Deep Dive into Disgust

YIWEN WANG* and YANFEI WANG*, Simon Fraser University, Canada

We comprehensively explore disgust as an affective phenomenon, covering the general definition, social and biological signals, research representations, appraisal descriptions, and other associated phenomena. Secondly, we focused on displaying various related datasets, including lab-controlled data, in-the-wild data, validated scientific datasets, and discrete media samples. The third part summarized the state-of-the-art research on detecting and synthesizing disgust and other basic emotions. In the last part, we bring up open questions and challenges faced.

Additional Key Words and Phrases: Disgust, emotion, facial expression, computation detection

1 GENERAL DESCRIPTION OF DISGUST

1.1 What is disgust

Disgust, considered one of the six basic emotions since Darwin, is defined as a profound sense of displeasure or revulsion triggered by encountering offensive, repulsive, or objectionable objects, individuals, or behaviors [1]. Initially rooted in distaste, which involves an oral rejection of unpleasant tastes, disgust has evolved to encompass a wider range of stimuli. It extends beyond its ancestral origins, manifesting not only in reactions to tangible, unpleasant substances but also in response to more abstract elements such as social or moral transgressions.

1.2 Signals of disgust

When feeling disgusted, people tend to show specific mouth-and-nose-centered movements, including a downward movement of the mouth, the uplifting of the upper lip, and a wrinkled nose. Sometimes, when the nausea is severe, a gag reflex can trigger an open mouth and a stuck-out tongue. As for the other part of the face, one may also lower their eyebrow and squint their eyes to make a grimace that expresses the displeasure feeling. To be more specific about the facial expression, a combination of AU4 and AU9//10 is often displayed [2]. Since the facial expression of disgust requires the lower half of the face to participate, the deformation of the mouth may affect the acoustics of speech. It has been proven that compared to natural speech, an expression of disgust needs a smaller mouth, thus lowering the formant frequencies [3].

As for physiological expressions, heart rate changes differ from different disgust stimuli. Gerlach et al. state that a lower heart rate is only observed when showing blood and physical injuries to the subjects [4]. Another experiment result shows that disgust only slightly lowers or does not change heart rate at all compared to the other 5 basic emotions [5]. The experiment result of Rohrmann et al shows that subjects' heart rates slightly increase while showing an amputation film, but decrease when showing a vomiting film [6]. fMRI result also shows such heterogeneous results of different kinds of stimuli affect different parts of the brain. The anterior insula responded to contamination and mutilation but not attacks, while the ventral visual areas responded to attacks and mutilations more strongly than contamination. The anterior insula exhibited a response to contamination and mutilation stimuli, but not to attacks. In contrast, the ventral visual areas demonstrated a stronger response to attacks and mutilations compared to contamination. [7].

1.3 Representation and classification

Though a basic emotion of disgust is, how different research theories represent it varies. In the Pleasure-Arousal-Dominance(PAD) space, disgust lies in the lower valence part due to its unpleasant nature. Other than this, it also triggers mild to average arousal and dominance. In the Plutchik wheel of emotion, disgust is described as a moderate emotion and is positioned as one of the primary emotions with the color green. Trust is at the opposite position of disgust, which suggests another extreme of the dimension. In Parrotts' emotion by groups, disgust is classified as a secondary emotion of anger, having revulsion and contempt, as its further subclasses [8]. The Emotion Annotation and Representation Language (EARL) proposed by HUMAINE placed disgust under the negative & forceful category with anger, annoyance, contempt, disgust, and irritation. All three kinds of models agree that disgust is a displeasure emotion and is associated with a moderate level of aversion.

1.4 Appraisal description and associated phenomena

Researchers widely agree that disgust functions as a mechanism for avoiding toxicity and illness, thereby serving a similar purpose across various cultures. However, this explanation addresses only one facet of "why people feel disgusted" – specifically, the "Consequence" in the OCC model – while the object and context aspects remain largely unexplored. The contentious issue surrounding disgust lies in the categorization of stimuli since it is a way of describing what is disgust. Table 1 illustrates different categorizations of disgust elicitors models over time.

The context in which disgust is experienced can vary significantly across cultures. For instance, in Chinese culture, the act of loudly slurping noodles is deemed impolite and may evoke disgust against the splattering soup and further possible contamination. In contrast, in Japanese culture, audibly sipping ramen is seen as a gesture of respect to the chef and is unlikely to cause discomfort. This highlights the cultural nuances that influence the perception and elicitation of disgust.

1.5 Induction and measurement methods

To study this emotion, certain lab-controlled inductions of disgust and valid measurement methods are needed. Let us look at induction methods first. The most commonly used induction method is displaying certain visionary content showing disgust elicitors, like pictures or video clips containing painful injuries, and bodily interaction with pus and feces [9]. Odors are another crucial stimulus that invokes disgust feelings, participants were asked to smell urine and rancid sweat [10]. Other times, participants will be asked

*Both authors contributed equally to this research.

Authors' address: Yiwen Wang, ywa422@sfu.ca; Yanfei Wang, ywa416@sfu.ca, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, Canada, V3J 2R9.

to imagine sorts of unpleasant things or experiences like "You are walking barefoot on concrete and you step on earthworms" [11].

After a desired amount of disgust has been triggered, a proper way of detecting and measuring the affective phenomenon. Self-reporting is probably one of the most common and direct ways of subjective measurement. Participants are asked to express their agreement with hypothetic scenarios like "It really bothers me when people sneeze without covering their mouths" [12]. A 25-item Disgust Scale-Revised [13] is used to gauge individual differences in how likely they are to experience disgust while facing different stimuli. Other than the facial Action Coding System(FACS) we mentioned before, facial electromyography(EMG) is used in more recent research to detect the contraction of the mouth muscles. Electrodes are placed at three facial muscles: the orbicularis oris, levator labii, and corrugator supercilli. The data is analyzed as the mean response of EMG [9]. Similarly, electrogastrography (EGG) monitors participants' gastric muscle contractions for disgust measurement [14]. Since disgust is related closely to nausea and vomiting, detecting stomach activity is also an efficient way of monitoring disgust.

2 SELECTED DATASETS OF DISGUST

2.1 FANE

FANE is an in-the-wild dataset for facial expression classification and emotion detection. It comprises static pictures of facial expressions collected from the internet and some expression databases. All images are manually labelled with emotion from nine categories: angry, confused, disgust, fear, happy, natural, sad, shy, and surprise. Figure 3 demonstrates some samples of disgust from FANE. The full dataset is available on kaggle.

The majority of FANE disgust dataset is real human expression, with a few samples being cartoon expressions(Fig. 3e). The real human samples vary in age (infant, adolescent, adult, elderly) and ethnicity (Caucasian, African, Asian), but mostly are Caucasian adults between the ages of 18 and 50. One notable pattern we can observe in the data is the scrunched grimace(Fig. 3a to 3d). Most of the samples frown, squint their eyes, and wrinkle their noses to express disgust, which is considered a mechanism of self-protection. We close our nasal passages by wrinkling our noses and shield our eyes by squinting, thus blocking out the odours and splash from the pathogenic stimuli. Some people exaggerate this expression by pinching their noses and covering their mouths (Fig. 3f and 3g). A gag reflex pattern with an open mouth and stuck-out tongue is also frequently observed in the dataset(Fig. 3h and 3i). Besides the signals of physical nausea and discomfort, we also see patterns of revulsion and hatred such as rolling eyes up and pulling down the corners of the mouth(Fig. 3l). There are some edge cases in which the expressions are more like anger or fear. For example, figure 3j shows an old lady staring straight ahead, which is a typical expression of anger. Figure 3k shows a man being shocked and fearful with a wide-open mouth and hands holding his head. At this moment, we are unable to determine whether these edge cases are genuinely signals of disgust or are wrongly labelled due to human mistakes. Despite the various expressions of disgust, we did not observe any significant pattern differences between age groups, race groups, or gender groups. The diversity of expressions we observe appears to be entirely individual.

2.2 RAVDNESS

RAVDNESS, which stands for *Ryerson Audio-Visual Database of Emotional Speech and Song*, is a lab-controlled database for emotion detection [15]. As the name suggests, the database comprises speech and songs in video-only, audio-only, and video-audio formats. The speech content was two simple neutral statements, *"Dogs are sitting by the door"* and *"Kids are talking by the door"*. Actors delivered the speech with eight emotions, including the six basic emotions, as well as neutral and calm as baseline emotions. Each emotion was performed at two intensity levels: normal and strong. We will not discuss the songs in this article because it does not include disgust emotion. The dataset was validated and rated by 247 untrained participants and had a high overall accuracy 80% for audio-video. The full dataset is available on zenodo.

The most common pattern of disgusted facial expression we can observe from RAVDNESS is the familiar grimacing we have seen in FANE. Social signals of anxiety also show up in the data. Some actors frequently blinked, moved their eyes, and looked away. Some would wave their head slightly but swiftly to express their dislike and denial. And some would breathe heavily during their speech due to rising heart rates. Interestingly, only one In terms of vocalization, the actors used a generally downward tone and low volume when performing at a neutral intensity. When it comes to the strong intensity, most actors would sound more agitated, with stronger rises and falls in pitch and much higher volume. Some actors will end their speech with a sigh like "Ew" and "Oh" to emphasize their disgust.

The dataset is homogeneous in terms of age and culture. The actors were all adults from the ages of 21 to 33. All actors were hired from Toronto, Canada and have English as their first language, which makes it difficult to observe differences in patterns across age groups and cultural backgrounds. It is worth noticing that just listening to the audio-only data makes it far more difficult to identify the disgust expression than looking at the video-only or the full audio video, probably because the speech content is neutral and there is no useful textual information. This also aligns with the rating results(75% for video-only and 60% for audio-only). Therefore, we can infer that visual signals are more important than audio signals for detecting disgust.

2.3 MELD

The *Multimodal EmotionLines Dataset* (MELD) is a multimodal dataset for emotion recognition in conversations [16]. MELD builds upon EmotionLines [17], a textual database with emotion labels, by incorporating video clips of the dialogues, thereby enriching the dataset with visual and audio modalities. MELD contains about 13,000 utterances from 1,433 dialogues collected from the famous TV series "Friend", and video clips for each utterance. Table 2 shows some sample utterances labelled as disgust from MELD. Figure 5 shows the scene of the corresponding video clips.

The utterances of disgust can be broadly categorized into four types by their content. The first one is directly expressing the disgust and aversion feeling of the talker, for example, *"He's like a big disgusting"*, and *"makes me wanna puke"*. The second is talking about the disgust stimuli, like *"You may wanna rethink the dirty underwear"*. The third type contains neutral or positive content, and it is difficult to recognize the genuine disgust emotion from these data. For example, at first glance at this utterance *"Great, now he's waving at me"*, we might think the sentiment is positive, but when combined with the visual data, audio data, and context from the video clips, it becomes clear that the character is expressing disgust. One common pattern of disgust in textual data is the use of exclamation words like *"Ew"* and *"Oh no"*. When the intensity of disgust increases, more exclamation words and even swear words appear in the dialogue, which becomes our fourth type. For example, in the scene shown in figure 5f, when the character is in extreme disgust, the content of the utterance is pure exclamation words, *"Ew! Ew! Ew! Ew! Ew! Ew!"*.

2.4 DIRTI

In most cases, we use not only social signals in expression but also the context and stimuli to recognize emotion. Therefore, studying what can induce disgust is also important. The *DIsgust-RelaTed-Images Database* (DIRTI) [18] is an emotion database of disgust stimuli. The database contains 240 static disgust-eliciting pictures and 60 neutral photos as the baseline. The photos are divided into six groups, food, animals, body products, injuries/infections, death, and hygiene. The dataset was validated by 200 participants by rating the intensity level of disgust and fear when seeing the picture. By observing the dataset and the validation results, we found that generally rotten and mouldy food is most likely to cause disgust. For disgust stimuli of high level, disgusting body products like faeces and phlegm are the most likely to cause disgust. Disgusting animals such as locusts and cockroaches are more likely to induce fear than disgust. The study also found female participants were more sensitive to disgust stimuli than male participants as they tended to report stronger feelings of disgust.

2.5 Summary of Observation

In terms of facial expression, the most common social signal of disgust seen in the datasets is a scrunched face with a wrinkled nose and squinted eyes. Signals of other emotions, such as anger and fear, can also occur in the expression of disgust. Regarding vocalization, people make low tones when they are in disgust but with low arousal. High volume and more rises and falls in tone are observed when people are in disgust at higher arousal. Exclamation and swear words are frequently seen in the textual content when people are at a high level of disgust. There is no significant difference between different groups of age, culture, and gender.

3 COMPUTATIONAL PROCESSING AND SYNTHESIS

For us humans, our brain considers multisensor information when detecting an affective phenomenon. Similarly, for computers to detect emotions like disgust, visual, acoustic, and language modalities are the most common inputs. Early research has concluded that models using multi-modality outperform their unimodal counterparts with an accuracy of 9.83% [19]. Thus Below we introduce 3 state-of-the-art research works that have been done within 10 years that could take all three main modalities into account when detecting emotions.

3.1 Emotion Detection

3.1.1 *Interpretable Dynamic Fusion Graph.*
Zadeh et al. [20] proposed an interpretable fusion model called Dynamic Fusion Graph (DFG) which could take in all three modalities mentioned above. They use CMU-MOSEI dataset, which consists of over 65 hours of YouTube videos with more than 1000 different solo speakers. The model takes in textual transcription, frames with bounding boxes, and COVAREP software extracted acoustic features. The underlying algorithm of this work is multiple Longshort Term Memory(LSTM) networks with an additional fusion component. Three separate subunits process three different modalities, and then every two pairs of modalities are combined to form bimodal dynamics. At last, three bimodal dynamics are fused to generate the final trimodal dynamic. All unimodal, bimodal, and trimodal are considered when doing the final classification task. Fig 1 shows that This model outperforms SOTA1 and SOTA2 in emotion detection on the same dataset with 85.2% overall F1 score. When it comes to detecting disgust, the F1 score is 76.6%. Disgust F1 score ranks third out of a total of 6 basic emotions.

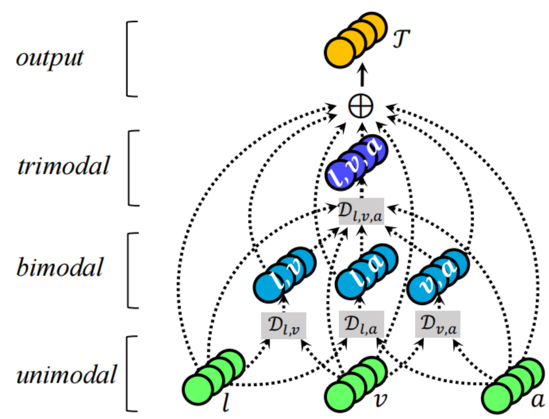


Fig. 1. The structure of Dynamic Fusion Graph (DFG) for three modalities of {(l)anguage, (v)ision, (a)coustic}. Dashed lines in DFG show the dynamic connections between vertices.

3.1.2 *Multi-head Attention Mechanism.*
Xi et al. [21] also developed another kind of deep learning modal that can detect all three kinds of modalities and yield an overall accuracy of 82.71% on the CMU-MOSI dataset, which also contains data originating from YouTube videos. A pre-trained VGG16 and LSTM are used to extract visual features, BERT is used to extract textual features, and a combination of spectrogram and convolutional neural network is used to extract audio features. After unimodal sentiment features are extracted, a

multi-head self-attention network is used to analyze the features. The correlation between every two pairs of unimodal features is also analyzed by another muti-head mutual attention network. At last, the output of these two parts is concatenated as the complete feature sent to make the final classification. Fig 2 shows the structure of the whole network. Unfortunately, a detailed detection accuracy for each sentiment is not revealed in this work.

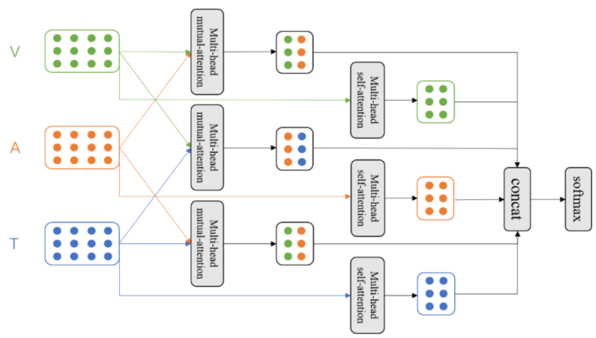


Fig. 2. The structure of Multi-Modal Multi-Head Attention (MMHA) model

3.1.3 Bi-Bimodal Modality Fusion. While the above two works use a take in all combinations of the three modalities, Han et al. [22] addressed that the importance of each modality is not the same, and textual information should be the most important one, thus only text-visual and text-acoustic bimodal should be considered. They also emphasized keeping mutual independence between each modality, thus the Bi-Bimodal Fusion Network(BBFN) they proposed contains a local regularizer that functions as a layer-wise feature space separator. Figure 7 shows the BBFN structure, which first uses Facet and Openface to extract visual features, BERT to extract textual features, and COVAREP to extract acoustic features. Then, it learns textual-related pairs to complement mutually. Lastly, all four representations are concatenated to give the final prediction. BBFN is doing quite well on negative//positive classification tasks with 86.2% accuracy on the CMU-MOSEI dataset, while on 7-class emotion detection, the accuracy falls to only 54.8%.

3.2 Emotion Synthesis

Other than detecting disgust and other affective phenomenon, emotion synthesis is also widely studied for the purpose of human-computer interaction, entertainment, and even therapeutics. Here we use a GAN-based model that can generate multiple human facial expressions as an example. Song et al. proposed a geometric guided generative adversarial network(G2GAN) that takes in neutral frontal-looking faces with no expressions and expressioned facial geometry heatmap as input and generates desired expressioned facial expressions with different intensities. The core component of this network is a residual generator that learns the residual between the neutral face and expressioned face as well as the the residual between authentic expressioned face and the synthesized expressioned face, and a dual-agent discriminator to distinguish between real images and generated images. Fig 8 shows synthesized disgust images with growing intensity. G2GAN is capable of generating compelling realistic images with desired complex identities and preserving expression-irrelevant attributes such as glasses. Quantitatively, the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) are used to measure the similarity between the synthesized images and the authentic ones on three different datasets. The average performance of SSIM is 0.867, and the average PSNR is 27.205.

4 OPEN DISCUSSION

Disgust is a special emotion overlooked by the affection study community. Unlike other basic emotions that have clear definitions and sufficient study, there is still a lot of confusion about disgust. There has been debate among psychologists as to whether disgust can be divided into physical and moral or only physical disgust exists. The prevailing view is there is core disgust (physical disgust) elicited by physical or chemical stimuli, which is from the human disease-avoidance mechanism. There is also moral disgust elicited by the violation of social norms to maintain sociality. Moral disgust is slightly different from physical disgust in terms of expression. Lee's study suggests that physical disgust is more to fear and moral disgust is more to anger [23]. People tend to exaggerate their expression of moral disgust for communication purposes because there is usually someone to blame for eliciting moral disgust. On the other hand, scientists including Bollard question that although disgust can be induced by non-physical stimuli like domestic violence and animal abuse, it is hard to prove that the genuine emotion people feel at the moment is moral disgust but not physical disgust [24]. In fact, diverse and bizarre stimuli can induce disgust and they can neither be categorized as physical contaminants nor moral violations. For example, some people with obsessive–compulsive disorder(OCD) can experience intense disgust and anxiety when seeing books not aligned. Misophonia patients can feel extreme disgust and irritation when confronted with human sounds like typing and chewing. A new study proposed that culture fusion can also stimulate disgust [25].

Probably due to the ambiguous definition of disgust, a huge insufficiency is seen in datasets of disgust. Despite the expressive differences between moral and physical disgust, there are few datasets specialized for disgust. Most of the time disgust appears as one of the six basic emotions, and its visual and textual data with various expressions are all assigned a simple label "Disgust". The benchmark datasets are also homogenous in ethnicity and age, being young adults in European and North American countries. This prevents us from further studying the effect of cultural differences on disgust. Meanwhile, most multimodal datasets are collected in the lab environment. The induced and acted emotions can be far away from the genuine ones. The high-quality and noise-free data is impossible to see in practical use cases. Consequently, state-of-the-art emotional recognition methods can not achieve satisfying performance when put into real-life applications. To address these issues, we need to conduct a more detailed categorization of the disgust emotion data and collect more spontaneous heterogeneous data approximate to real life for the disgust emotion study.

REFERENCES

[1] M. Pagel, *Encyclopedia of Evolution*, ser. Encyclopedia of Evolution. Oxford University Press, 2002, no. v. 1. [Online]. Available: https://books.google.ca/books?id=7MV9AQAACAAJ

[2] P. Ekman and E. Rosenberg, "What the face reveals: Basic and applied studies of spontaneous expression using the facial action coding system (facs)," 02 1998.

[3] C. S. Chong, J. Kim, and C. Davis, "Disgust expressive speech: The acoustic consequences of the facial expression of emotion," *Speech Communication*, vol. 98, 12 2017.

[4] A. Gerlach, G. Spellmeyer, C. Vögele, R. Huster, S. Stevens, G. Hetzel, and J. Deckert, "Blood-injury phobia with and without a history of fainting: Disgust sensitivity does not explain the fainting response," *Psychosomatic medicine*, vol. 68, pp. 331–9, 03 2006.

[5] R. Levenson, L. Carstensen, W. Friesen, and P. Ekman, "Emotion, physiology, and expression in old age," *Psychology and aging*, vol. 6, pp. 28–35, 03 1991.

[6] S. Rohrmann, "Cardiovascular indicators of disgust," *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, vol. 68, pp. 201–8, 07 2008.

[7] P. Wright, G. He, N. Shapira, W. Goodman, and Y. Liu, "Disgust and the insula: fmri responses to pictures of mutilation and contamination," *Neuroreport*, vol. 15, pp. 2347–51, 11 2004.

[8] W. G. Parrott, Ed., *Emotions in Social Psychology: Key Readings.* Psychology Press, 2000.

[9] A. Shenhav and W. Mendes, "Aiming for the stomach and hitting the heart: Dissociable triggers and sources for disgust reactions," *Emotion (Washington, D.C.)*, vol. 14, 11 2013.

[10] A. Gilbert, A. Fridlund, and J. Sabini, "Hedonic and social determinants of facial displays to odors," *Chemical Senses - CHEM SENSES*, vol. 12, pp. 355–363, 06 1987.

[11] J. Haidt, C. Mccauley, and P. Rozin, "Individual differences in sensitivity to disgust: A scale sampling seven domains of disgust elicitors," *Personality and Individual Differences*, vol. 16, pp. 701–713, 05 1994.

[12] J. Park, M. Schaller, and C. Crandall, "Pathogen-avoidance mechanisms and the stigmatization of obese people," *Evolution and Human Behavior*, vol. 28, pp. 410–414, 11 2007.

[13] B. Olatunji, N. Williams, D. Tolin, J. Abramowitz, C. Sawchuk, J. Lohr, and L. Elwood, "The disgust scale: Item analysis, factor structure, and suggestions for refinement," *Psychological assessment*, vol. 19, pp. 281–97, 09 2007.

[14] K. Meissner, E. Muth, and B. Herbert, "Bradygastric activity of the stomach predicts disgust sensitivity and perceived disgust intensity," *Biological psychology*, vol. 86, pp. 9–16, 09 2010.

[15] S. R. Livingstone and F. A. Russo, "The ryerson audio-visual database of emotional speech and song (ravdess): A dynamic, multimodal set of facial and vocal expressions in north american english," *PLOS ONE*, vol. 13, no. 5, pp. 1–35, 05 2018. [Online]. Available: https://doi.org/10.1371/journal.pone.0196391

[16] S. Poria, D. Hazarika, N. Majumder, G. Naik, E. Cambria, and R. Mihalcea, "Meld: A multimodal multi-party dataset for emotion recognition in conversations," 10 2018.

[17] S.-Y. Chen, C.-C. Hsu, C.-C. Kuo, K. Huang, and L.-W. Ku, "Emotionlines: An emotion corpus of multi-party conversations," 05 2018.

[18] A. Haberkamp, J. Glombiewski, F. Schmidt, and A. Barke, "The disgust-related-images (dirti) database: Validation of a novel standardized set of disgust pictures," *Behaviour Research and Therapy*, vol. 89, 11 2016.

[19] S. Poria, E. Cambria, R. Bajpai, and A. Hussain, "A review of affective computing: From unimodal analysis to multimodal fusion," *Information Fusion*, vol. 37, 02 2017.

[20] A. Bagher Zadeh, P. P. Liang, S. Poria, E. Cambria, and L.-P. Morency, "Multimodal language analysis in the wild: CMU-MOSEI dataset and interpretable dynamic fusion graph," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, I. Gurevych and Y. Miyao, Eds. Melbourne, Australia: Association for Computational Linguistics, Jul. 2018, pp. 2236–2246. [Online]. Available: https://aclanthology.org/P18-1208

[21] C. Xi, G. Lu, and J. Yan, "Multimodal sentiment analysis based on multi-head attention mechanism," in *Proceedings of the 4th International Conference on Machine Learning and Soft Computing*, ser. ICMLSC '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 34–39. [Online]. Available: https://doi.org/10.1145/3380688.3380693

[22] W. Han, H. Chen, A. Gelbukh, A. Zadeh, L.-p. Morency, and S. Poria, "Bi-bimodal modality fusion for correlation-controlled multimodal sentiment analysis," in *Proceedings of the 2021 International Conference on Multimodal Interaction*, 2021, pp. 6–15.

[23] S. Lee and P. Ellsworth, *Maggots and morals: Physical disgust is to fear as moral disgust is to anger1*, 01 2011.

[24] M. Bollard, "Is there such a thing as genuinely moral disgust?" *Review of Philosophy and Psychology*, vol. 13, pp. 1–22, 06 2022.

[25] B. Cheon, G. Christopoulos, and Y. Y. Hong, "Disgust associated with culture mixing: Why and who?" *Journal of Cross-Cultural Psychology*, vol. 47, pp. 1268–1285, 11 2016.

[26] P. Rozin and A. Fallon, "A perspective on disgust," *Psychological Review*, vol. 94, pp. 23–41, 01 1987.

[27] B. Olatunji, J. Haidt, D. Mckay, and B. David, "Core, animal reminder, and contamination disgust: Three kinds of disgust with distinct personality, behavioral, physiological, and clinical correlates," *Journal of Research in Personality*, vol. 42, pp. 1243–1259, 10 2008.

[28] J. Tybur, D. Lieberman, R. Kurzban, and P. DeScioli, "Disgust: Evolved function and structure," *Psychological review*, vol. 120, 12 2012.

Authors	Year of Publication	Elicitors	Comments
Rozin & Fallon [26]	1987	1. bitter and sour subjects 2. body products, certain foods, and certain animals 3. mutilation 4. interpersonal contamination 5. sociomoral disgust	-
Olatunji et al. [27]	2008	1. core disgust (e.g. prevent consuming certain kinds of subjects) 2. mutilation 3. contamination	Pure distaste in Rozin & Fallon's theory is not included for it is not considered as an emotion-related disgust
Tybur et al. [28]	2013	1. pathogen 2. sexual 3. sociomoral	Contamination is not included and there is a potential overlapping between the sexual and sociomoral category

Table 1. Disgust Elicitor Categorisation Models

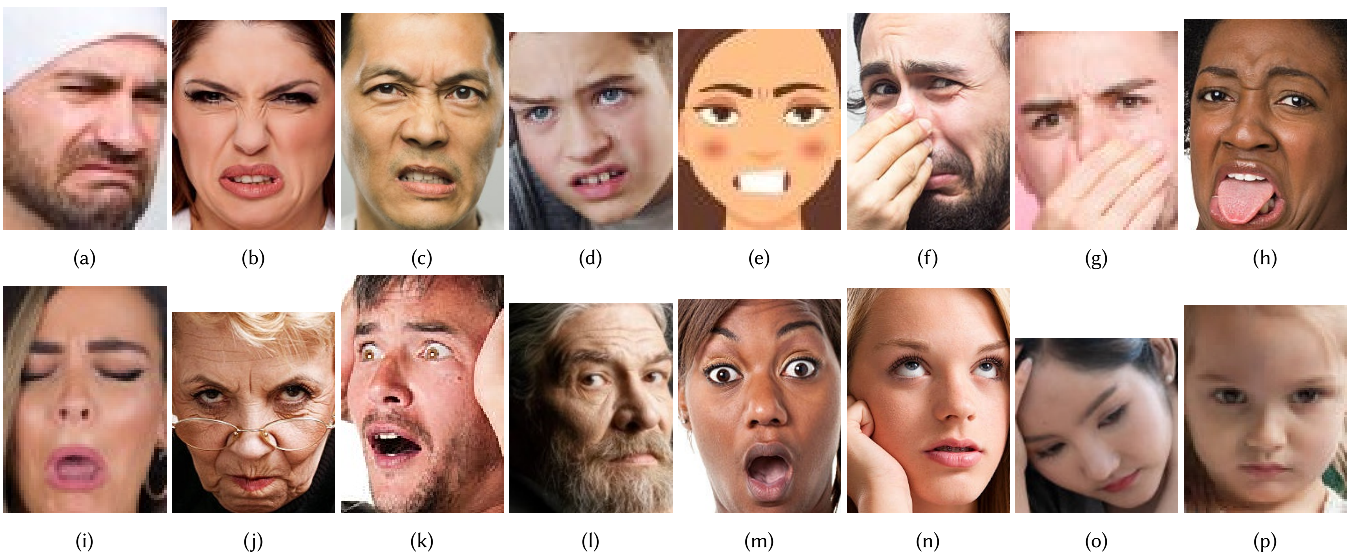


Fig. 3. Samples of disgust expression from FANE dataset. The full dataset is available at https://www.kaggle.com/datasets/furcifer/fane-facial-expressions-and-emotion-dataset

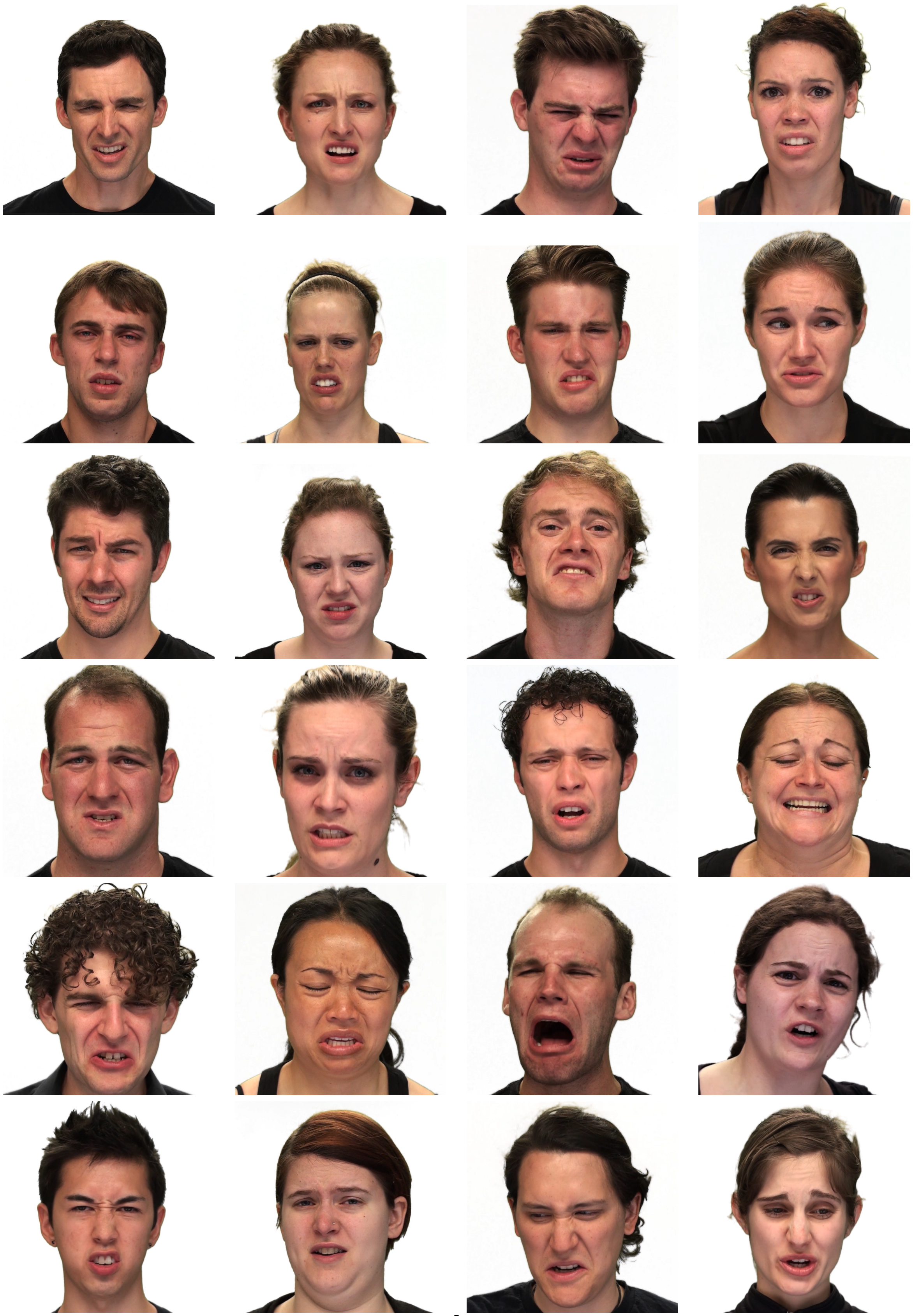
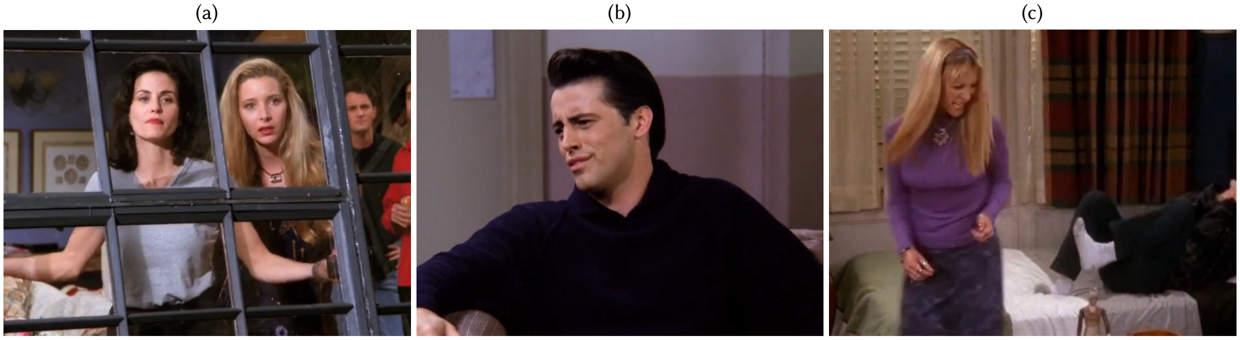


Fig. 4. Actors performing speech with disgust from RAVDNESS. The full dataset is available at https://zenodo.org/records/1188976.

Scene Fig	Utterance	Speaker	Clip ID
5a	Makes me wanna puke!	Monica	dia431_utt19
5b	He's like a big disgusting...	Rachel	dia398_utt12
5c	Well, you may wanna rethink the dirty underwear.	Chandler	dia111_utt17
5d	Great, now he's waving back.	Monica	dia149_utt9
5e	The Celtics? Ha. They couldn't hit a boat if...wait. They suck, alright?	Joey	dia300_utt8
5f	Ew! Ew! Ew! Ew! Ew! Ew!	Phoebe	dia321_utt10

Table 2. Sample utterances from MELD with disgust emotion label

(a) (b) (c)



(d) (e) (f)

Fig. 5. Scenes for the sample utterances from MELD displayed in Table 2. The full dataset is available at https://affective-meld.github.io/

Fig. 6. Sample images from DIRTI dataset. The full dataset is available at https://zenodo.org/records/167037.

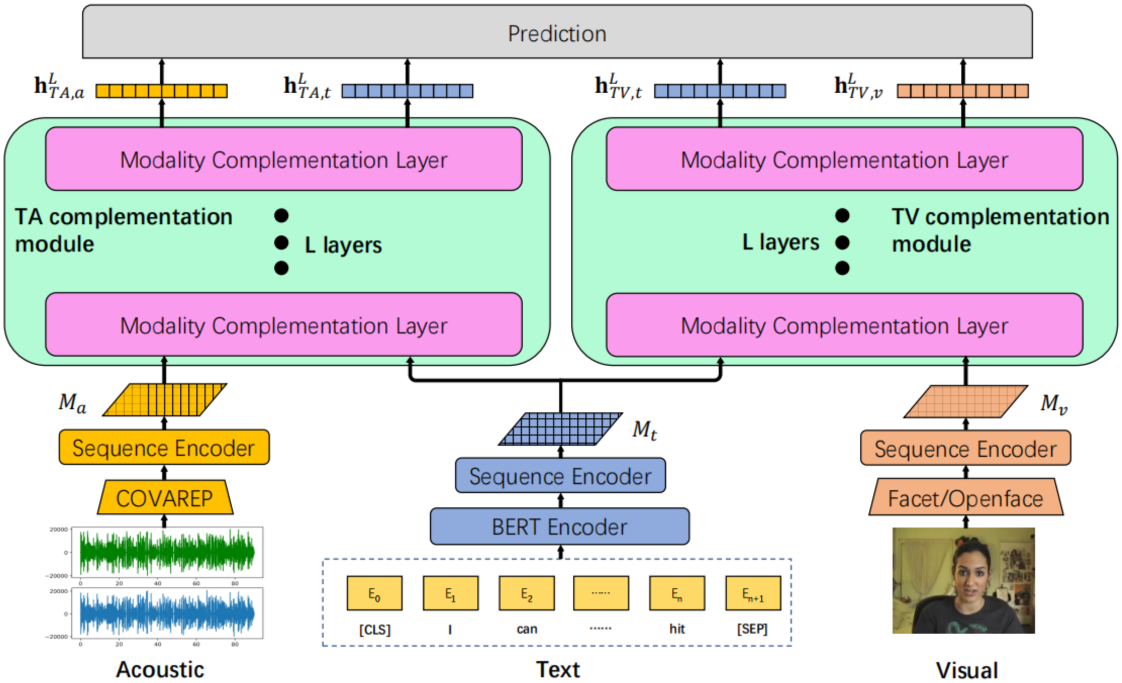


Fig. 7. The structure of Bi-Bimodal Fusion Network(BBFN) model

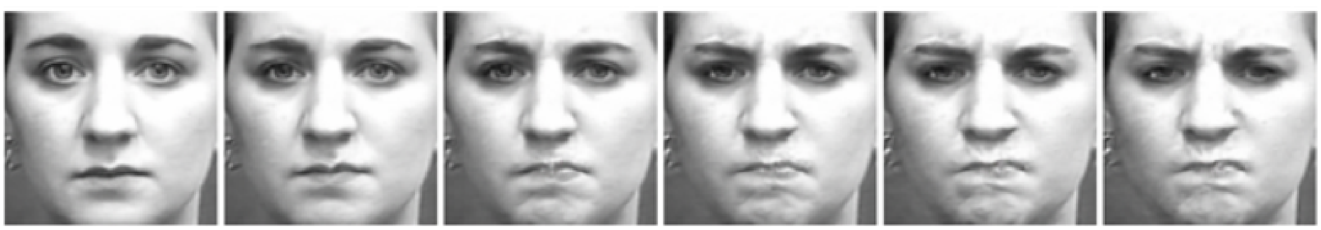


Fig. 8. Sythesized facial expression interpolation of disgust. Images in the left-most column are the source images, and the remainder are synthesized results.