



EÖTVÖS LORÁND UNIVERSITY

FACULTY OF INFORMATICS

DEPT. OF ARTIFICIAL INTELLIGENCE

Personality Recognition based on Visual Cues and Facial Movement Features

Supervisor:

Ádám Fodor

Teaching Assistant

Author:

Muhammad Luqman Hakim

Computer Science MSc

Budapest, 2022

Contents

1	Introduction	1
2	Methodology	3
2.1	Action Unit	4
2.1.1	ME-GraphAU	4
2.1.2	OpenFace	5
2.2	Head Pose	5
2.3	Face Movement	6
2.4	Models	7
3	Experimental Setup	9
3.1	Datasets	9
3.2	Evaluation Metrics	10
3.3	Experimental Setting	10
4	Experiments and Results	12
4.1	Reproducing ME-GraphAU Results	12
4.2	Retraining ME-GraphAU on the Combined Dataset	15
4.3	Personality Recognition using AUs Features	16
4.4	Personality Recognition using Head Pose and Face Movement Features	18
4.5	Personality Recognition using Fusion Feature	20
5	Conclusion	22
	Bibliography	23
	List of Figures	28
	List of Tables	29

Abstract

In order to make intelligent system having “human-like” manner in human-system interaction, machine learning (ML) and computer vision (CV) communities are interested in automatic sentiment analysis, emotion recognition, and personality recognition. Studies on personality computing vary from mono-modality data (audio, visual, or text) to multi-modality data such as audiovisual. This study aims to show the contribution of Action Units (AU), head pose and face movements in personality recognition. First Impression V2 database is used in the experiments. I used ME-GraphAU [1] and OpenFace [2] to extract essential Action Units features. I use 3DDFA-V2 [3] method to extract face landmarks and head pose features. Then, face landmarks are used to calculate face movement on each consecutive two frames. I train those features on several well-known models such as LSTM, GRU, and Transformer. The results indicate the relevance of these features and encourage their usage in the field of personality recognition.

Chapter 1

Introduction

A person can be recognized by his/her personality. It reveals some traits of how an individual think, feels, and acts. Numerous fields, including marketing, education, training, and human resource management, get benefits from personality. One specific example in the hiring process, a recruiter usually asks about candidates' personality or gives them a personality test because it is considered to be one of important factors in a working environment.

Apart from that, emotion and personality are two of the key clues in social communication that determine how we feel and act when we engage with others. We need to give intelligent systems similar characteristics if we want them to interact and communicate with us in a "human-like" manner. For those reasons, machine learning (ML) and computer vision (CV) communities are becoming more interested in sentiment analysis, emotion recognition, and personality recognition.

Studies on personality computing range from mono-modality (audio, visual, or text) to multi-modality such as audiovisual. In a text-based study, Hernández et al. worked on personality recognition based on Spanish text [4]. They surveyed 120 participants. They asked the participants to do a personality test and write some paragraphs. They trained a model to detect personality traits (Dominance, Influence, Steadiness, and Compliance) based on the written paragraph.

Some studies worked on text-based personality recognition which gets their data from social media [5] [6]. Another work is based on text in Curriculum Vitae (CV) [7]. They utilize machine learning algorithms to obtain the vital information needed to evaluate the applicants' personalities.

In the case of visual modality-based, Xiaodong Duan et al. [8] developed a person-

ality recognizer based on facial appearance. It was inspired by psychological findings that facial appearance can be viewed as the key clue in evaluating a human’s personality. They experimented with the study on a customized CelebAMask-HQ dataset [9]. Another study extract deep visual activity-based features extracted only from key-dynamic images [10]. They experimented on ELEA [11] and ChaLearn First Impression Dataset [12].

Fallahnezhad et al. built a personality recognizer using audio modality [13]. They asked participants to take the NEO-FFI questionnaire to define their personalities. For feature extraction, they used OpenSMILE toolkit [14] and ComParE 2013 audio feature set [15]. Liu et al. [16] audio-based study did the experiments on the standard SSPNet dataset [17]. They extracted essential audio features, i.e., signal power roll-off, pitch, and pause rate, and filtered out undesirable adulterations be it noise, silence, or uncertain pitch segments.

Regarding audiovisual modality-based models, Gorbova et al. [18] did experiments on the FI dataset. They extracted audio features (MFCCs, chroma, and tonality) using the Python library Librosa. Additionally, they used OpenFace [2] to extract AUs as visual features. Another study [19] employed audio, visual, and text modalities. They did the experiments on the First Impression dataset. To extract audio features, they used eGeMAPS [20] and Wav2Vec [21]. To extract visual features, they used OpenFace [2] and FAb-Net [22]. To extract text features, they used BERT [23] and RoBERTa [24].

In this study, I explored visual and movement features to recognize personality traits. I particularly utilized Action Units (AUs) and head pose for the visual feature and face movements for the movement feature. Hopefully, this study will help expanding previous studies that utilized AUs features by considering head pose and face movement and see how they contribute on recognizing human’s personality.

Chapter 2

Methodology

Figure 2.1 shows the big picture of the methods of this study. The first step is extracting feature representations from raw video data. Based on those extracted features, I trained LSTM, GRU, and Transformer, three well-known models that can process sequential data. Finally, the models will predict the Big-Five (OCEAN) apparent personality traits, which are Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism.

In this study, I consider 3 types of features for personality recognition, which are action units, head pose, and face movement. Regarding action units, I used 2 tools to extract them, ME-GraphAU [1] and OpenFace [2]. The details about ME-GraphAU and OpenFace will be described on section 2.1.1 and 2.1.2 respectively. Furthermore, I discuss about extracting head pose features on section 2.2 and calculating the face movement feature on section 2.3. The models that were used for personality recognition are described on section 2.4.

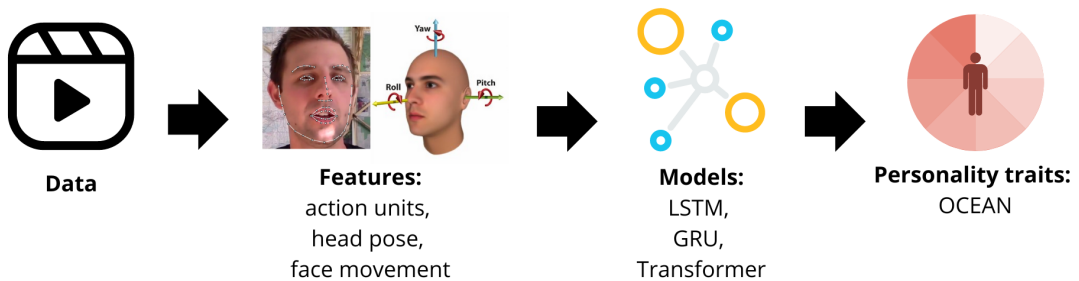


Figure 2.1: Methods for personality recognition

2.1 Action Unit

The Facial Action Coding System (FACS) is a comprehensive, anatomically based system for describing all visually discernible facial movements. It breaks down facial expressions into individual components of muscle movement, called Action Units (AUs). Action units are commonly used by behavioral scientists and computer scientists. Each AU describes one observable movement of a facial feature (e.g. eyebrows) by facial muscles. Some examples of AUs taken from [25] are shown in fig. 2.2.

One of the main reasons why AUs are a great fit as features for a machine learning model is it focuses only on facial muscle activities. It avoids any potential bias that may come from the image background and other appearance-based information that can be retrieved from raw visual data such as gender, age, ethnicity, etc.

AUs are usually labelled in two ways: presence (indicating if the AU is visible) and intensity (indicating how intense an AU is). In this study, I focus on the presence of AUs, which are a binary values: "0" means that the AU is not visible, while "1" implies its visibility.

Upper Face Action Units					
AU1	AU2	AU4	AU5	AU6	AU7
					
Inner Brow Raiser	Outer Brow Raiser	Brow Lowerer	Upper Lid Raiser	Cheek Raiser	Lid Tightener
*AU41	*AU42	*AU43	AU44	AU45	AU46
					
Lip Droop	Slit	Eyes Closed	Squint	Blink	Wink

Figure 2.2: Several examples of action units [25].

2.1.1 ME-GraphAU

ME-GraphAU [1] deep learns a unique graph to explicitly describe the relationship between each pair of AUs of the target facial display. Firstly, it encodes each AU's activation status and its association with other AUs into a node feature. Then, it learns a pair of multi-dimensional edge features to describe multiple task-specific relationship cues between each pair of AUs.

During both node and edge feature learning, this method also considers the influence of the unique facial display on AUs' relationship by taking the full face

representation as an input. Experimental results on BP4D and DISFA datasets show that both node and edge feature learning modules provide large performance improvements for CNN and Transformer-based backbones. The best setup achieved the state-of-the-art of AU recognition results.

This method will output the intensity of each AU represented in a value between 0 and 1. It will generate the intensity of 11 AUs (AU1, AU2, AU4, AU6, AU7, AU10, AU12, AU14, AU15, AU17, AU23) which later will be the features for personality recognition. Before further usage, the outputs of this method will be normalized using z-score normalization.

2.1.2 OpenFace

OpenFace [2] is an open-source toolkit for extracting features on the face such as facial landmarks, facial action units, gaze tracking, etc. It is one of the most popular tool for those tasks. Many researchers have utilized it in their studies. Concerning on AUs, it can produce 35 AUs values. 17 AUs for the intensity value (ranging on a 5-point scale). 18 AUs for the presence value (0 or 1).

To have a fair comparison, I will only consider 11 AUs which are common with the ME-GraphAU tool. They are AU1, AU2, AU4, AU6, AU7, AU10, AU12, AU14, AU15, AU17, and AU23. Furthermore, the outputs of this method will be normalized using z-score normalization before further usage.

2.2 Head Pose

Head pose includes the angle of yaw, pitch, and roll. Yaw is a rotation around the vertical axis, while pitch and roll are the rotations around the front-to-back and side-to-side axis, respectively. See figure 2.3 for the illustration. I used 3DDFA-V2 [3] to extract the head pose. It predicts three angles corresponding to yaw, pitch, and roll for each frame.

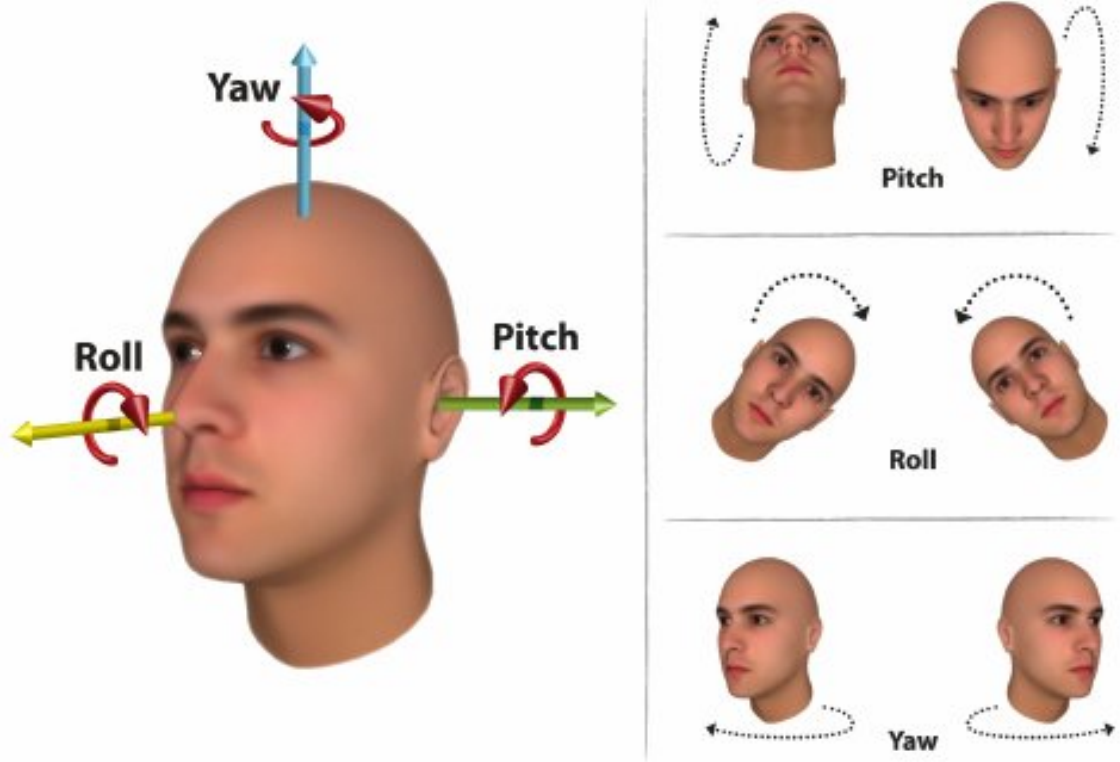


Figure 2.3: Head pose (yaw, pitch, roll) [26]

2.3 Face Movement

Face movement is derived from the spatial movement of the middle point of face landmarks. To get the face movement, I initially need to detect face landmarks. I used 3DDFA-V2 [3] to generate 68 points of face landmark for each frame (see fig. 2.4). Then, I get the middle point by calculating the mean of those 68 points. Finally, I obtained the movement feature by calculating the spatial difference between two consecutive middle points. In the end, I will have a sequence of movement feature for each video clip.

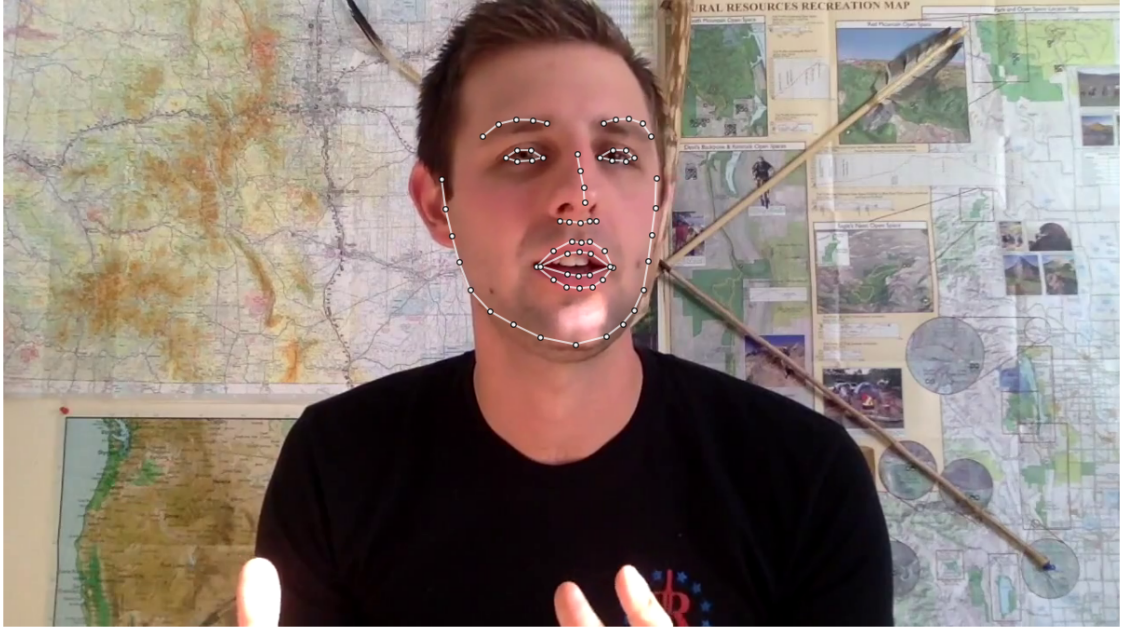


Figure 2.4: 68 points of face landmark

2.4 Models

I used 3 prominent models that can process sequential data, including LSTM, GRU, and Transformer. Long Short Term Memory (LSTM) is one variant of RNN. It is able to learn long-term dependencies. It was introduced by Hochreiter and Schmidhuber [27], and was refined and popularized by many people in following works. It is frequently used today and works incredibly well on a wide range of applications. Figure 2.5 (a) shows the architecture of LSTM model [28].

Gated Recurrent Unit (GRU) is a variation of LSTM and was introduced by Cho et al. [29]. It unites forget gates and input gates into a single update gate. It also combines the cell state and hidden state, and makes some other changes. The resulting model is simpler than standard LSTM model, and has been becoming more popular today. Figure 2.5 (b) shows the architecture of GRU model [28].

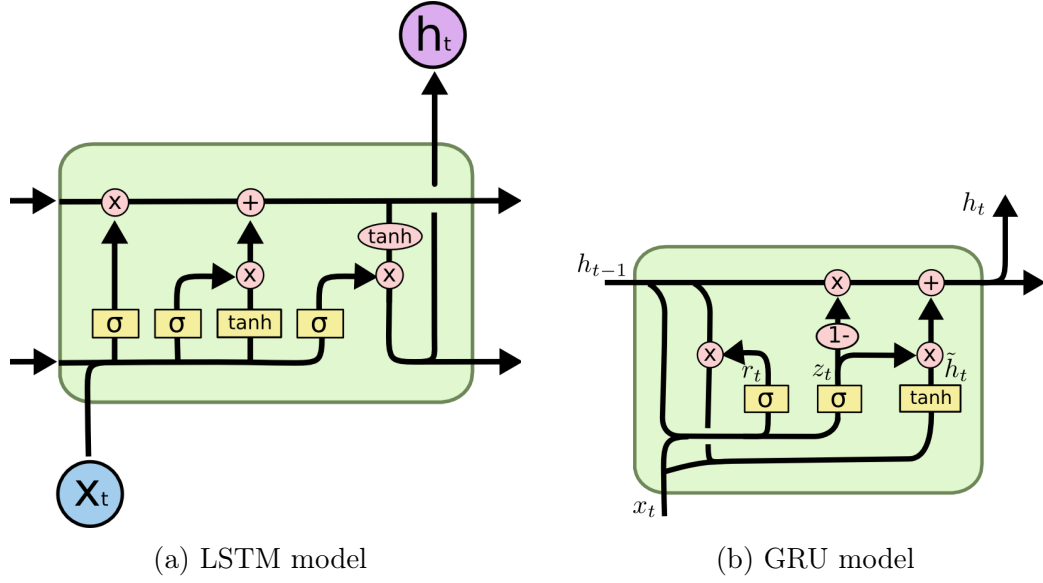


Figure 2.5: LSTM and GRU model [28]

Transformer, introduced by Vaswani et al. [30], is a model that uses attention mechanism. It consists of encoder and decoder components. The encoding component is a stack of 6 encoders. The decoding component is also a stack of 6 decoders. One key benefit of Transformer is its internal computation can work in parallel. Figure 2.6 shows the architecture of Transformer model [31].

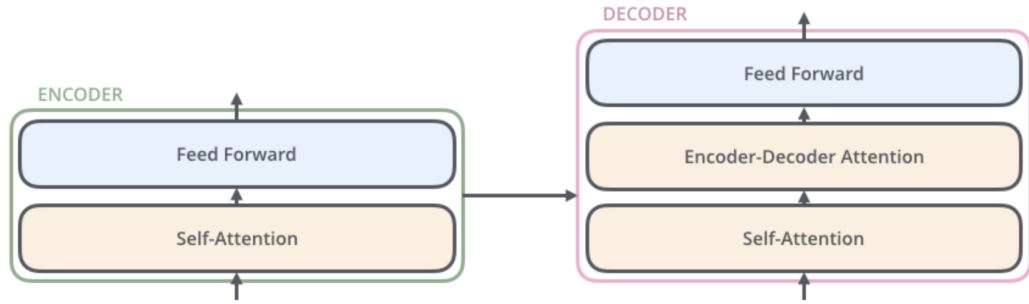


Figure 2.6: Encoder-Decoder Transformer model [30]

Chapter 3

Experimental Setup

This chapter describes how I set up the experiment during this study. In section 3.1, I discuss about the database I used for the experiments. In section 3.2, I explain about the metrics to measure the performance of the experiments. In section 3.3, I discuss about the models, hyperparameters, and other experimental strategy.

3.1 Datasets

ChaLearn First Impression V2: This work was experimented on the ChaLearn First Impression V2 (FI) for personality recognition [12]. It is the largest publicly available in-the-wild database on the topic. The FI database was released in the context of a computational challenge, where the goal was to automatically recognise the Big-Five (OCEAN) apparent personality traits of single individuals in videos: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism.

The dataset comprises 10,000 short video clips (average duration of 15s each, divided into training, validation, and test sets) extracted from more than 3,000 different YouTube videos of people talking to a camera. It was annotated using crowdsourcing through pairwise comparisons, which were further converted to continuous values [32]. The labels correspond to human annotators’ impressions. Therefore, the Big-five traits are considered as first impressions or perceived personalities.

BP4D: BP4D (Binghamton-Pittsburgh 3D Dynamic Spontaneous Facial Expression Database) [33] consists of 328 videos (about 140,000 facial frames). Its participants are 41 young adults (23 females and 18 males) who were asked to respond to 8 emotion elicitation tasks. For each task, there are both 3D and 2D

videos. It also includes manually annotated action units (FACS AU), automatically tracked head pose, and 2D/3D facial landmarks.

DISFA: DISFA database [34] (Denver Intensity of Spontaneous Facial Action) consists of 130,815 frames from 27 participants (12 females and 15 males) who were watching Youtube videos. Those frames are high resolution images with 1024×768 size. The database also includes 66 facial landmark points of each image in the database. It is a suitable database to develop computer algorithms for automatic AU detection and their intensities. The intensity of AU is in 0-5 scale. Each frame in DISFA is annotated with occurrence labels of several AUs and were manually scored by two human FACS (Facial Action Coding System) experts.

3.2 Evaluation Metrics

In personality recognition, referring to most works on FI dataset, I used “1-Mean Absolute Error (MAE)” as a performance metric. MAE formula is shown on equation 3.1. In AU detection task, which is associated on reproducing and retraining ME-GraphAU, I used F1 score as the main performance metric. F1 score formula is shown on equation 3.2. F1 score is the most common metric for this task.

$$MAE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i) \quad (3.1)$$

$$F1 = 2 * \frac{precision * recall}{precision + recall} \quad (3.2)$$

3.3 Experimental Setting

Regarding LSTM and GRU, I used the bidirectional variant (denoted as BiLSTM and BiGRU). Regarding Transformer, I used only the Encoder component which will encode the input data into a more meaningful representation. I did not use the decoder component because I did not need sequential output in this case. After passing through LSTM, GRU, and Transformer Encoder layers, the output will be fed into a fully-connected layer that will map to personality traits.

Grid search was performed to determine the hyperparameters of the models. Concerning the training on AUs (ME-GraphAU and OpenFace) features, I used 3

LSTM layers (256 units for each layer), dropout of 0.1. I used a Bidirectional GRU with 2 layers (256 units for each layer), dropout of 0.2. For Transformer, I used 6 stacks of encoder layers (256 units for each layer), dropout of 0.1.

Regarding the training on head pose and face movement features, I used BiLSTM and BiGRU with 2 layers (128 units for each layer), dropout of 0.2. For Transformer, I used 6 stacks of encoder layers (256 units for each layer), dropout of 0.2.

For the training on the combined features, I used BiLSTM and BiGRU with 2 layers (256 units for each layer), dropout of 0.1. For Transformer, I used 6 stacks of encoder layers (256 units for each layer), dropout of 0.2.

The RELU and Sigmoid activation functions are set in every model and in the output layer, respectively. ADAM optimizer and Mean Squared Error (MSE) loss function is used for all experiments. The models are trained for a maximum of 50 epochs and employed an early stopping strategy.

Chapter 4

Experiments and Results

I present the results of this study throughout this chapter. Prior training the model for personality recognition, I tested the ME-GraphAU tool, because it is relatively new and the authors claim it has state-of-the-art results. I tried to reproduce the ME-GraphAU results in order to see the reliability of the tool. Then, I evaluated the model’s performance in a cross-dataset setting, and retrained their model using an extended dataset to see its generalization attribute. I discuss about the reproduction of ME-GraphAU in section 4.1. The results of retraining ME-GraphAU on the combined (BP4D and DISFA) dataset will be described in section 4.2.

In the training phase of personality recognition, I trained the models on individual feature and a combination of features. I trained the models on AUs only. The results are described in section 4.3. Then, I trained the models based on head pose and face movement features. The details are in section 4.4. Finally, I did an experiment based on the fusion feature which is described in section 4.5.

4.1 Reproducing ME-GraphAU Results

Luo et al. [1] trained and evaluated the model of ME-GraphAU on BP4D and DISFA dataset independently. They make the code and pretrained model available so that others can reproduce their results in the paper. Prior to the reproduction, there are some required data preprocessing steps. Raw video data is converted into sequence of frames. Then face detection and alignment is applied on each frame using MTCNN [35]. To preserve the aspect ratio a 224 x 224 crop is used on the face region. See figure 4.1 to see frame image before and after transformation.

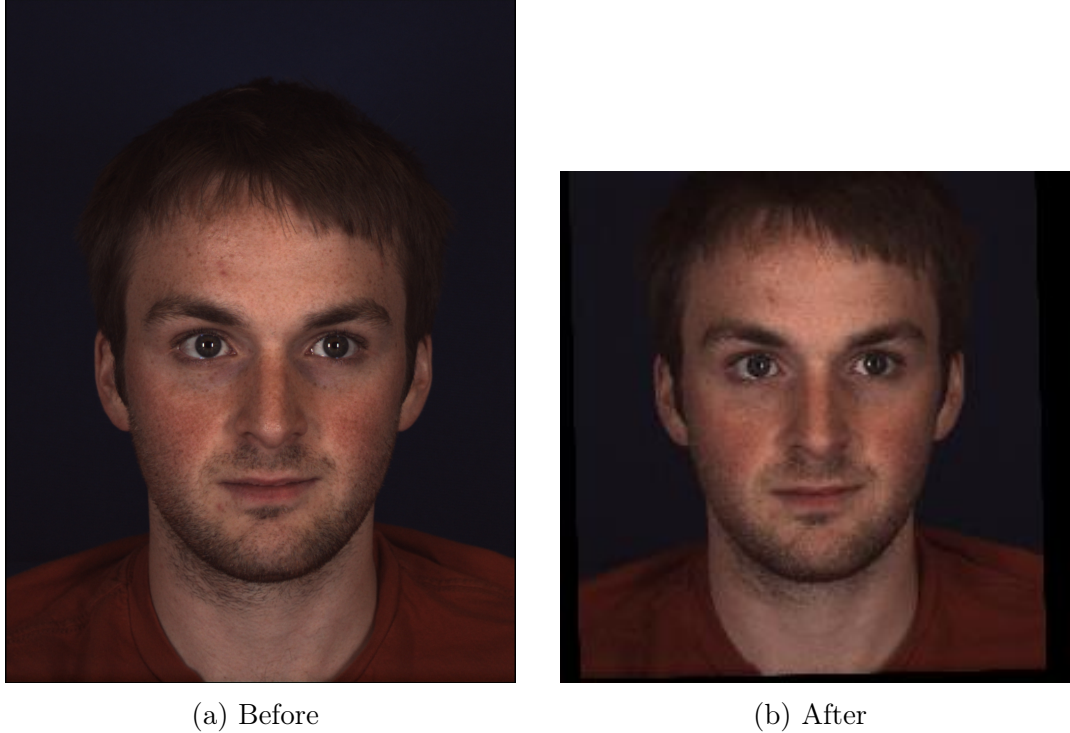


Figure 4.1: Frame image transformation using face detection and alignment

Table 4.1 shows the reproduction results executed on the BP4D dataset. F1 scores for each evaluated AU are presented. The second column contains the values derived from the paper. The third and the fourth column show F1 scores of the reproduction experiment with and without face alignment, respectively.

The achieved F1 score is just slightly less than the score presented in the original paper. Besides, the result of the experiment without face alignment is slightly better in this case. One possible reason is the participants do not have significant head pose variations on this dataset. As a consequence, face alignment does not remarkably help improving the result.

Action Unit	F1 (from the paper)	F1 (with alignment)	F1 (without alignment)
AU1	52.7	37.8	40.5
AU2	44.3	32.9	38.8
AU4	60.9	55.6	52.6
AU6	79.9	79.1	79.7
AU7	80.1	79.1	79.8
AU10	85.3	84.1	85.4
AU12	89.2	88.3	87.6
AU14	69.4	67.7	69.4
AU15	55.4	47.8	51.0
AU17	64.4	62.2	63.4
AU23	49.8	44.7	46.2
AU24	55.1	55.0	55.9
Average	65.5	61.2	62.5

Table 4.1: F1 score for reproduction of ME-GraphAU on BP4D dataset

Table 4.2 shows the reproducing results executed on DISFA dataset. F1 scores for each evaluated AU are presented. The second column contains the values derived from the paper. The third and the fourth column show F1 scores of the reproduction experiment with and without face alignment, respectively. On average, the achieved F1 score exceeds the target values in the original paper. Based on table 4.1 and table 4.2, ME-GraphAU is considered to be a reliable tool and is worth to use for future experiments.

Action Unit	F1 (from the paper)	F1 (with alignment)	F1 (without alignment)
AU1	52.5	58.3	56.6
AU2	45.7	51.1	47.3
AU4	76.1	77.0	76.2
AU6	51.8	51.4	54.4
AU9	46.5	47.6	50.6
AU12	76.1	76.1	76.0
AU25	92.9	93.5	92.8
AU26	57.6	64.7	59.3
Average	62.4	65.0	64.1

Table 4.2: F1 score for reproduction of ME-GraphAU on DISFA dataset

4.2 Retraining ME-GraphAU on the Combined Dataset

In most cases, a model is designed specifically for certain databases. As a result, when tested on different datasets, that model is often overfitting because their accuracy changes with the changes in the domain of dataset. For example in personality dataset, the tasks or the activities that the participants do differ from one dataset to other datasets. For that reason, it is important to train the model in combined data and see if it performs better.

While being introduced, the model of ME-GraphAU was trained on BP4D and DISFA datasets separately. I retrained the model on the combination of those two datasets. I only consider the AUs that both datasets have in common in the annotation, including AU1, AU2, AU4, AU5, AU6, AU9, AU12, AU15, AU17, and AU20. After the training phase, the model is evaluated on the test set of combined data, BP4D only, and DISFA only.

Table 4.3 shows the evaluation results of ME-GraphAU calculated in a cross-dataset setting and using the combined dataset. All values presented are F1 scores. There are some missing values in the table because the model was not trained on that particular AU.

In the second column, the ME-GraphAU model was trained on the BP4D dataset and evaluated on the DISFA dataset. In the fifth column, the model was trained on the combined dataset and was evaluated on the DISFA dataset. As shown in those two columns, the model trained on the combined dataset performed better in all AUs except in AU15. In the third column, the model was trained on the DISFA dataset and was evaluated on the BP4D dataset. In the fourth column, the model was trained on the combined dataset and was evaluated on the BP4D dataset. As shown in those two columns, the model trained on the combined dataset performed better in all AUs. To conclude, the model trained on the combined dataset is more generalized than that trained a single dataset.

Action Unit	Trained on BP4D Evaluated on DISFA	Trained on DISFA Evaluated on BP4D	Trained on combined	
			Evaluated on BP4D	Evaluated on DISFA
AU1	23.2	40.1	48.5	46.5
AU2	11.2	33.2	44.3	43.0
AU4	40.1	30.1	57.0	61.2
AU5	-	-	20.7	50.6
AU6	28.1	56.0	75.8	39.4
AU9	-	23.4	28.7	41.7
AU12	37.9	67.7	85.0	63.6
AU15	30.7	-	47.3	22.8
AU17	21.8	-	58.2	35.0
AU20	-	-	10.4	13.3

Table 4.3: Evaluation of ME-GraphAU on cross dataset and combined dataset

4.3 Personality Recognition using AUs Features

Table 4.4 shows the comparison between AUs that are extracted by ME-GraphAU and OpenFace. Overall, AUs by ME-GraphAU produced better performance in all models. It implies that AUs by ME-GraphAU have better representation as features. The highest score is obtained by GRU (0.8878), followed by Transformer (0.8870) and LSTM (0.8862). To see more detail regarding AUs by ME-GraphAU, table 4.5, table 4.6, and table 4.7 shows the personality-wise results using LSTM, GRU, Transformer respectively. To see more detail regarding AUs by OpenFace, table 4.8, table 4.9, and table 4.10 shows the personality-wise results using LSTM, GRU, Transformer respectively.

AU extraction method	Model	AVG 1-MAE
ME-GraphAU	LSTM	0.8862
	GRU	0.8878
	Transformer	0.8870
OpenFace	LSTM	0.8737
	GRU	0.8761
	Transformer	0.8777

Table 4.4: Evaluation of different sequence models on the FI dataset for the personality recognition task

As shown on table 4.5, all personality traits obtained good scores of 1-MAE. The highest score is achieved on Agreeableness (0.8955). Several personalities (Openness

and Extroversion) have moderate positive correlation between predicted and ground-truth personality. The rest (Conscientiousness, Agreeableness, Neuroticism) have low positive correlation. All correlation coefficients are statistically significant.

	O	C	E	A	N
1-MAE	0.8891	0.8802	0.8847	0.8955	0.8818
R	0.30	0.27	0.33	0.16	0.27
p-value	<0.05	<0.05	<0.05	<0.05	<0.05

Table 4.5: Results of personality recognition using AUs (ME-GraphAU) on LSTM

As shown on table 4.6, all personality traits obtained good scores of 1-MAE. The highest score is achieved by Agreeableness (0.8935). All personalities (Openness, Conscientiousness, Extroversion, Agreeableness, Neuroticism) have low positive correlation between predicted and ground-truth personalities. All correlation coefficients are statistically significant except on Agreeableness.

	O	C	E	A	N
1-MAE	0.8841	0.8761	0.8787	0.8935	0.8782
R	0.05	0.06	0.06	0.02	0.06
p-value	<0.05	<0.05	<0.05	0.33	<0.05

Table 4.6: Results of personality recognition using AUs (ME-GraphAU) on GRU

As shown on table 4.7, all personality traits obtained good scores of 1-MAE. The highest score is achieved by Agreeableness (0.8948). Several personalities (Openness, Conscientiousness, Extroversion) have moderate positive correlation between predicted and ground-truth personality. The rest (Agreeableness and Neuroticism) have low positive correlation. All correlation coefficients are statistically significant.

	O	C	E	A	N
1-MAE	0.8902	0.8818	0.8865	0.8948	0.8819
R	0.32	0.30	0.35	0.12	0.28
p-value	<0.05	<0.05	<0.05	<0.05	<0.05

Table 4.7: Results of personality recognition using AUs (ME-GraphAU) on Transformer

As shown on table 4.8, all personality traits obtained good scores of 1-MAE. The highest score is achieved by Conscientiousness (0.8890). All personalities (Openness, Conscientiousness, Extroversion, Agreeableness and Neuroticism) apparently do not have correlation between predicted and ground-truth personalities.

	O	C	E	A	N
1-MAE	0.8733	0.8890	0.8816	0.8711	0.8701
R	-0.01	-0.01	0	-0.01	-0.01
p-value	0.73	0.68	0.89	0.83	0.53

Table 4.8: Results of personality recognition using AUs (OpenFace) on LSTM

As shown on table 4.9, all personality traits obtained good scores of 1-MAE. The highest score is achieved by Conscientiousness (0.8895). Similar to LSTM results (fig. 4.8), all personalities (Openness, Conscientiousness, Extroversion, Agreeableness and Neuroticism) apparently do not have correlation between predicted and ground-truth personalities.

	O	C	E	A	N
1-MAE	0.8743	0.8895	0.8829	0.8739	0.8712
R	0.01	-0.01	-0.01	0	-0.02
p-value	0.60	0.69	0.65	0.88	0.40

Table 4.9: Results of personality recognition using AUs (OpenFace) on GRU

As shown on table 4.10, all personality traits obtained good scores of 1-MAE. The highest score is achieved by Conscientiousness (0.89). Similar to LSTM results (table 4.8) and GRU (table 4.9), all personalities (Openness, Conscientiousness, Extroversion, Agreeableness and Neuroticism) apparently do not have correlation between predicted and ground-truth personalities.

	O	C	E	A	N
1-MAE	0.8746	0.89	0.8811	0.8722	0.8705
R	0.04	0	0.03	0.02	0.01
p-value	0.07	0.92	0.22	0.37	0.74

Table 4.10: Results of personality recognition using AUs (OpenFace) on Transformer

4.4 Personality Recognition using Head Pose and Face Movement Features

Concerning on head pose and face movement features, table 4.11 shows the comparison on each model. The highest score is obtained by GRU (0.8832), followed by LSTM (0.8814) and Transformer (0.8812). It achieved better score compared to

AUs by Openface, but not as good as the results of AUs by ME-GraphAU. To see the details in personality-wise, table 4.12, table 4.13, and table 4.14 show the results using LSTM, GRU, Transformer respectively.

Model	AVG 1-MAE
LSTM	0.8814
GRU	0.8832
Transformer	0.8812

Table 4.11: Results of personality recognition using head pose and face movement on all models

As shown on table 4.12, all personality traits obtained good scores of 1-MAE. The highest score is achieved by Agreeableness (0.8928). All personalities (Openness, Conscientiousness, Extroversion, Agreeableness and Neuroticism) have low positive correlation between predicted and ground-truth personality. All correlation coefficients are statistically significant.

	O	C	E	A	N
1-MAE	0.8841	0.8757	0.8775	0.8928	0.8770
R	0.10	0.11	0.07	0.07	0.09
p-value	<0.05	<0.05	<0.05	<0.05	<0.05

Table 4.12: Results of personality recognition using head pose and face movement on LSTM

As shown on table 4.13, all personality traits obtained good scores of 1-MAE. The highest score is achieved by Agreeableness (0.8942). All personalities (Openness, Conscientiousness, Extroversion, Agreeableness and Neuroticism) have low positive correlation between predicted and ground-truth personality. All correlation coefficients are statistically significant.

	O	C	E	A	N
1-MAE	0.8855	0.8777	0.8795	0.8942	0.8791
R	0.12	0.12	0.12	0.08	0.01
p-value	<0.05	<0.05	<0.05	<0.05	<0.05

Table 4.13: Results of personality recognition using head pose and face movement on GRU

As shown on table 4.14, all personality traits obtained good scores of 1-MAE. The highest score is achieved by Agreeableness (0.8939). All personalities (Openness,

Conscientiousness, Extroversion, Agreeableness and Neuroticism) apparently tend not to have correlation between predicted and ground-truth personality.

	O	C	E	A	N
1-MAE	0.8839	0.8732	0.8784	0.8939	0.8765
R	0.07	0.03	0.01	0.02	0.03
p-value	<0.05	0.18	0.57	0.44	0.15

Table 4.14: Results of personality recognition using head pose and face movement on Transformer

4.5 Personality Recognition using Fusion Feature

This section discuss the evaluation results using fusion feature, which is a combination of AUs (ME-GraphAU), head pose, and face movement. Table 4.15 shows the comparison on each model. The highest score is obtained by Transformer (0.8857), followed by LSTM (0.8819) and GRU (0.8806). To see the details in personality-wise, table 4.16, table 4.17, and table 4.18 show the results using LSTM, GRU, Transformer respectively.

Comparing to the previous results (table 4.4 and table 4.11), fusion feature did not improve the score. Head pose and face movement did not help AUs to make richer feature representation. Some possible reasons are camera matrix is currently not considered in calculating head pose and face movement and the database has very different camera angles, which introduce noise. As a result, these two types of feature are not discriminative enough.

Model	AVG 1-MAE
LSTM	0.8819
GRU	0.8806
Transformer	0.8857

Table 4.15: Results of personality recognition using fusion feature on all models

As shown on table 4.16, all personality traits obtained good scores of 1-MAE. The highest score is achieved by Agreeableness (0.8914). All personalities (Openness, Conscientiousness, Extroversion, Agreeableness and Neuroticism) have low positive correlation between predicted and ground-truth personality. All correlation coefficients are statistically significant.

	O	C	E	A	N
1-MAE	0.8856	0.8778	0.8777	0.8914	0.8770
R	0.21	0.19	0.18	0.11	0.18
p-value	<0.05	<0.05	<0.05	<0.05	<0.05

Table 4.16: Results of personality recognition using fusion features on LSTM

As shown on table 4.17, all personality traits obtained good scores of 1-MAE. The highest score is achieved by Agreeableness (0.8916). All personalities (Openness, Conscientiousness, Extroversion, Agreeableness and Neuroticism) apparently tend not to have correlation between predicted and ground-truth personality.

	O	C	E	A	N
1-MAE	0.8812	0.8747	0.8784	0.8916	0.8772
R	0.02	0.06	0.06	-0.04	0.02
p-value	0.33	<0.05	<0.05	0.12	0.44

Table 4.17: Results of personality recognition using fusion features on GRU

As shown on table 4.18, all personality traits obtained good scores of 1-MAE. The highest score is achieved by Agreeableness (0.8947). Extroversion have moderate positive correlation between predicted and ground-truth personality. The rest (Openness, Conscientiousness, Agreeableness and Neuroticism) have low positive correlation. All correlation coefficients are statistically significant.

	O	C	E	A	N
1-MAE	0.8888	0.8801	0.8835	0.8947	0.8816
R	0.28	0.25	0.30	0.17	0.27
p-value	<0.05	<0.05	<0.05	<0.05	<0.05

Table 4.18: Results of personality recognition using fusion features on Transformer

Chapter 5

Conclusion

In summary, I have reproduced the results of ME-GraphAU, retrained it on the combined database, and build the personality recognizer based on AUs, head pose, and face movement. The reproduction of ME-GraphAU went very well. It exceeded the expected target metric when executed on one database.

In terms of retraining ME-GraphAU on the combined database, the retrained model shows better generalization than I expected. A model that is trained on a combined database performed better than a model that is trained on a certain database and is evaluated on a different database.

When it comes to comparing AUs by ME-GraphAU with AUs by OpenFace, the results show that ME-GraphAU features have better representation convincingly. It is highly recommended for future studies consider this tool for AUs extraction.

In all experiments, evaluation results on all models and features obtained good scores. Thus, it is recommended to consider AUs (particularly by ME-GraphAU), head pose, and face movement in personality recognition study. In my experiments, training model on AUs by ME-GraphAU alone achieved the best result. The highest score is achieved by the GRU model. It suggested that a single AUs feature is enough.

Finally, head pose and face movement features did not noticeably contribute in this study. Although they generally can achieve good results, combining those features with AUs (ME-GraphAU) did not improve the results. One possible reason is calculating head pose and face movement by not considering the camera matrix will introduce noise because the database has very different camera angles.

Bibliography

- [1] Cheng Luo et al. “Learning Multi-dimensional Edge Feature-based AU Relation Graph for Facial Action Unit Recognition”. In: *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*. 2022, pp. 1239–1246.
- [2] Tadas Baltrusaitis et al. “OpenFace 2.0: Facial Behavior Analysis Toolkit”. In: *2018 13th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2018)*. 2018, pp. 59–66. DOI: 10.1109/FG.2018.00019.
- [3] Jianzhu Guo et al. “Towards Fast, Accurate and Stable 3D Dense Face Alignment”. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2020.
- [4] Yasmín Hernández et al. “Machine Learning Approach for Personality Recognition in Spanish Texts”. In: *Applied Sciences* 12.6 (2022). ISSN: 2076-3417. DOI: 10.3390/app12062985. URL: <https://www.mdpi.com/2076-3417/12/6/2985>.
- [5] Candida S. Punla et al. “Are we there yet?: An analysis of the competencies of BEED graduates of BPSU-DC”. In: *International Multidisciplinary Research Journal* 4.3 (Sept. 2022), pp. 50–59.
- [6] Zhe Wang et al. “Encoding Text Information with Graph Convolutional Networks for Personality Recognition”. In: *Applied Sciences* 10.12 (2020). ISSN: 2076-3417. DOI: 10.3390/app10124081. URL: <https://www.mdpi.com/2076-3417/10/12/4081>.
- [7] Leong Dickmond, Vazeerudeen Abdul Hameed, and Muhammad Ehsan Rana. “A Study of Machine Learning Based Approaches to Extract Personality Information from Curriculum Vitae”. In: *2021 14th International Conference*

- on *Developments in eSystems Engineering (DeSE)*. 2021, pp. 82–85. DOI: 10.1109/DeSE54285.2021.9719496.
- [8] Xiaodong Duan et al. “Personality Recognition Method based on Facial Appearance”. In: *2022 3rd International Conference on Computer Vision, Image and Deep Learning and International Conference on Computer Engineering and Applications (CVIDL and ICCEA)*. 2022, pp. 710–715. DOI: 10.1109/CVIDLICCEA56201.2022.9824658.
- [9] Tero Karras et al. “Progressive Growing of GANs for Improved Quality, Stability, and Variation”. In: *CoRR* abs/1710.10196 (2017). arXiv: 1710.10196. URL: <http://arxiv.org/abs/1710.10196>.
- [10] Cigdem Beyan et al. “Personality Traits Classification Using Deep Visual Activity-Based Nonverbal Features of Key-Dynamic Images”. In: *IEEE Transactions on Affective Computing* 12.4 (2021), pp. 1084–1099. DOI: 10.1109/TAFFC.2019.2944614.
- [11] Dairazalia Sanchez-Cortes et al. “A Nonverbal Behavior Approach to Identify Emergent Leaders in Small Groups”. In: *IEEE Transactions on Multimedia* 14.3 (2012), pp. 816–832. DOI: 10.1109/TMM.2011.2181941.
- [12] Víctor Ponce-López et al. “ChaLearn LAP 2016: First Round Challenge on First Impressions - Dataset and Results”. In: *Computer Vision – ECCV 2016 Workshops*. Ed. by Gang Hua and Hervé Jégou. Cham: Springer International Publishing, 2016, pp. 400–418. ISBN: 978-3-319-49409-8.
- [13] Mohsen Fallahnezhad, Mansour Vali, and Mehdi Khalili. “Automatic Personality Recognition from reading text speech”. In: *2017 Iranian Conference on Electrical Engineering (ICEE)*. 2017, pp. 18–23. DOI: 10.1109/IranianCEE.2017.7985447.
- [14] Florian Eyben, Martin Wöllmer, and Björn Schuller. “Opensmile: The Munich Versatile and Fast Open-Source Audio Feature Extractor”. In: *Proceedings of the 18th ACM International Conference on Multimedia*. MM ’10. Firenze, Italy: Association for Computing Machinery, 2010, 1459–1462. ISBN: 9781605589336. DOI: 10.1145/1873951.1874246. URL: <https://doi.org/10.1145/1873951.1874246>.

- [15] Bjorn Schuller et al. “The INTERSPEECH 2013 computational paralinguistics challenge: Social signals, conflict, emotion, autism”. In: *Proceedings INTERSPEECH 2013, 14th Annual Conference of the International Speech Communication Association*. Lyon, France, 2013, pp. 148–152.
- [16] Zhen-Tao Liu et al. “Speech Personality Recognition Based on Annotation Classification Using Log-Likelihood Distance and Extraction of Essential Audio Features”. In: *IEEE Transactions on Multimedia* 23 (2021), pp. 3414–3426. DOI: 10.1109/TMM.2020.3025108.
- [17] Gelareh Mohammadi and Alessandro Vinciarelli. “Automatic personality perception: Prediction of trait attribution based on prosodic features extended abstract”. In: *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*. 2015, pp. 484–490. DOI: 10.1109/ACII.2015.7344614.
- [18] Jelena Gorbova et al. “Integrating Vision and Language for First-Impression Personality Analysis”. In: *IEEE MultiMedia* 25.2 (2018), pp. 24–33. DOI: 10.1109/MMUL.2018.023121162.
- [19] Ádám Fodor et al. “Multimodal Sentiment and Personality Perception Under Speech: A Comparison of Transformer-based Architectures”. In: *Understanding Social Behavior in Dyadic and Small Group Interactions*. Ed. by Cristina Palmero et al. Vol. 173. Proceedings of Machine Learning Research. PMLR, 2022, pp. 218–241. URL: <https://proceedings.mlr.press/v173/fodor22a.html>.
- [20] Florian Eyben et al. “The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing”. In: *IEEE Transactions on Affective Computing* 7.2 (2016), pp. 190–202. DOI: 10.1109/TAFFC.2015.2457417.
- [21] Steffen Schneider et al. “wav2vec: Unsupervised Pre-training for Speech Recognition”. In: *CoRR* abs/1904.05862 (2019). arXiv: 1904.05862. URL: <http://arxiv.org/abs/1904.05862>.
- [22] Olivia Wiles, A. Sophia Koepke, and Andrew Zisserman. “Self-supervised learning of a facial attribute embedding from video”. In: *CoRR* abs/1808.06882 (2018). arXiv: 1808.06882. URL: <http://arxiv.org/abs/1808.06882>.

- [23] Jacob Devlin et al. “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”. In: *CoRR* abs/1810.04805 (2018). arXiv: 1810.04805. URL: <http://arxiv.org/abs/1810.04805>.
- [24] Yinhan Liu et al. “RoBERTa: A Robustly Optimized BERT Pretraining Approach”. In: *CoRR* abs/1907.11692 (2019). arXiv: 1907.11692. URL: <http://arxiv.org/abs/1907.11692>.
- [25] Fernando De la Torre and Jeffrey F. Cohn. “Facial Expression Analysis”. In: *Visual Analysis of Humans*. 2011.
- [26] Euclides N Arcoverde Neto et al. “Enhanced real-time head pose estimation system for mobile device”. In: *Integr. Comput. Aided Eng.* 21.3 (Apr. 2014), pp. 281–293.
- [27] Sepp Hochreiter and Jürgen Schmidhuber. “Long Short-Term Memory”. In: *Neural Comput.* 9.8 (1997), 1735–1780. ISSN: 0899-7667. DOI: 10.1162/neco.1997.9.8.1735. URL: <https://doi.org/10.1162/neco.1997.9.8.1735>.
- [28] Christopher Olah. *Understanding LSTM Networks*. 2015. URL: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/> (visited on 12/02/2022).
- [29] Kyunghyun Cho et al. “Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation”. In: *CoRR* abs/1406.1078 (2014). arXiv: 1406.1078. URL: <http://arxiv.org/abs/1406.1078>.
- [30] Ashish Vaswani et al. “Attention Is All You Need”. In: *CoRR* abs/1706.03762 (2017). arXiv: 1706.03762. URL: <http://arxiv.org/abs/1706.03762>.
- [31] Jay Alammar. *The Illustrated Transformer*. URL: <https://jalammar.github.io/illustrated-transformer/> (visited on 12/02/2022).
- [32] Baiyu Chen et al. “Overcoming Calibration Problems in Pattern Labeling with Pairwise Ratings: Application to Personality Traits”. In: *Computer Vision – ECCV 2016 Workshops*. Ed. by Gang Hua and Hervé Jégou. Cham: Springer International Publishing, 2016, pp. 419–432. ISBN: 978-3-319-49409-8.
- [33] Xing Zhang et al. “BP4D-Spontaneous: a high-resolution spontaneous 3D dynamic facial expression database”. In: *Image and Vision Computing* 32.10 (2014). Best of Automatic Face and Gesture Recognition 2013, pp. 692–706. ISSN: 0262-8856. DOI: <https://doi.org/10.1016/j.imavis.2014.06>.

002. URL: <https://www.sciencedirect.com/science/article/pii/S0262885614001012>.
- [34] S. Mohammad Mavadati et al. “DISFA: A Spontaneous Facial Action Intensity Database”. In: *IEEE Transactions on Affective Computing* 4.2 (2013), pp. 151–160. DOI: 10.1109/T-AFFC.2013.4.
- [35] Kaipeng Zhang et al. “Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks”. In: *IEEE Signal Processing Letters* 23.10 (2016), pp. 1499–1503. DOI: 10.1109/LSP.2016.2603342.

List of Figures

2.1	Methods for personality recognition	3
2.2	Several examples of action units [25].	4
2.3	Head pose (yaw, pitch, roll) [26]	6
2.4	68 points of face landmark	7
2.5	LSTM and GRU model [28]	8
2.6	Encoder-Decoder Transformer model [30]	8
4.1	Frame image transformation using face detection and alignment . . .	13

List of Tables

4.1	F1 score for reproduction of ME-GraphAU on BP4D dataset	14
4.2	F1 score for reproduction of ME-GraphAU on DISFA dataset	14
4.3	Evaluation of ME-GraphAU on cross dataset and combined dataset .	16
4.4	Evaluation of different sequence models on the FI dataset for the personality recognition task	16
4.5	Results of personality recognition using AUs (ME-GraphAU) on LSTM	17
4.6	Results of personality recognition using AUs (ME-GraphAU) on GRU	17
4.7	Results of personality recognition using AUs (ME-GraphAU) on Transformer	17
4.8	Results of personality recognition using AUs (OpenFace) on LSTM .	18
4.9	Results of personality recognition using AUs (OpenFace) on GRU . .	18
4.10	Results of personality recognition using AUs (OpenFace) on Transformer	18
4.11	Results of personality recognition using head pose and face movement on all models	19
4.12	Results of personality recognition using head pose and face movement on LSTM	19
4.13	Results of personality recognition using head pose and face movement on GRU	19
4.14	Results of personality recognition using head pose and face movement on Transformer	20
4.15	Results of personality recognition using fusion feature on all models .	20
4.16	Results of personality recognition using fusion features on LSTM . . .	21
4.17	Results of personality recognition using fusion features on GRU . . .	21
4.18	Results of personality recognition using fusion features on Transformer	21