Multimodal Compositional Learning for Diseased Face Generation

1 General Info

Project Title: Multimodal Compositional Learning for Diseased Face Generation
Supervisors: Azade Farshad, Yousef Yeganeh
Contact Email: azade.farshad@tum.de, y.yeganeh@tum.de

2 Background and Motivation

Facial image generation and manipulation have significant applications in biomedicine, particularly for visualizing the impact of diseases on facial features, which is particularly challenging due to the scarcity and privacy of training data [5]. Profiling such attributes in images and synthesizing new samples could significantly impact healthcare by enabling early diagnosis and disease monitoring [9]. Therefore, there is a need for novel approaches that can leverage existing images along with prior medical information like medical knowledge graphs (KGs) [1], which are structured representations of biomedical concepts and their relations. Compositional learning [8] can represent visual attributes along with the rest of prior knowledge in a multimodal setting. For example, a medical KG can indicate that Down syndrome is a genetic disorder likely to cause a flattened appearance to the face, almond-shaped eyes, and a short neck.

3 Project Abstract

The aim of this project is to develop a system that can not only generate realistic facial images with various disease categories but also evaluate the real and synthetic images based on the learned profiles of the diseases and identify the severity and progression of these diseases. We propose to first learn the disease profiles based on negative (healthy [4]) and positive (diseased [2]) samples as well as medical knowledge graphs [1] and large language models [10]. These profiles will describe the facial attributes of people with different diseases at different ages. We will then use existing generative models such as Stable Diffusion [7] and condition them on the disease profiles. The project aims to produce facial images that reflect the existence and progression of diseases at different stages [6] or transform healthy faces to diseased ones and vice versa, which will be evaluated against real-world face disease datasets [3, 2] using our evaluation method.

4 Technical Prerequisites

- Good background in machine learning and deep learning
- Experience in PyTorch

- Experience in Python
- Experience with Generative Models

5 Benefits

- Weekly supervision and discussions
- Possible novelty of the research
- The results of this work are intended to be published in a conference or journal

6 Work packages and Time-plan

	Description	# Students
WP1	Familiarizing with the literature.	4
WP2	Implementing the baselines	4
WP3	Improving the baselines and validation on relevant datasets	4
	Midterm Presentation	4
WP4	Implementing the model	4
WP5	Finalizing the results and evaluation	4
	Final Presentation	4

Table 1: Project Timeline

References

[1] Payal Chandak, Kexin Huang, and Marinka Zitnik. Building a knowledge graph to enable precision medicine. *Nature Scientific Data*, 2023.

[2] Bo Jin. Disease-specific faces, 2020.

[3] Bo Jin, Leandro Cruz, and Nuno Gonçalves. Deep facial diagnosis: Deep transfer learning from face recognition to facial diagnosis. *IEEE Access*, 8:123649–123661, 2020.

[4] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Large-scale celebfaces attributes (celeba) dataset. *Retrieved August*, 15(2018):11, 2018.

[5] Anwesha Mohanty, Alistair Sutherland, Marija Bezbradica, and Hossein Javidnia. High fidelity synthetic face generation for rosacea skin condition from limited data. *arXiv preprint arXiv:2303.04839*, 2023.

[6] Chi Nhan Duong, Khoa Luu, Kha Gia Quach, and Tien D Bui. Longitudinal face modeling via temporal deep restricted boltzmann machines. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5772–5780, 2016.

[7] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.

[8] Ugur Sahin, Hang Li, Qadeer Khan, Daniel Cremers, and Volker Tresp. Enhancing multimodal compositional reasoning of visual language models with generative negative mining. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 5563–5573, 2024.

[9] Kuan Wang and Jiebo Luo. Detecting visually observable disease symptoms from faces. *EURASIP Journal on Bioinformatics and Systems Biology*, 2016:1–8, 2016.

[10] Michihiro Yasunaga, Antoine Bosselut, Hongyu Ren, Xikun Zhang, Christopher D Manning, Percy S Liang, and Jure Leskovec. Deep bidirectional language-knowledge graph pretraining. *Advances in Neural Information Processing Systems*, 35:37309–37323, 2022.