Emo3D: Translating Emotion Descriptions to 3D Facial Expressions, Benchmark Dataset and Evaluation Metrics

Anonymous ACL submission

Abstract

Exisiting 3D facial emotion modeling have been constrained by limited emotion classes and insufficient datasets. This paper introduces "Emo3D", an extensive "Text-Image-Expression dataset" spanning a wide spectrum of human emotions, each paired with images and 3D blendshapes. Leveraging advanced Language Models (LLMs), we generate a diverse array of textual descriptions, facilitating the capture of a broad spectrum of emotional expressions. Using this unique dataset, we conduct a comprehensive evaluation of language model fine-tuning and CLIP-based models for 3D facial expression synthesis. We also introduce a new evaluation metric for this task to more directly measure the conveyed emotion. Our new evaluation metric, Emo3D, demonstrates its superiority over Mean Squared Error (MSE) metrics in assessing visual-text alignment and semantic richness in 3D facial expressions associated with human emotions. "Emo3D" has great applications in animation design, virtual reality, and emotional human-computer interaction.

1 Introduction

Automatic translation of character emotions into 3D facial expressions is an important task in digital media, owing to its potential to enhance user experience and realism. Facial Expression Generation (FEG) has a wide range of applications across various industries, including game development, animation, film production, and virtual reality. Previous studies in this domain have primarily focused on generating facial expressions for 2D or 3D characters, often relying on a limited set of predefined classes (Siddiqui, 2022) or driven by audio cues (Karras et al., 2017; Peng et al., 2023). However, there is a growing need for better control in the generation of complex and diverse human facial expressions. Recent studies (Zou et al., 2023; Zhong et al., 2023; Ma et al., 2023) have made notable progress in this area through the use of text prompts, offering a more direct approach to address the challenge of limited control that has been prevalent in earlier works (Siddiqui, 2022; Karras et al., 2017; Peng et al., 2023).

The primary issue with recent works using text prompts is **(i)** their limited focus on textual descriptions of emotions, as many studies have not deeply explored emotional context or offered a comprehensive solution that integrates both textual descriptions and 3d facial expression generation, creating a noticeable gap in the field (Zhong et al., 2023; Zou et al., 2023). Moreover, there is **(ii)** a scarcity of datasets containing emotional text alongside corresponding 3d facial expressions, impeding the development and training of facial expression generation (FEG) models for practical applications (Zhong et al., 2023; Zou et al., 2023; Ma et al., 2023). Additionally, **(iii)** the absence of reliable benchmarks and standardized evaluation metrics in this research area further complicates the assessment of FEG models.

Contributions: This paper tackles key challenges in Facial Expression Generation, focusing on generating expressions from textual emotion descriptions. Our contributions towards addressing the gaps in the field of Facial Expression Generation are as follows: **(i) Emo3d-dataset:** We present the Emo3D-dataset, specifically developed to bridge the gap between textual emotion descriptions and 3D facial expression generation. This dataset provides a rich compilation of annotated emotional texts alongside matching 3D expressions for effective training and assessment of FEG models. **(ii) Baseline Models:** We propose baseline FEG models as benchmarks for future research in this field. These models offer a reference point for evaluating new advancements and assessing progress. **(iii) Evaluation Metric:** To address the absence of standardized evaluation metrics in FEG, we introduce a new metric designed for the unique challenges of capturing the complexities of human emotions.

2 Related Work

Audio-based emotion extraction: Facial expression generation methods often utilize audio data, leveraging the semantic, tonal, and expressive qualities of voice for 3D generation. 'Audio-driven Facial Animation' (Karras et al., 2017) learns to map audio waveforms to 3D facial coordinates, identifying a latent code for expression variations beyond audio cues. 'EmoTalk' (Peng et al., 2023) focuses on creating 3D facial animations driven by speech, aligning expressions with both content

and emotion.

CLIP-based baselines: The utility of CLIP's language-and-vision feature space (Radford et al., 2021) in text-to-image generation has been highlighted in several works. MotionCLIP (Tevet et al., 2022) leverages CLIP for a feature space that accommodates dual modalities, enabling out-of-domain actions and motion integration into CLIP's latent space. The 4D Facial Expression Diffusion Model (Zou et al., 2023) uses this latent space for text-driven control. Also, ExpCLIP (Zhong et al., 2023) bridges text with facial expressions through semantic alignment.

Additionally, (Li et al., 2023) introduced CLIPER, a unified framework for both static and dynamic facial expression recognition, utilizing CLIP and introducing multiple expression text descriptors (METD) for fine-grained expression representations, achieving state-of-the-art performance by a two-stage training paradigm which involves learning METD and fine-tuning the image encoder for discriminative features.

Metrics: While numerous metrics exist for 2D image generation, a notable gap persists in effective metrics for 3D facial expression generation. R-precision, used to measure the alignment between input text and output image, was adopted by (Cong et al., 2023) following the approach in (Xu et al., 2017). This involved utilizing a CLIP model fine-tuned on the entire dataset, with R-Precision calculations based on (Park et al., 2021).

3 Dataset

We introduce the Emo3d-dataset, an assembly of 150,000 instances. Each instance comprises a triad: textual description, corresponding image, and blendshape scores created as follows:

(i) Emotion descriptions: For generating emotion-specific textual descriptions, we prompt GPT-3.5 (OpenAI, 2023), with a focus on eight principal emotions for this purpose. Secondly, we again hard-prompt GPT-3.5 language model to derive emotion distributions for these textual elements. This process entailed the construction of specialized prompts for GPT, leading to the generation of eight-dimensional vectors representing distinct emotion distributions, detailed in Figure 1.**a**. More details on the linguistic characteristics of the generated data is provided in the Appendix A.

(ii) 2D Image Generation: Subsequently, we utilize DALL-E 3 (Ramesh et al., 2022), a hierarchical image generation model, to create images that align with the generated textual descriptions.

(iii) Blendshape scores estimation: We employ Mediapipe framework (Lugaresi et al., 2019) to synthesize blendshape scores corresponding to the generated images based on textual descriptions.

Primitive Emotion Faces: Additionally, for intrinsic evaluation purposes, we construct a dataset of primitive emotions comprising singular emotion words, each paired with corresponding images that portray males and females exhibiting three distinct intensity levels of emotion. Utilizing Mediapipe (Lugaresi et al., 2019), we subsequently extract blendshape scores for the facial expressions depicted in these images. The emotional distributions associated with these individual words are derived using Emolex (Mohammad, 2018).

4 Method

4.1 Models

In this section, we propose several baseline models for the task of translating emotion descriptions into 3D facial expressions. This includes **(i)** fine-tuning of pre-trained language models, **(ii)** CLIP-based approaches, and **(iii)** Emotion-XLM, an architecture we have developed to enhance LM functionality.

Pretrained LM Baselines: We utilize BERT (Devlin et al., 2019) and Glot500, a highly multilingual variant of XLM-RoBERTa (ImaniGooghari et al., 2023), as the backbones. To map LM outputs into a designated target space, we incorporate a Multi-Layer Perceptron (MLP). During the training phase, we process textual data through the LMs to obtain encoded latent representations of the $[CLS]$ token. The MLP is then trained with tuples $T = (b, l) \mid b \in \mathbb{R}^{768}, l \in \mathbb{R}^{52}$, where b denotes the LM output and l represents the corresponding blendshape scores.

Emotion-XLM: In our architecture, we extend this MLP structure to XLM-RoBERTa as the backbone, introducing an additional emotion-extractor unit. Within this framework, we feed the transformer output into the emotion-extractor unit to extract the distribution of emotions alongside the corresponding one-hotted vector. Representing the input space as $B = \{b \mid b \in \mathbb{R}^{768}\}$, the emotion-extractor unit produces an output $E = \{(v, o) \mid v \in \mathbb{R}^8, o \in \mathbb{R}^8\}$, where v indicates emotion intensities in $V = \{[v_1, \ldots, v_8] \mid v_i \in [0, 1], i = 1, \ldots, 8\}$, and o is the one-hotted vector of v. Pairs of vectors are then passed to the MLP unit, where the vectors are concatenated with themselves and the text embedding before being fed to the regression unit. Consequently, the regression unit functions as $\mathbb{F}(.) : \mathbb{R}^{784} \rightarrow \mathbb{R}^{52}$. To enhance robustness in the regression phase, 50% of the ground truths of emotions are replaced with the outputs from the emotion-extractor unit.

$$\mathcal{L} = \lambda_1 \mathcal{L}_{Blendshape} + \lambda_2 \mathcal{L}_{Emotion} \quad (1)$$

Overall, for the training phase, the model is trained using triples of $T = (e, v, l) \mid e \in \mathbb{R}^{768}, v \in \mathbb{R}^8, l \in \mathbb{R}^{52}$, and two MSE losses over blendshapes and the extracted emotions are summed up with coefficients 1 and 5, respectively. This model is illustrated in Figure 1.**c**.

CLIP Baseline: We employed a Multi-Layer Perceptron (MLP) structure on the CLIP (Radford et al., 2021) backbone. What distinguishes this model from Pretrained LM Baselines is the incorporation of both image and text embeddings during training, effectively doubling the size of our dataset.

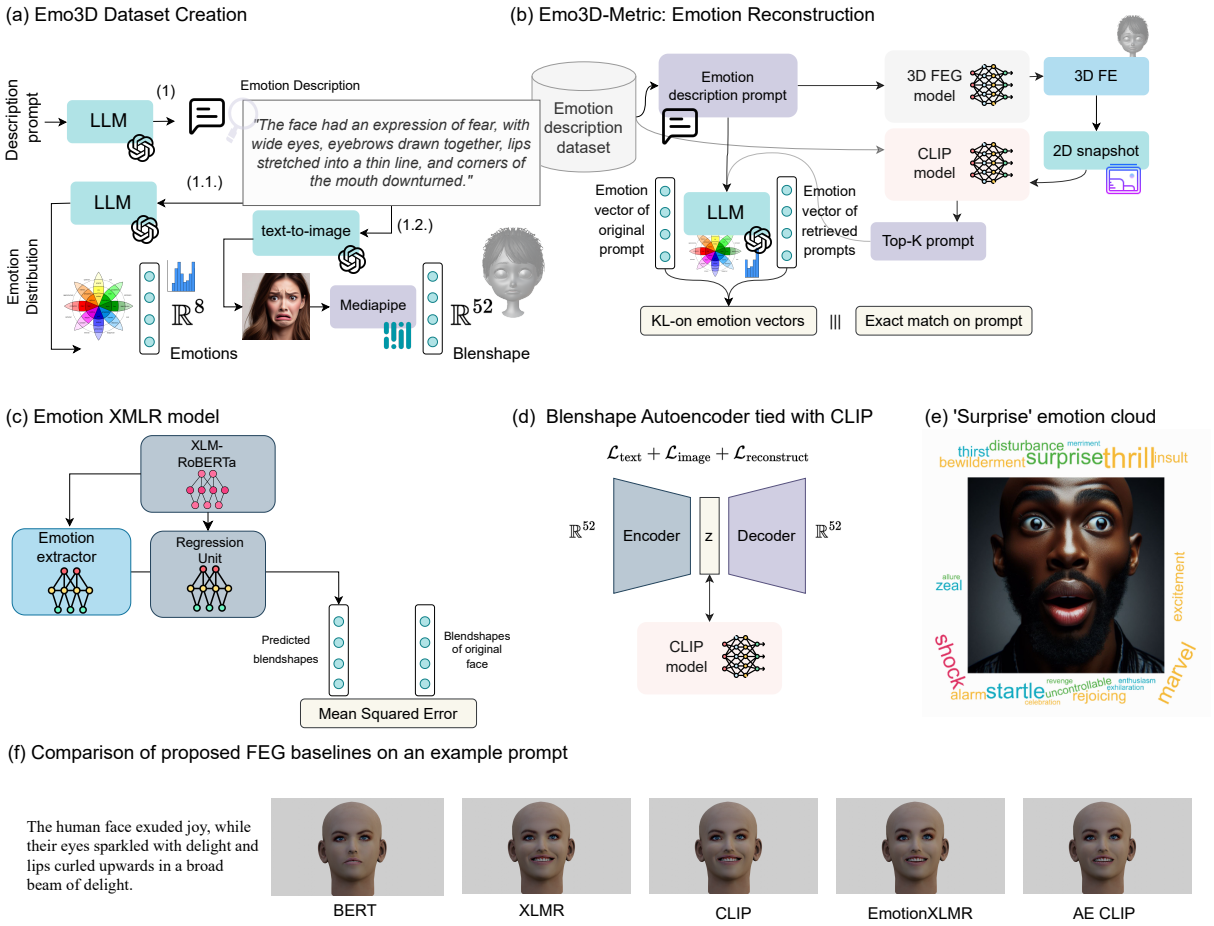

Figure 1: (a) **Emo3D-datasets Generation Pipeline**: Textual data describing human emotions is initially generated using GPT (OpenAI, 2023). We then utilize DALL-E models (Ramesh et al., 2022) to synthesize human faces. Each image undergoes face blendshape extraction using MediaPipe (Lugaresi et al., 2019). Furthermore, we employ GPT (OpenAI, 2023) to extract the emotion distribution for each prompt. (b) **Emo3D Metric**: Our methodology entails selecting n prompts with a balanced emotion distribution. We generate facial expressions using a text-utilizing FEG model for a given input prompt. We project the 3D face model onto a 2D image and employ zero-shot CLIP to identify the k nearest text prompts. Subsequently, we compute the Kullback-Leibler (KL) divergence between the emotion distribution of the input text and these k prompts. (c) **Emotion-XLM**: The model uses emotion ground truth to predict facial blendshapes. An Emotion Extractor guides the Regression model with the Teacher-Forcing technique at a 50% ratio. Both units are trained via Mean Squared Error (MSE) loss. (d) **Autoencoder CLIP**: The model concurrently reconstructs facial expressions while aligning their latent representation with corresponding text and image representations in the CLIP space. (e) **'Surprise' Emotion Word Cloud**: closest words to 'surprise' using Emolex (Mohammad, 2018) based on cosine similarity of emotion distribution. (f) **FEG model Comparison** for the proposed FEG baselines on an example prompt.

VAE CLIP We employed a Variational Autoencoder (VAE) to align blendshape scores with their corresponding text and image CLIP (Radford et al., 2021) embeddings, as illustrated in Figure 1**d**. The encoder maps blendshape scores to their respective text and image representations using cosine similarity, and the decoder sample from the latent space and reconstructs the blendshape scores. The reconstruction loss is defined by Mean Squared Error (MSE).

$$\mathcal{L}_{text} = 1 - \cos\left(CLIP_{text}, z\right) \quad (2)$$

$$\mathcal{L}_{image} = 1 - \cos\left(CLIP_{image}, z\right) \quad (3)$$

$$\mathcal{L} = \mathcal{L}_{recon} + \lambda_{text}\mathcal{L}_{text} + \lambda_{image}\mathcal{L}_{image} \quad (4)$$

Here, cos(a, b) denotes the cosine similarity between two vectors a and b.

Additionally, in the pursuit of a comprehensive comparison, we made diligent attempts to establish contact with the Expclip team through email and include the model in our comparison. Regrettably, as of the present moment, we have not received a response

from them.

4.2 Emo3D Metric

We introduce a new 3D Facial Expression Generation (FEG) metric for evaluating the reconstruction of the original emotion vector from 2D snapshots of the generated 3D faces. We create a test set comprising diverse emotion prompts uniformly selected. To evaluate any proposed FEG model, we generate the corresponding blenshape scores of the input text and project the 3D face model onto a 2D image. Using zero-shot CLIP (Radford et al., 2021), we identify the k-nearest text prompts related to the image. We calculate the emotion distribution for the original prompt and the top-K prompts. This is followed by computing the Kullback-Leibler (KL) divergence between the emotion vector of the original prompt and the average emotion vector of the top-K retrieved prompts. We refer to the normalized KL bounded between 0 and 1 as the "Emo3D metric":

$$D_{KL}(\phi \,||\, \bar{\psi}) = \sum_i \phi(i) \cdot \log\left(\frac{\phi(i)}{\bar{\psi}(i)}\right) \tag{5}$$

$$\text{Emo3D Metric} = \frac{1}{1 + e^{-D_{KL(\phi \,||\, \bar{\psi})}}} \tag{6}$$

where ϕ represents the emotion distribution of the input prompt, and $\bar{\psi}$ represents the mean emotion distribution of the top-k prompts. The steps for Emo3D calculation are outlined in Figure 1**b**. In our evaluation of the FEG models, we provide both the Emo3D Metric and the MSE scores of the 3D models for comparison purposes.

5 Results

The FEG model performances are provided in Table 1. It becomes evident that the CLIP With Regression Unit model demonstrates superior performance when evaluated using our Emo3D metric. Our results indicate that the MSE and Emo3D metrics do not consistently align. When we examined the 3D model outputs, we observed that samples that performed better according to Emo3D metric also demonstrated a closer visual resemblance to the input prompt, in contrast to samples that showed better performance based on MSE, similar to Figure 1**f**. This can be because in our metric, Emo3D prioritizes visual-text alignment and proximity, tending to capture richer semantic information than distance metrics in 3D space using MSE.

6 Conclusion

In this paper, we introduced 'Emo3D,' a comprehensive 'Text-Image-Expression dataset' that covered a wide range of human emotions and their textual descriptions, paired with images and 3D blendshapes. Our use of Language Models (LLMs) to generate prompts captured a variety of emotional expressions and descriptions. To the best of our knowledge, 'Emo3D' stood out as the most comprehensive FEG dataset, encompassing sufficiently diverse and complex emotional descriptions. Furthermore, we developed an efficient evaluation metric to provide 3D image synthesis models with a reliable benchmark. Throughout our work, we tested several unimodal and multimodal models as baselines to encourage new entrants to the field. The significance of 'Emo3D' lay in its potential to advance 3D facial expression synthesis, holding promising implications for animation, virtual reality, and emotional human-computer interaction.

Model	MSE	Emo3D
BERT	0.03	0.796
XLMRoBERTa	0.04	0.789
Autoencoder CLIP	0.002	0.776
Emotion-XLM	0.035	0.756
CLIP	0.014	0.737

Table 1: Performance comparison of FEG models using MSE vs. Emo3D metrics.

Comparison of Emo3D with existing datasets: Emo3d-dataset integrates textual, visual, and blendshape modalities, providing a more holistic representation of emotional expressions compared to single-modal datasets (Saravia et al., 2018; Mollahosseini et al., 2019; Chen et al., 2023). Our dataset comprises 90,000 images and 60,000 texts. It can also be employed for emotion recognition in text and images, thanks to the emotion distributions associated with each sample. The Emo3d-dataset shares similarities with other existing datasets, particularly TEAD(Zhong et al., 2023) and TA-MEAD(Ma et al., 2023), in terms of modality integration and a focus on emotional expressions.

The TA-MEAD(Ma et al., 2023) dataset, designed for 2D facial expression generation (FEG), provides emotion descriptions for videos, along with Action Unit (AU)(Ekman and Friesen) intensity annotations for each video. In contrast, our Emo3d-dataset offers a unique perspective by concentrating on textual emotion expressions, corresponding images, and blendshape scores.

The TEAD(Zhong et al., 2023) dataset, designed for 3D FEG, features situation descriptions, our Emo3d-dataset distinguishes itself by emphasizing emotion descriptions. Additionally, our dataset includes a distinctive feature with corresponding images for each text, providing a richer and more comprehensive resource. The Emo3d-dataset, comprising 150,000 samples, stands out significantly in scale when compared to ExpClip, which consists of 50,000 samples.

7 Limitations and future work

While our dataset exhibits positive attributes, it is not without errors stemming from the processes involved in its production. Specifically, the use of Mediapipe to

obtain blendshape sores introduced inaccuracies, particularly in the representation of certain emotions and facial expressions. To enhance the dataset in future endeavors, collaboration with skilled animators could be sought to refine and design more accurate blendshape scores.

References

Keyu Chen, Changjie Fan, Wei Zhang, and Yu Ding. 2023. 135-class emotional facial expression dataset.

Yuren Cong, Martin Renqiang Min, Li Erran Li, Bodo Rosenhahn, and Michael Ying Yang. 2023. Attribute-centric compositional text-to-image generation. *IEEE Trans. Pattern Anal. Mach. Intell.*

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Paul Ekman and Wallace V Friesen. Facial action coding system. *Environmental Psychology & Nonverbal Behavior*.

Ayyoob ImaniGooghari, Peiqin Lin, Amir Hossein Kargaran, Silvia Severini, Masoud Jalili Sabet, Nora Kassner, Chunlan Ma, Helmut Schmid, André Martins, François Yvon, and Hinrich Schütze. 2023. Glot500: Scaling multilingual corpora and language models to 500 languages. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1082–1117, Toronto, Canada. Association for Computational Linguistics.

Tero Karras, Timo Aila, Samuli Laine, Antti Herva, and Jaakko Lehtinen. 2017. Audio-driven facial animation by joint end-to-end learning of pose and emotion. *ACM Transactions on Graphics (TOG)*.

Sheng Li, Jinpeng Wang, Wei Zhao, Yucong Chen, and Kunpeng Du. 2023. Cliper: A unified vision-language framework for in-the-wild facial expression recognition. *arXiv preprint arXiv:2303.00193*.

Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris Mcclanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg, Matthias Grundmann, and Google Research. 2019. Mediapipe: A framework for building perception pipelines. *IEEE Trans. Vis. Comput. Graph.*

Yifeng Ma, Suzhen Wang, Yu Ding, Bowen Ma, Tangjie Lv, Changjie Fan, Zhipeng Hu, Zhidong Deng, and Xin Yu. 2023. Talkclip: Talking head generation with text-guided expressive speaking styles. *IEEE Trans. Circuit Syst. Video Technol.*

Saif M. Mohammad. 2018. Word affect intensities. In *Proceedings of the 11th Edition of the Language Resources and Evaluation Conference (LREC-2018)*, Miyazaki, Japan.

Ali Mollahosseini, Behzad Hasani, and Mohammad H. Mahoor. 2019. Affectnet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing*, 10(1):18–31.

OpenAI. 2023. Gpt-4 technical report.

Dong Huk Park, Samaneh Azadi, Xihui Liu, Trevor Darrell, and Anna Rohrbach. 2021. Benchmark for compositional text-to-image synthesis. *IEEE Trans. Image Process.*

Ziqiao Peng, Haoyu Wu, Zhenbo Song, Hao Xu, Xiangyu Zhu, Jun He, Hongyan Liu, and Zhaoxin Fan. 2023. Emotalk: Speech-driven emotional disentanglement for 3d face animation. *IEEE Trans. Circuit Syst. Video Technol.*

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning transferable visual models from natural language supervision. *Proceedings of Machine Learning Research*, 139:8748–8763.

Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. Hierarchical text-conditional image generation with clip latents.

Elvis Saravia, Hsien-Chi Toby Liu, Yen-Hao Huang, Junlin Wu, and Yi-Shin Chen. 2018. CARER: Contextualized affect representations for emotion recognition. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3687–3697, Brussels, Belgium. Association for Computational Linguistics.

J. Rafid Siddiqui. 2022. Explore the expression: Facial expression generation using auxiliary classifier generative adversarial network.

Guy Tevet, Brian Gordon, Amir Hertz, Amit H. Bermano, and Daniel Cohen-Or. 2022. Motionclip: Exposing human motion generation to clip space. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 13682 LNCS:358–374.

Tao Xu, Pengchuan Zhang, Qiuyuan Huang, Han Zhang, Zhe Gan, Xiaolei Huang, and Xiaodong He. 2017. Attngan: Fine-grained text to image generation with attentional generative adversarial networks. *IEEE Conf. Comput. Vis. Pattern Recog.*

Yicheng Zhong, Huawei Wei, Peiji Yang, and Zhisheng Wang. 2023. Expclip: Bridging text and facial expressions via semantic alignment.

Kaifeng Zou, Sylvain Faisan, Boyang Yu, Sébastien Valette, and Hyewon Seo. 2023. 4d facial expression diffusion model. *IEEE Trans. Pattern Anal. Mach. Intell.*

A Appendix

A.1 More analysis of Emo3D-dataset

To further explore the linguistic characteristics of each emotion category, this appendix presents four detailed tables. Table 2 presents statistical analyses of the dataset, offering further insights into its characteristics. Table 3 lists the most frequently occurring words within each category, providing insights into the vocabulary most closely associated with different emotional states. Tables 4 and 5 delve deeper into the semantic landscape of each emotion by showcasing the most frequent synsets (sets of words with similar meanings) within each category.

Emotion	Number of words	Number of unique words	Avg word len	Avg sentence len
Neutral	14805	1684	7.277609	13.555218
Happiness	31405	1519	6.687916	12.375832
Surprise	33690	1559	6.576492	12.152983
Sadness	32878	2220	6.816656	12.633311
Anger	16097	1288	6.419271	11.838541
Disgust	19917	1306	6.766280	12.532560
Fear	15120	1245	6.161111	11.322222
Contempt	7535	958	6.982349	12.964698

Table 2: Dataset Statistics by emotion Category

Neutral	Happiness	Surprise	Sadness	Anger	Disgust	Fear	Contempt
emotion	happiness	surprise	sadness	anger	contempt	fear	contempt
expression	eyes	eyes	eyes	eyes	expression	eyes	expression
confusion	smile	wide	expression	furrowed	disgust	wide	eyes
person	joy	mouth	downturned	expression	look	mouth	lips
furrowed	wide	emotion	emotion	lips	lips	expression	disdain
one	expression	open	mouth	brow	eyes	furrowed	look
eyes	bright	eyebrows	deep	rage	mouth	lips	emotion
hint	expressing	raised	person	narrowed	furrowed	pale	mouth
random	emotion	shock	lips	eyebrows	nose	open	feeling
look	person	slightly	sorrow	brows	disdain	look	sneer

Table 3: Most Frequent Words for each emotion

Neutral	Happiness	Surprise	Sadness
demonstration.n.05	feeling.n.01	astonishment.n.01	area.n.01
cognitive state.n.01	communication.n.02	feeling.n.01	feeling.n.01
combination.n.07	area.n.01	demonstration.n.05	sadness.n.01
confusion.n.02	positive stimulus.n.01	cognitive state.n.01	negative stimulus.n.01
communication.n.02	demonstration.n.05	combination.n.07	region.n.01
feeling.n.01	collection.n.01	emotion.n.01	sagacity.n.01
countenance.n.01	emotional state.n.01	communication.n.02	unhappiness.n.02
sagacity.n.01	sagacity.n.01	sagacity.n.01	countenance.n.01
small indefinite quantity.n.01	facial expression.n.01	rejoinder.n.01	communication.n.01
hair.n.01	countenance.n.01	hair.n.01	rejoinder.n.01

Table 4: Most frequent synsets for each emotion

Anger	Disgust	Fear	Contempt
feature.n.02	dislike.n.02	fear.n.01	dislike.n.02
anger.n.01	demonstration.n.05	feature.n.02	demonstration.n.05
communication.n.02	area.n.01	emotion.n.01	area.n.01
sagacity.n.01	facial expression.n.01	sagacity.n.01	hair.n.01
countenance.n.01	communication.n.02	anxiety.n.02	communication.n.02
demonstration.n.05	region.n.01	countenance.n.01	region.n.01
communication.n.01	rejoinder.n.01	communication.n.01	disrespect.n.01
hair.n.01	countenance.n.01	rejoinder.n.01	countenance.n.01
rejoinder.n.01	communication.n.01	hair.n.01	communication.n.01
feeling.n.01	disrespect.n.01	appearance.n.01	rejoinder.n.01

Table 5: Most Frequent synsets for each emotion