] dataset, which consists of videos of people facing the camera. These participants are from different nations, ages, and ethnicities. The dataset initially consisted of 3,000 videos which were again processed in 10,000 clips. The training, validation, and test datasets include 6,000, 2,000, and 2,000 video clips. Each video clip is labelled with Big-Five traits ranging from 0 to 1. For model development, ten frames were extracted from each video.

### B. Network Architecture

In CNN-RNN architecture development, the CNN part was developed using pre-trained deep learning models, trained initially on the ImageNet Classification problem (ILSVRC). Two deep learning architectures were designed, developed, and tested to compare the XAI technique findings. VGG19 [31] model is used for the first model, and the second, the

ResNet152 [31] model, is used for the CNN branch. These two models were selected because these are the most common pre-trained models used in several previous works [24], [27], [29]. RNN branch consists of one Gated Recurrent Units (GRU) layer to capture the temporal information, and Fig. 2 illustrates the network's architecture.

### C. Network Parameters

Following are the network parameters used in the current study, finalised after a few experiments conducted with the dataset.

- Batch size = 4
- Early Stop counter: 20
- Maximum number of epochs = 200
- Optimiser: Adam
- Learning rate =  $1 \times e^{-5}$
- Loss: Mean Absolute Error

All experiments were conducted on a precision server with Nvidia RTX 3090 24 GB.

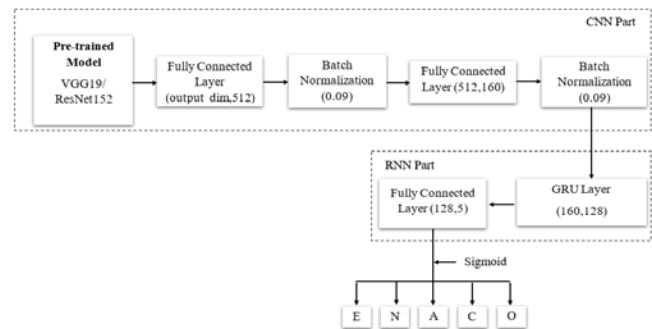


Fig. 2. Deep CNN-RNN network architecture

### D. Visualisation

This study followed the following steps to determine which features (human or background) mainly affected personality prediction in the CNN-RNN deep learning models.

**Step 1:** Removed background data from the raw frames extracted from the video clips (10 frames for each video). Python library Rembg [32] was used to detect human beings from the raw frames. This library uses U2-Net [33] deep learning architecture to detect an object.

$P_{human}$  = Pixels that correspond to the human detected from the raw frames

**Step 2:** Calculated the bitwise intersection between heatmaps and the output of step 1.

Instead of using COLORMAP\_JET [34], the most popular colour map for feature importance visualisation, we used COLORMAP\_BONE [34]. COLORMAP\_BONE, as seen in Fig. 3, uses black and white to represent low and high intensities in pixels. Moderate intensities receive grey colour (in between black and white). Thus, it is more convenient to

identify which features affect more to the output with different intensities.

$$I_1 = P_{human} \cap P_{heatmap_1} \quad (1)$$

$P_{heatmap_1}$  = Pixels which were highlighted by the CAM visualisation technique

$$I_2 = P_{human} \cap P_{heatmap_2} \quad (2)$$

$P_{heatmap_2}$  =  $P_{heatmap}$  pixel values (R, G, B) greater than or equal to 100 (higher intensities)

**Step 3:** Calculated fractions  $f_1$  and  $f_2$ :

$f_1$ : Pixels highlighted by CAM and belongs to the area where the human being exists in the frame / the pixels highlighted by CAM

$f_2$ : Pixels highlighted by CAM with high intensity and belongs to the area where the human being exists in the frame / the pixels highlighted by CAM with high intensity

$$f_1 = \frac{I_1}{P_{heatmap_1}} \times 100 \% \quad (3)$$

$$f_2 = \frac{I_2}{P_{heatmap_2}} \times 100 \% \quad (4)$$

**Step 4:** Followed the above steps for all video files in the test dataset; Measured the average of  $f_1$  and  $f_2$ .

$$F_1 = \frac{\sum_{i=0}^n f_1}{n} \quad (5)$$

$$F_2 = \frac{\sum_{i=0}^n f_2}{n} \quad (6)$$

where  $n = 2000$  (size of the test dataset)

**Step 5:** Repeated the same process for all personality traits.



Fig. 3. Opencv colormaps

### III. RESULTS AND DISCUSSION

The models were trained ten times, and the highest accurate model was selected for feature visualisation. Table II summarises the highest accuracy of each model (VGG19-based CNN-RNN model) and Table III (ResNet152-based CNN-RNN model). The accuracy of the model is calculated using the following equation:

$$\text{Accuracy} = 1 - \frac{1}{N} \sum_{i=1}^N |\text{target}_{ij} - \text{output}_{ij}| \quad (7)$$

$N$  = number of videos, the target is the respective ground-truth value, and output is the predicted value from the model for a given video.

TABLE II. VGG19-BASED MODEL ACCURACY

Big Five Personality Trait	Accuracy
E	90.23%
N	89.95%
A	90.52%
C	90.04%
O	90.16%
Mean Accuracy	90.18%

TABLE III. RESNET152-BASED MODEL ACCURACY

Big Five Personality Trait	Accuracy
E	90.53%
N	90.20%
A	90.21%
C	91.15%
O	90.42%
Mean Accuracy	90.50%

The ResNet152-based model outperforms the VGG19-based model by achieving approximately 90% accuracy for all the traits. While VGG19 based model achieved more than 90% accuracy for all the traits except for neuroticism.

#### A. Visualisation Techniques Results

As mentioned in the methodology section, we calculated the  $F_1$  and  $F_2$  values for all five personality traits with two architectures.

TABLE IV. VGG-19 BASED MODEL F SCORES

Big Five Personality Trait	$F_1$ score	$F_2$ score
E	35.36%	35.76%
N	36.19%	35.40%
A	36.19%	35.40%
C	36.19%	35.40%
O	36.19%	54.00%

Table IV conveys that human data (excluding background) affect personality prediction by 35% - 36%, with  $F_1$  score and  $F_2$  score, except for the Openness trait, which is 54% for  $F_2$ . Hence, with VGG19 openness trait is more influenced by human data with high intensity.

Table V conveys that nearly 35% - 38% ( $F_1$ ) of the facial and non-facial data (excluding background) affected the personality prediction. Furthermore, with  $F_2$  scores, it is 36% to 51%. Since the  $F_2$  scores were calculated using pixels with high intensities in the heatmap, and we can conclude that ResNet152 identified more human data than VGG19 for Extraversion, Openness, Neuroticism, and Conscientiousness traits. In both architectures, the Agreeableness trait is more

affected by background data than other traits (Tables IV and V). In the ResNet152-based model, Extraversion and Openness traits were less affected by background information than other traits (Table V).

TABLE V. RESNET152-BASED MODEL F SCORE VALUE

Big Five Personality Trait	F score	F2 score
E	38.19%	51.44%
N	36.25%	46.95%
A	32.52%	36.08%
C	35.58%	44.14%
O	37.37%	47.68%

Tables IV and V express that the human data (facial and non-background data) affected the personality prediction by nearly 40%, implying that the image's background affected the apparent personality by almost 60%, with  $F_1$  scores. Nevertheless,  $F_2$  scores confirm that high intensities (heatmap) were allocated to human data with ResNet152. The previous works to explain the outputs of the CNN-based APD models also concluded that facial, non-facial, and background data affected the prediction. Zhang and co-workers [24] concluded that different features affect different CNN models designed to predict the apparent personality. As per their demonstration, background data are highlighted as features contributing more to the model prediction. Wei and co-workers [27] also concluded that different models highlighted different features. The models they designed using ResNet and VGGFace-based architectures highlighted different image regions of the input image. Also, they concluded that VGGFace-based architecture is more prone to background data. The current work's quantitative results also convey that models identified different features from the input. Furthermore, ResNet152 is more towards human data rather than the background.

#### IV. CONCLUSION

The primary goal of the current study is to explain the output of the CNN-RNN-based APD models using the CAM as the XAI technique. The results convinced that the models' output is based on the background rather than non-background data (human data, including facial and non-facial data). Usually, the human data (facial and non-facial data excluding background) affects the personality prediction more than the background. However, the findings imply a different conclusion. Even past researchers highlighted this fact with various XAI techniques for CNN-based APD. Hence, the current study with the CNN-RNN APD model also concludes that the background is more influential for APD than human data with the CAM visualisation technique. Also, the models acted differently in the current study because they produced different  $F_1$  and  $F_2$  scores. Furthermore, for Extraversion, Openness, Neuroticism, and Conscientiousness ResNet152 based CNN-RNN models recorded higher  $F_2$  values than  $F_1$ , which implies that more contributing features are from human data. The study's conclusions are derived from an assessment of the deep learning architectures employed and the efficacy of the background removal procedure.

#### ACKNOWLEDGMENT

Our special thanks to the Faculty of Computing of General Sir John Kotelawala Defence University for providing access to servers to experiment.

#### REFERENCES

- [1] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, "Explainable AI: A Review of Machine Learning Interpretability Methods," *Entropy*, vol. 23, no. 1, p. 18, Dec. 2020, doi: 10.3390/e23010018.
- [2] M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," arXiv:1311.2901 [cs], Nov. 2013, Accessed: Sep. 20, 2020. [Online]. Available: <http://arxiv.org/abs/1311.2901>
- [3] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps," arXiv:1312.6034 [cs], Apr. 2014, Accessed: Sep. 19, 2020. [Online]. Available: <http://arxiv.org/abs/1312.6034>
- [4] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for Simplicity: The All Convolutional Net," arXiv:1412.6806 [cs], Apr. 2015, Accessed: Sep. 24, 2020. [Online]. Available: <http://arxiv.org/abs/1412.6806>
- [5] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning Deep Features for Discriminative Localization," arXiv:1512.04150 [cs], Dec. 2015, Accessed: Sep. 24, 2020. [Online]. Available: <http://arxiv.org/abs/1512.04150>
- [6] A. Shrikumar, P. Greenside, and A. Kundaje, "Learning Important Features Through Propagating Activation Differences," arXiv:1704.02685 [cs], Oct. 2019, Accessed: Feb. 04, 2021. [Online]. Available: <http://arxiv.org/abs/1704.02685>
- [7] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," *Int J Comput Vis*, vol. 128, no. 2, pp. 336–359, Feb. 2020, doi: 10.1007/s11263-019-01228-7.
- [8] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You?: Explaining the Predictions of Any Classifier," arXiv:1602.04938 [cs, stat], Aug. 2016, Accessed: Jan. 17, 2021. [Online]. Available: <http://arxiv.org/abs/1602.04938>
- [9] S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," arXiv:1705.07874 [cs, stat], Nov. 2017, Accessed: Feb. 05, 2021. [Online]. Available: <http://arxiv.org/abs/1705.07874>
- [10] D. V. Day and S. B. Silverman, "PERSONALITY AND JOB PERFORMANCE: EVIDENCE OF INCREMENTAL VALIDITY," *Personnel Psychology*, vol. 42, no. 1, pp. 25–36, Mar. 1989, doi: 10.1111/j.1744-6570.1989.tb01549.x.
- [11] M. R. Barrick, G. L. Stewart, and M. Piotrowski, "Personality and job performance: Test of the mediating effects of motivation among sales representatives.," *Journal of Applied Psychology*, vol. 87, no. 1, pp. 43–51, 2002, doi: 10.1037/0021-9010.87.1.43.
- [12] G. M. Hurtz and J. J. Donovan, "Personality and Job Performance: The Big Five Revisited," 2000.
- [13] I. Inceoglu and P. Warr, "Personality and job engagement," *Journal of Personnel Psychology*, vol. 10, pp. 177–181, 2011, doi: 10.1027/1866-5888/a000045.
- [14] S. Dhelim, N. Aung, M. A. Bouras, H. Ning, and E. Cambria, "A survey on personality-aware recommendation systems," *Artif Intell Rev*, vol. 55, no. 3, pp. 2409–2454, Mar. 2022, doi: 10.1007/s10462-021-10063-7.
- [15] K. M. Lee, W. Peng, S.-A. Jin, and C. Yan, "Can Robots Manifest Personality?: An Empirical Test of Personality Recognition, Social Responses, and Social Presence in Human-Robot Interaction," *Journal of Communication*, vol. 56, no. 4, pp. 754–772, Dec. 2006, doi: 10.1111/j.1460-2466.2006.00318.x.
- [16] R. Kirby, J. Forlizzi, and R. Simmons, "Affective social robots," *Robotics and Autonomous Systems*, vol. 58, no. 3, pp. 322–332, Mar. 2010, doi: 10.1016/j.robot.2009.09.015.
- [17] M. S. Juhan and N. Ismail, "Character Design towards Narrative Believability of Boboiboy in the Malaysian Animated Feature Film Boboiboy: The Movie (2016)," in *Social Sciences*, 2016, p. 10.

- [18] T. W. Smith and J. MacKenzie, "Personality and Risk of Physical Illness," *Annual Review of Clinical Psychology*, vol. 2, no. 1, pp. 435–467, 2006, doi: 10.1146/annurev.clinpsy.2.022305.095257.
- [19] J. A. Reid, "Crime and Personality: Personality Theory and Criminality Examined," *Inquiries Journal*, vol. 3, no. 01, 2011, Accessed: Sep. 17, 2020. [Online]. Available: <http://www.inquiriesjournal.com/articles/1690/crime-and-personality-personality-theory-and-criminality-examined>
- [20] C. Salazar, J. Aguilar, J. Monsalve-Pulido, and E. Montoya, "Affective recommender systems in the educational field. A systematic literature review," *Computer Science Review*, vol. 40, p. 100377, May 2021, doi: 10.1016/j.cosrev.2021.100377.
- [21] Y. Mehta, N. Majumder, A. Gelbukh, and E. Cambria, "Recent Trends in Deep Learning Based Personality Detection," arXiv:1908.03628 [cs], Aug. 2019, Accessed: Jun. 05, 2020. [Online]. Available: <http://arxiv.org/abs/1908.03628>
- [22] V. Ponce-López et al., "ChaLearn LAP 2016: First Round Challenge on First Impressions - Dataset and Results," in *Computer Vision – ECCV 2016 Workshops*, Cham, 2016, pp. 400–418. doi: 10.1007/978-3-319-49409-8\_32.
- [23] O. P. John and S. Srivastava, *The Big-Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives*. 1999.
- [24] C.-L. Zhang, H. Zhang, X.-S. Wei, and J. Wu, "Deep Bimodal Regression for Apparent Personality Analysis," in *Computer Vision – ECCV 2016 Workshops*, Cham, 2016, pp. 311–324. doi: 10.1007/978-3-319-49409-8\_25.
- [25] C. Ventura, D. Masip, and A. Lapedriza, "Interpreting CNN Models for Apparent Personality Trait Regression," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, HI, USA, Jul. 2017, pp. 1705–1713. doi: 10.1109/CVPRW.2017.217.
- [26] P. Ekman, W. Friesen, and S. Ancoli, "Facial-Sign-Of-Emotional-Experience.pdf," *Journal of Personality and Social Psychology*, vol. 39, pp. 1125–1134, 1980.
- [27] X.-S. Wei, C.-L. Zhang, H. Zhang, and J. Wu, "Deep Bimodal Regression of Apparent Personality Traits from Short Video Sequences," *IEEE Trans. Affective Comput.*, vol. 9, no. 3, pp. 303–315, Jul. 2018, doi: 10.1109/TAFFC.2017.2762299.
- [28] K. Yang and N. Glaser, "Prediction of Personality First Impressions With Deep Bimodal LSTM," *CS 231N Spring 2017*, p. 10, 2017.
- [29] Y. Li et al., "CR-Net: A Deep Classification-Regression Network for Multimodal Apparent Personality Analysis," *Int J Comput Vis*, vol. 128, no. 12, pp. 2763–2780, Dec. 2020, doi: 10.1007/s11263-020-01309-y.
- [30] M. Waskom, "seaborn: statistical data visualization," *JOSS*, vol. 6, no. 60, p. 3021, Apr. 2021, doi: 10.21105/joss.03021.
- [31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.
- [32] D. Gatis, "Rembg." Jan. 27, 2022. Accessed: Jan. 27, 2022. [Online]. Available: <https://github.com/danielgatis/rembg>
- [33] X. Qin, Z. Zhang, C. Huang, M. Dehghan, O. R. Zaiane, and M. Jagersand, "U<sup>2</sup>-Net: Going Deeper with Nested U-Structure for Salient Object Detection," *Pattern Recognition*, vol. 106, p. 107404, Oct. 2020, doi: 10.1016/j.patcog.2020.107404.
- [34] "OpenCV: ColorMaps in OpenCV." [https://docs.opencv.org/3.4/d3/d50/group\\_imgproc\\_colormap.html](https://docs.opencv.org/3.4/d3/d50/group_imgproc_colormap.html) (accessed Jan. 15, 2023).