Skeleton-based Personality Recognition using Laban Movement Analysis

Ziya Erkoç[*†] ZIYA.ERKOC@BILKENT.EDU.TR
Serkan Demirci[†] SERKAN.DEMIRCI@BILKENT.EDU.TR
Sinan Sonlu SINAN.SONLU@BILKENT.EDU.TR
Uğur Güdükbay GUDUKBAY@CS.BILKENT.EDU.TR
Department of Computer Engineering, Bilkent University, Ankara 06800, Turkey

Abstract

Personality is expressed through multiple behavioral elements, including body movement. Using a feature transformation based on Laban Movement Analysis, we present a model for estimating individuals' Big Five personality traits. Our approach achieves higher performance than other methods without exposing image-level information to the network, which otherwise can leave the system susceptible to bias and result in ethical issues. With the ever-increasing role of computers in our daily lives, human-computer interaction and human understanding have become significant. Our system enables better human understanding for intelligent agents and personal assistants through personality estimation. We utilize Graph Convolutional Networks, commonly used for action recognition for this task.
Keywords: Personality Recognition, Laban Movement Analysis, Deep Learning

1. Introduction

From personal assistants to intelligent home appliances, accurate human-computer interaction is essential for many applications. Commanding the household's service robot through natural-language speech and gesture is no longer mere science fiction. Unlike interpersonal communication, human-computer interaction happens through explicit rules and logic, yet we expect the computer to understand us as other people do. Such accuracy requires introducing psychological communication elements into human-computer interaction.

Human intentions are not complete without emotions and personality; as a result, for successful human-computer interaction, the computer should have a good understanding of the individual. Even a simple sentence such as "Leave me alone." requires a different interpretation and response when we are happy or sad. An introverted individual may need solitude, while an extrovert may require attention in the same situation. A computer that can sense the individual's psychological state can respond highly accurately. Services can better customize digital content, virtual tutors can more intelligently adapt to user's understanding level, personal assistants can make more appropriate suggestions, and conversational agents can behave more realistically.

Involving many parameters, capturing the inner-state of the individual requires an in-depth examination, and thus data-driven approaches yield better performance. On the

[*] Corresponding author
[†] Equal contribution

other hand, data can be susceptible to bias and may introduce privacy issues, and a well-equipped system should resolve these issues successfully. Utilizing mid-level features, we filter out unnecessary information that may expose private details. To this end, we follow a well-known movement analysis framework that also reduces the risk of network bias. For example, body shape is influential in personality perception (Hu et al., 2018), but such associations follow stereotypes and do not apply to human-computer interaction.

Capturing the body shape will not help a system that tries to understand the user's inner-state to recommend music; conversely, it would harm the user's feelings. Similarly, age, gender, weight, height, and ethnicity can influence personality perception (Melamed, 1992; Faith et al., 2001; Chan et al., 2012; Löckenhoff et al., 2014) but do not carry much information about the user's inner-self. We acknowledge that appearance can expose the personality in some instances. For example, people can use different hair models, clothes, and accessories to communicate their feelings. On the other hand, there is no guarantee that every messy hair indicates an irresponsible personality. The choices about appearance can be mere stylistic. As a result, we focus on pose-related features only.

One of our goals is to show that accurate personality prediction is possible without using features that reveal the individual's identity. We do not utilize image-level features and vocal data while delivering state-of-the-art performance. To this end, we analyze movement style to capture personality. While movement style can be acted as well, we believe it is more reliable due to its biological foundation. For example, a slanted posture has a more concrete connection to low enthusiasm, probably caused by low energy or boredom. Similarly, reserved nature is best represented by enclosing and retreating movements since they relate to self-protection. Posture and mood are so closely related that one can feel better by mimicking a cheerful body pose (Peper et al., 2016). Although many resources that focus on body language try to uncover the relationship between the body and the inner-self, there are no definite rules about pose-personality; thus, we utilize data to associate self-reported personality with pose-related features.

This work focuses on Big Five personality recognition using landmark-based skeletal animation extracted from real-life videos. Skeletal landmark positions are shown to be effective in affect perception (Pollick et al., 2001, 2002), and we advocate that they are also beneficial for personality recognition. Our goal is to utilize behavioral cues only, without collecting and using image-level information that potentially raises privacy concerns and network bias. For this reason, in addition to the skeletal pose data, we utilize mid-level features based on Laban Movement Analysis (LMA). LMA captures the style of movement using minimal parameters that express the change in skeletal configuration. We utilize Graph Convolutional Networks (GCNs), commonly used for action recognition, to estimate each Big Five personality dimension in a regression model. The resulting model offers an effective solution for human understanding used in numerous applications, including personal assistants, conversational agents, targeted content marketing, and medical diagnosis, where preserving the individual's privacy is essential.

Our contributions include

- novel mid-level features following LMA derived from skeletal animation landmarks,
- a novel personality estimation network that combines GCNs with a regression task, and

- state-of-the-art performance without exposing image-level features to the network and by only using skeleton data.

The organization of the rest of the paper is as follows. Section 2 examines the research of similar interest. In Section 3, we introduce the details of our network model, including feature calculations. Section 4 presents experimental results. Finally, we provide a conclusion with a discussion on future directions in Section 5.

2. Related Work

This section reviews the related works in three categories: *personality recognition*, *action recognition*, and *Laban movement analysis*.

2.1. Personality Recognition

Recent works in personality recognition use deep neural networks to infer the personality traits (Palmero et al., 2021; Dotti et al., 2020a,b; Suen et al., 2019; Gorbova et al., 2018; Shao et al., 2021; Aslan et al., 2021; Li et al., 2020; Wei et al., 2018; Salam et al., 2021). The majority of the works combine multi-modal features. Palmero et al. (2021) combine visual, audio, and metadata features like gender, age, ethnicity, and perceived attractiveness to train a transformer model. SMART-SAIR (Salam et al., 2021), winner of ICCV'21 Understanding Social Behavior in Dyadic and Small Group Interactions Challenge, combines visual features, textual features (transcripts), and gender information to train personalized models. Aslan et al. (2021) train different models to fuse image, audio, and textual features, and they combine the trained models using an attention-based model. Unlike these methods, we want to only focus on the skeleton data because this information can be easily acquired from videos using pose estimation techniques. However, other information such as audio, transcript, and demographic information may not be available when these methods are applied in a real-world setting.

Dotti et al. (2020b) propose a CNN based personality recognition system. They convert the skeleton data extracted from video sequences into an image and feed it into the CNN architecture. They utilize the first few layers of VGG19 with pre-trained weights. They use both per-individual skeleton data and the distance between different individuals. The usage of CNN might suffer from the loss of spatial information between the joints. They also extend their work in Dotti et al. (2020a) and use metric learning to improve the accuracy. We use GCNs to capture the spatial and temporal information between skeleton joints. While they focus on the classification problem, we approach personality recognition as a regression task. This approach enables us to have a deeper understanding of the individual's personality. Additionally, we can use regression results to synthesize animation, as most work that focuses on generation utilize quantitative personality input. Thus, we can visually compare the input video of our network to the generated animation based on the regression results as a future task.

Suen et al. (2019) use only facial features to predict personality. They also consider the task as a classification problem and utilize a CNN. They use face images and the facial landmarks of each person to calculate the corresponding OCEAN personality factors. The direct use of image-level features for the face might develop some bias, as mentioned in Serna et al. (2021). Hence, we focus only on the skeleton data in our solution.

2.2. Action Recognition

We examine the Action Recognition literature because we consider recognizing the action and personality embedded in a movement as similar tasks. Thus we aim to benefit from the methods developed for Action Recognition, which is a field more extensively studied than Personality Recognition. We specifically focused on Action Recognition Networks that operate on skeleton data using GCNs.

Shi et al. (2019a) utilize a GCN, but they use directed graphs for that purpose. Initially, the skeleton is an undirected graph; they convert it to a Directed Acyclic Graph (DAG). In their DAG, the vertex closer to the root points to the other vertex farther away. They define the root as the center of gravity. Following these rules, they construct the graph and introduce convolution operations for the DAG. The resulting network performs competitively on the Action Recognition datasets.

Liu et al. (2020) introduce an Action Recognition Network, called *MS-G3D*. The network contains Graph Convolution operations to encode spatial information between the joints. Furthermore, they apply a convolution operation on the time domain to capture the temporal information between successive frames. We selected it as our backbone network because of its high performance in Action Recognition challenges. Due to its unique structure, their network can handle sophisticated relations in both spatial and time domains, which is essential for the Personality Recognition task.

Previous works also use additional hand-crafted features like motion vectors or angles. They assume a multiple-stream approach where they train multiple neural networks and ensemble the results (Simonyan and Zisserman, 2014; Shi et al., 2019a,b). However, we did not implement multiple deep networks for each Laban feature. We preferred having a single network for all features because Laban features' prediction shortens, and interactions become more pronounced in this way. Originally, MS-G3D also enabled motion-vectors, but it does so by creating a separate network for that purpose (Liu et al., 2020). We improve it to fuse Laban features at the first layer of the network.

2.3. Laban Movement Analysis

LMA is initially used for dance choreography design and adapted to general human motion analysis. LMA concepts examine the body pose during motion concerning the self and the surrounding environment. The relation between LMA and OCEAN personality is well-studied. Sonlu et al. (2021) utilize LMA based animation modification parameters to express different personalities in conversational agents. Durupinar et al. (2016) establish a connection between OCEAN personality and LMA parameters based on perceived traits in atomic actions, using animations crafted with the help of LMA experts. The mapping validation is based on participants filling out a standard personality questionnaire. They focus on the apparent personality, and the peer-rated results may differ from the self-reported personality (Ready et al., 2000). However, we believe the correlation between LMA parameters and personality is not solely about perception. There is a strong bond between the apparent posture and the inner-self (Peper et al., 2016), and the body mainly reveals

Table 1: Correlation map between Laban features and OCEAN parameters. x signifies that two entities are correlated.

	O	C	E	A	N
Flow	x	x	x		x
Time		x	x	x	x
Space	x	x	x		x
Weight				x	

the truth about feelings (Fast, 1970). We adapt perception-based LMA parameters to self-reported personality and show that the use of Laban features improves the performance in personality recognition.

Application of LMA in the study of personality is not uncommon (Levy and Duke, 2003). However, there are no definite rules about the quantification of LMA parameters. In this work, we propose Laban-based features that can be derived from the skeletal pose, inspired by the use of Laban Effort parameters in Durupinar et al. (2016). These parameters are *Flow*, *Time*, *Space* and *Weight*. *Flow* signifies body tension or vibration; *Time* shows how urgent the movement is, *Space* shows the relation to the surrounding environment, and *Weight* is the measure of the impact of the movement. Table 1 shows the correlated Laban features and OCEAN factors. We can infer that each Laban feature will help the network learn a particular OCEAN parameter from this information.

3. Method

Our method consists of three stages: *Input Preprocessing*, *Calculation of Laban Features*, and *Action Recognition* (cf. Figure 1).

3.1. Input Preprocessing

We apply several preprocessing techniques from the Action Recognition literature. We transform the skeleton to align hip-spine bone with one axis and shoulder bone with another axis, as in Liu et al. (2020). We use all of the frames in each video instead of skipping some. Since we cannot fit the whole video into the memory, we divide each video into clips of 2000 frames. That way, each 2000-frame segment corresponds to 80 seconds of the video. We take the average of all predictions for a person and use it as the final prediction.

3.2. Calculation of Laban Features

To extract Laban features, we need both joint and bone graphs. Joint graph is provided with the dataset, but the bone graph is not given; hence, we need to calculate it based on the joint graph. The bone information is necessary because we need rotation angles, which are easier to calculate over the bones. We apply the approach used by Liu et al. (2020). Figure 3 shows the joint and bone graphs. The joint graph is an undirected graph whose structure is identical to SMPL model (Loper et al., 2015). We generate the bone graph out of the joint graph by converting it to a Directed Acyclic Graph. That is, for an $edge_{ij}$, $node_i$ is directed to $node_j$. As a feature vector of $node_i$, we hold the unit vector pointing

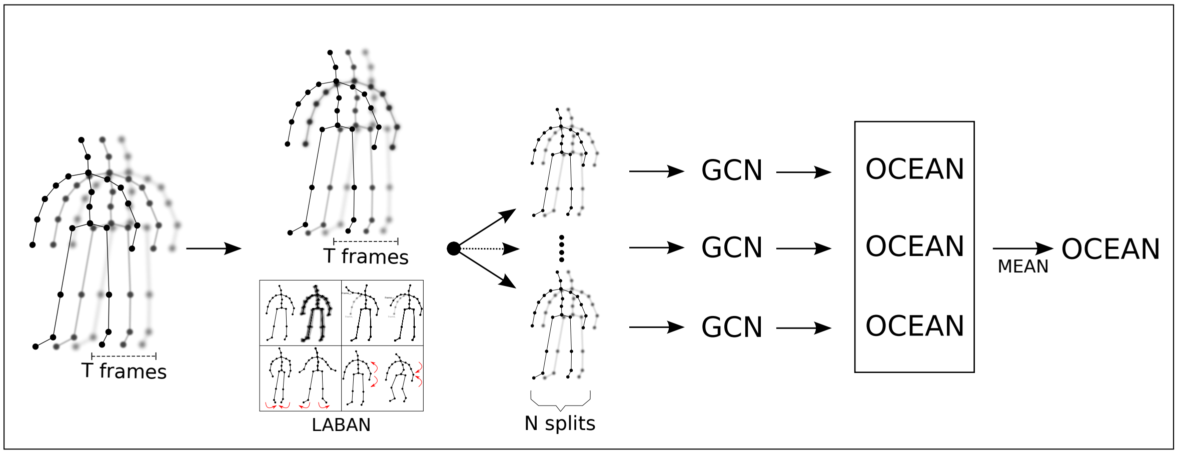

Figure 1: The overview of our method. Our input is a long sequence of the skeleton. We enrich the skeleton data with Laban features and divide the long data into short chunks. We fed each chunk into a Graph Convolution Network (GCN). Our final prediction is the average estimated OCEAN traits made for a person.

from $node_i$ to $node_j$. For instance, there is a directed edge going from $node_{22}$ to $node_{20}$, which means inside $node_{22}$, we hold the unit vector between $node_{22}$ and $node_{20}$. Formally, assume a directed edge $edge_{ij}$, then $bone\_vector_i = normalize(pos_j - pos_i)$. The $bone\_vector$ corresponds to the bone between these two joints and we use it to calculate necessary angles.

Although we show the graphs individually, we can store all the information in one graph because the joints of both graphs are identical. After creating the joint and bone graphs, we calculate the Laban features as follows:

Space and Weight: Both of these features correspond to bone angles. Hence, we treat them equally. Specifically, the calculation of bone angles is straightforward. Because we have the bone vectors available, we calculate the dot product of the bone vectors with each dimension and apply the *arccos* function to calculate the angles.

Flow: We calculate this feature as the 2^{nd} numerical derivative of the bone rotations. We expect it to capture the vibration of the motion approximately.

Time: This feature signifies the movement speed of the person. To represent this, we calculate the 1^{st} numerical derivative of the joint positions for each joint.

After calculating all immediate features, we concatenate them with the three-dimensional joint-position vector. As a result, we create a 15-dimensional vector for each joint. So, a graph with 24-nodes (see Figure 3 (a)) where each node contains 15-dimensional information is sent to the Action Recognition Network (ARN). Unlike other methods that run different networks for each feature, we create a single graph structure and execute it in a single network.

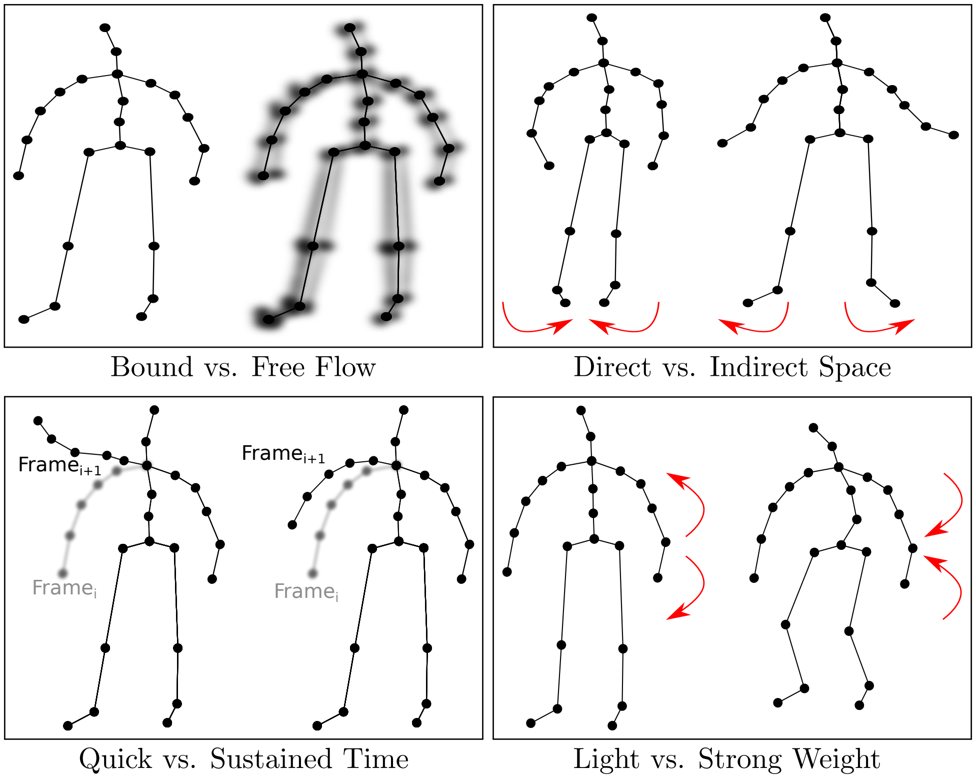

Figure 2: Skeleton images representing each end of Laban features. We show Free Flow using blurred joints as it corresponds to vibrating motion. For Space and Weight, arrows indicate the general direction of limb rotations. For Time, we overlay two consecutive frames in each image; the Quick one has more displacement while the Sustained one has less.

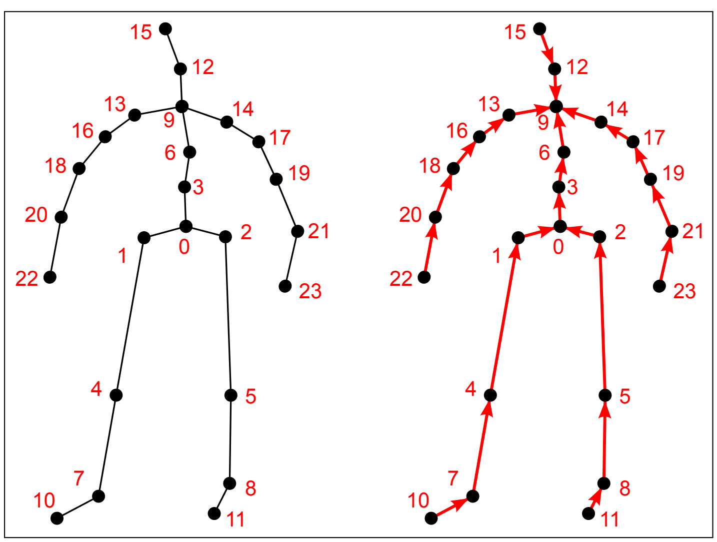

Figure 3: Joint graph (left) and Bone graph (right).

3.3. Action Recognition

We use an ARN as the backbone of our architecture. Our skeleton input is in the form of a graph structure. Hence, we utilize Graph Convolution-based ARN. Although our preliminary research inspired us to choose the MS-G3D architecture (Liu et al., 2020), the backbone ARN use is arbitrary, and we can replace it with a different network of similar nature. We utilized the ARN, which performs well in Action Recognition tasks, with minimal modification. We observe strong over-fitting at early epochs, and therefore, we inserted Dropout layers to several parts (i.e., especially the temporal convolution layers) of MS-G3D as a solution.

During training, we monitor training and validation losses. We save the network's weights when the validation loss is minimum and use it during inference. We use Stochastic Gradient Descent (SGD) for optimization with a learning rate of 0.05 and weight decay of 0.001. The batch size we selected was 16, and we ran the network as many as 200 epochs. We have also adjusted the parameters of the underlying ARN (MS-G3D). We selected the GCN scale as 13 and the G3D scale as 6. We also parameterized the number of layers at each Spatio-Temporal Graph Convolution (STGC) block. Initially, there are three blocks. We set the number of layers at each block as 12, 24, and 48, respectively. For further details, please refer to our Github implementation on `https://github.com/Rgtemze/PersonalityRecognition`.

4. Results

To train, validate and test our network, we use the UDIVA v0.5 dataset proposed by Palmero et al. (2021, 2022), which includes both self and peer-reported personality labels in addition to audiovisual features of the participants in face-to-face dyadic interactions. Please refer to the related work for details of the participants. We conduct our experiments on a computer with an NVIDIA Tesla P100 GPU. Table 2 summarizes the performance of our models in comparison to the UDIVA challenge winner, SMART-SAIR (Salam et al., 2021), and the baseline (Palmero et al., 2021). We use the results of other methods as reported in the corresponding sources because we use the same test dataset. We experiment with the existing ARN, called MS-G3D, without Laban features. We had to optimize its hyper-parameters to improve the results. Our method with Laban features seems to outperform the vanilla MS-G3D network, which does not use any Laban features. In addition, we can achieve lower loss values than the SMART-SAIR and the baseline. The Laban features help the network learn the personality traits. Moreover, our method only uses the skeleton data while others use various features. We showed that it is possible to perform very well with only the skeleton data.

To see the contribution of each Laban feature, we conduct an ablation study as shown in Table 3. Since the Space and Weight parameters stem from the same mathematical foundation, we treat them equally. Removing any of the Laban features considerably reduces the model's overall performance. Moreover, removing some features can decrease the loss value of some parameters. We expect such behavior because one feature might interfere with the correct prediction of another parameter. We can see a performance boost if we rule out that feature.

We observe that the figures in Tables 1 and 4 are correlated, showing that our results are on par with the user studies provided in Durupinar et al. (2016) and Sonlu et al. (2021). For instance, these user studies showed that the Flow feature correlates with O, C, E, N parameters. We can observe in Table 4 that the absence of this parameter caused a significant loss increase for most of these parameters compared to the lack of other Laban features. However, for the Agreeableness parameter, ruling out this parameter did not cause an increase in loss. In addition, a combination of the Space and Weight features cover all of the OCEAN parameters; their absence created a significant loss increase in all of the parameters.

We have provided a scatter plot for each person in the test set (see Figure 4). As we divide the videos into chunks, we end up with a different prediction for each chunk. Then, the predicted prediction for a person is defined as the mean of all predictions in all 2000-frame chunks. As we know that the training data is centered around a zero mean, our model focuses on correctly predicting the data around the mean. Hence, it predicts people whose personality traits fall into that range. However, our method could not correctly predict personality trait values too far from the mean.

Table 2: Comparison of our method with the existing networks. Values are Mean Square Error (MSE) loss results on the UDIVA v0.5 test dataset.

Method	Mean	O	C	E	A	N
Challenge Winner (Salam et al., 2021)	0.769	0.711	0.723	0.867	0.548	0.997
Baseline (Palmero et al., 2021)	0.818	0.744	0.794	0.886	0.653	1.012
Ours (MS-G3D)	0.833	**0.703**	0.695	0.871	0.665	1.229
Ours (Laban)	**0.723**	0.887	**0.564**	**0.590**	0.718	**0.859**

Table 3: Ablation study of our proposed approach for Laban features. Values are MSE loss results on the UDIVA test dataset.

Method	Mean	O	C	E	A	N
Laban	**0.723**	0.887	**0.564**	**0.590**	0.718	**0.859**
Laban w/o Flow	0.78	0.76	0.723	0.673	**0.541**	1.202
Laban w/o Time	0.747	0.700	0.653	0.682	0.542	1.157
Laban w/o Space and Weight	0.76	0.755	0.685	0.83	0.64	0.89

Table 4: Per-column inverse ranking extracted from Table 3.

Method	O	C	E	A	N
Laban w/o Flow	(1)	(1)	(3)	(3)	(1)
Laban w/o Time	(3)	(3)	(2)	(2)	(2)
Laban w/o Space and Weight	(2)	(2)	(1)	(1)	(3)

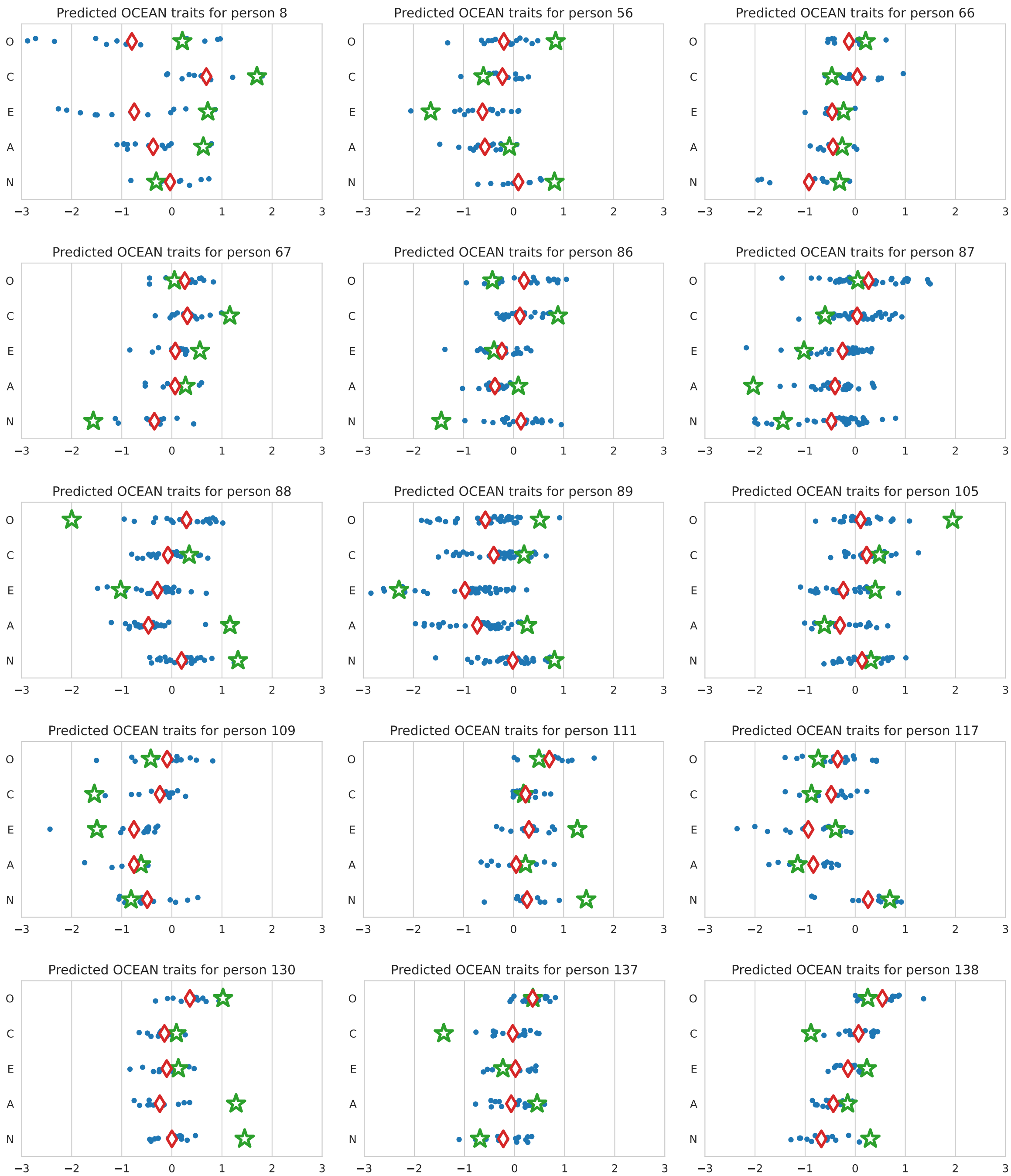

Figure 4: Scatter plots showing the predictions for the people in UDIVA test set using our best model. Each blue dot corresponds to the predicted value for that 2000-frame length chunk. Diamond shape (◇) is the mean of all predictions for that person. Star shape (⋆) is the ground truth.

5. Conclusion and Future Directions

In this work, we extend an existing ARN, called MS-G3D, and modify it into a Personality Recognition Network. We utilize mappings from psychology and use hand-crafted Laban features, which increase personality recognition accuracy. We show that using these features significantly decreases the MSE loss and helps to outperform the baseline performance and the challenge winner. In addition, our method achieves this performance by only using the skeleton information, unlike other methods using multi-modal data. We believe various applications can benefit from not utilizing the private information that can reveal the individual's identity. For example, personality recognition can be used in public areas for targeted marketing; thus, public screens can display more relevant information to the customer's personality. Our approach can guarantee that no private information is revealed to the network in this scenario. We admit that utilizing image-level features can improve performance in some instances. We leave studying the interrelation of the audiovisual features with Laban features as a future task. An ablation study of a personality recognition network that utilizes LMA-based skeletal features and video input can reveal important details about how personality is expressed through different communication elements.

We verify the correlation between Laban features and OCEAN parameters with the ablation study. By adding extra features calculated from the skeletal pose, we show that the network can estimate people's personalities more accurately. Finding more such features and utilizing them in human understanding networks seems to be a promising direction. For example, the individual's facial expression can be examined utilizing high-level descriptors regarding the changes during an interaction. We use our Laban features to focus on one individual at a time; modeling the dyadic interaction in similar parameters can improve the results. For example, the proximity of one participant's joints to the other's can reveal information about extroversion. Network modules can also calculate Laban features automatically, which we leave as future work. Another future direction is to utilize our Personality Recognition Network in adversarial training to form a system that automatically generates personality-rich animation. Then, it would be possible to look closer at how personality is expressed through body language; thus, we can utilize it better for recognition. We consider body language to be more connected to the inner-self. Wearing certain accessories, speaking specific words, or making various facial expressions can make an individual look like a particular persona, but in the end, it is behavior that reveals the true nature of the personality.

References

Süleyman Aslan, Uğur Güdükbay, and Hamdi Dibeklioğlu. Multimodal assessment of apparent personality using feature attention and error consistency constraint. *Image and Vision Computing*, 110:Article no. 104163, 9 pages, 2021.

Wayne Chan, Robert R McCrae, Filip De Fruyt, Lee Jussim, Corinna E Löckenhoff, Marleen De Bolle, Paul T Costa Jr, Angelina R Sutin, Anu Realo, Jüri Allik, et al. Stereotypes of age differences in personality traits: Universal and accurate? *Journal of Personality and Social Psychology*, 103(6):1050–1066, 2012.

Dario Dotti, Esam Ghaleb, and Stylianos Asteriadis. Temporal triplet mining for personality recognition. In *Proceedings of the 15th IEEE International Conference on Automatic Face and Gesture Recognition*, FG 2020, pages 379–386. IEEE, 2020a.

Dario Dotti, Mirela Popa, and Stylianos Asteriadis. Being the center of attention: A person-context CNN framework for personality recognition. *ACM Transactions on Interactive Intelligent Systems*, 10(3):Article no. 19, 20 pages, 2020b.

Funda Durupinar, Mubbasir Kapadia, Susan Deutsch, Michael Neff, and Norman I Badler. PERFORM: Perceptual approach for adding OCEAN personality to human motion using Laban Movement Analysis. *ACM Transactions on Graphics*, 36(1):Article no. 6, 16 pages, 2016.

Myles S Faith, Jonathan Flint, Christopher G Fairburn, Guy M Goodwin, and David B Allison. Gender differences in the relationship between personality dimensions and relative body weight. *Obesity Research*, 9(10):647–650, 2001.

Julius Fast. *Body Language*. Simon and Schuster, New York, NY, 1970.

Jelena Gorbova, Egils Avots, Iiris Lüsi, Mark Fishel, Sergio Escalera, and Gholamreza Anbarjafari. Integrating vision and language for first-impression personality analysis. *IEEE MultiMedia*, 25(2):24–33, 2018.

Ying Hu, Connor J Parde, Matthew Q Hill, Naureen Mahmood, and Alice J O'Toole. First impressions of personality traits from body shapes. *Psychological Science*, 29(12): 1969–1983, 2018.

Jacqyln A Levy and Marshall P Duke. The use of Laban movement analysis in the study of personality, emotional state and movement style: An exploratory investigation of the veridicality of "body language". *Individual Differences Research*, 1(1):39–63, 2003.

Yunan Li, Jun Wan, Qiguang Miao, Sergio Escalera, Huijuan Fang, Huizhou Chen, Xiangda Qi, and Guodong Guo. CR-Net: A deep classification-regression network for multimodal apparent personality analysis. *International Journal of Computer Vision*, 128(12):2763–2780, 2020.

Ziyu Liu, Hongwen Zhang, Zhenghao Chen, Zhiyong Wang, and Wanli Ouyang. Disentangling and unifying graph convolutions for skeleton-based action recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, CVPR '20, pages 143–152, 2020.

Corinna E Löckenhoff, Wayne Chan, Robert R McCrae, Filip De Fruyt, Lee Jussim, Marleen De Bolle, Paul T Costa Jr, Angelina R Sutin, Anu Realo, Jüri Allik, et al. Gender stereotypes of personality: Universal and accurate? *Journal of Cross-Cultural Psychology*, 45(5):675–694, 2014.

Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. SMPL: A skinned multi-person linear model. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia)*, 34(6):Article no. 248, 16 pages, October 2015.

Tuvia Melamed. Personality correlates of physical height. *Personality and Individual Differences*, 13(12):1349–1350, 1992.

Cristina Palmero, Javier Selva, Sorina Smeureanu, Julio C.S. Jacques Junior, Albert Clapés, Alexa Moseguí, Zejian Zhang, David Gallardo, Georgina Guilera, David Leiva, et al. Context-aware personality inference in dyadic scenarios: Introducing the UDIVA dataset. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision Workshops*, WACVW '21, pages 1–12, 2021.

Cristina Palmero, German Barquero, Julio C. S. Jacques Junior, Albert Clapés, Johnny Núñez, David Curto, Sorina Smeureanu, Javier Selva, Zejian Zhang, David Saeteros, David Gallardo-Pujol, Georgina Guilera, David Leiva, Feng Han, Xiaoxue Feng, Jennifer He, Wei-Wei Tu, Thomas B. Moeslund, Isabelle Guyon, and Sergio Escalera. Chalearn LAP challenges on self-reported personality recognition and non-verbal behavior forecasting during social dyadic interactions: Dataset, design, and results. In *Understanding Social Behavior in Dyadic and Small Group Interactions*, volume 173 of *Proceedings of Machine Learning Research (PMLR)*, pages 4–52, 2022.

Erik Peper, Annette Booiman, I-Mei Lin, and Richard Harvey. Increase strength and mood with posture. *Biofeedback*, 44(2):66–72, 2016.

Frank E Pollick, Helena M Paterson, Armin Bruderlin, and Anthony J Sanford. Perceiving affect from arm movement. *Cognition*, 82(2):B51–B61, 2001.

Frank E. Pollick, Vaia Lestou, Jungwon Ryu, and Sung-Bae Cho. Estimating the efficiency of recognizing gender and affect from biological motion. *Vision Research*, 42(20):2345–2355, 2002.

Rebecca E Ready, Lee Anna Clark, David Watson, and Kelley Westerhouse. Self-and peer-reported personality: Agreement, trait ratability, and the "self-based heuristic". *Journal of Research in Personality*, 34(2):208–224, 2000.

Hanan Salam, Oya Celiktutan, Viswonathan Manoranjan, Iman Ismail, and Mukherjee. Understanding social behavior in dyadic and small group interactions challenge fact sheet: Automatic self-reported personality recognition track. In *Understanding Social Behavior in Dyadic and Small Group Interactions Challenge at ICCV*, 2021.

Ignacio Serna, Alejandro Peña, Aythami Morales, and Julian Fierrez. InsideBias: Measuring bias in deep networks and application to face gender biometrics. In *Proceedings of the 25th International Conference on Pattern Recognition*, ICPR '21, pages 3720–3727. IEEE, 2021.

Zilong Shao, Siyang Song, Shashank Jaiswal, Linlin Shen, Michel Valstar, and Hatice Gunes. Personality recognition by modelling person-specific cognitive processes using graph representation. In *Proceedings of the 29th ACM International Conference on Multimedia*, MM '21, page 357–366, New York, NY, USA, 2021. Association for Computing Machinery.

Lei Shi, Yifan Zhang, Jian Cheng, and Hanqing Lu. Skeleton-based action recognition with directed graph neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, CVPR '19, pages 7912–7921, 2019a.

Lei Shi, Yifan Zhang, Jian Cheng, and Hanqing Lu. Two-stream adaptive graph convolutional networks for skeleton-based action recognition. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, CVPR '19, pages 12026–12035, 2019b.

Karen Simonyan and Andrew Zisserman. Two-stream convolutional networks for action recognition in videos. *arXiv preprint arXiv:1406.2199*, 2014.

Sinan Sonlu, Uğur Güdükbay, and Funda Durupinar. A conversational agent framework with multi-modal personality expression. *ACM Transactions on Graphics*, 40(1):Article no. 7, 16 pages, 2021.

Hung-Yue Suen, Kuo-En Hung, and Chien-Liang Lin. TensorFlow-based automatic personality recognition used in asynchronous video interviews. *IEEE Access*, 7:61018–61023, 2019.

Xiu-Shen Wei, Chen-Lin Zhang, Hao Zhang, and Jianxin Wu. Deep bimodal regression of apparent personality traits from short video sequences. *IEEE Transactions on Affective Computing*, 9(3):303–315, 2018.