Generating Acne Images for Simulated Multimodal N-of-1 Trials

Advanced Machine Learning Seminar (WS 2023/2024)

Manuel Hettich

Hasso Plattner Institute for Digital Engineering
manuel.hettich@student.hpi.uni-potsdam.de

This project explores the training and evaluation of different generative models designed to create identity-preserving facial images annotated with acne severity levels based on a study protocol incorporating time dependencies. Leveraging the capabilities of deep generative models, including Diffusion Models, our work aims to address the need for a more nuanced and multimodal representation of N-of-1 trial outcomes in dermatological studies to train and evaluate unsupervised models on image categorization. By generating high-fidelity images that accurately reflect varying degrees of acne severity, this project contributes to the broader field of medical image synthesis, offering a novel approach for enhancing the visual datasets available for medical research and education.

1 Introduction

Visual representations play a crucial role in medical diagnoses, offering invaluable insights into patient conditions that are often challenging to quantify through scalar measures alone. Among various conditions, acne vulgaris stands out due to its widespread prevalence and the subjective nature of its severity assessment. Traditional approaches to evaluating acne severity take a lot of manual effort and rely heavily on scalar outcomes derived from clinical observations.

In recent years, generative models, particularly Generative Adversarial Networks (GANs) and Diffusion Models, have shown great promise in their ability to create realistic images across various domains, including medical imaging. These models offer a pathway to generating detailed, accurate representations of medical conditions, such as acne, in a controlled and scalable manner with fewer privacy concerns compared to real-life images while preserving their medical accuracy. This project is centered around the exploration and application of these advanced generative models to create a dataset of facial images that reflect a range of identity-preserving acne severity levels, guided by predefined study protocols. Building on this project, the ultimate goal is to improve unsupervised learning models capable of rating patient images in order to automatically evaluate acne severity levels in clinical trials and research.

1.1 Contributions

Throughout the project, my supervisors, Juliana Schneider and Thomas Gärtner, provided unwavering support, ensuring regular updates were maintained every few weeks to track our progress and overcome any new challenges. Their guidance was instrumental in navigating the complexities of the project. Additionally, I would like to express my gratitude to the course organizer, Sumit Shekhar, without whom this project wouldn't have been possible.

2 Context

The generation of medical images through computational models presents a unique set of challenges and opportunities. On one hand, the demand for extensive, diverse, and accurate medical imaging datasets is ever-increasing, driven by the need for more sophisticated diagnostic tools and the development of AI-based analysis techniques. On the other hand, the creation and compilation of these datasets are hindered by privacy concerns, the scarcity of cases for certain conditions, and the potential for inherent biases and lack of diversity in the data collected.

Generative models, such as GANs and, more recently, Diffusion Models, have emerged as powerful tools for synthesizing high-fidelity images. These models have the potential to augment existing datasets or create new ones from scratch, thereby overcoming the limitations posed by traditional data collection methods. In the domain of dermatology, specifically in the study of acne severity, these models can be adapted to generate facial images that accurately mimic a wide spectrum of acne, ranging from mild to very severe.

2.1 Background

Acne is a common skin condition characterized by the appearance of pimples, blackheads, whiteheads, and sometimes deeper cysts or nodules, primarily on the face, chest, and back. It results from the clogging of hair follicles with oil and dead skin cells, leading to inflammation and bacterial infection. While acne is most prevalent among teenagers due to hormonal changes, it can affect people of all ages, causing not only physical discomfort but also emotional distress due to its impact on appearance. Effective treatment varies, ranging from topical creams to oral medication, depending on the severity and type of acne.

2.2 Multimodal N-of-1 trials

In an advance to personalize healthcare, Fu et al. introduce a framework that incorporates multimodal N-of-1 trials, utilizing deep learning models and statistical inference to analyze health outcomes assessed through images, audio, or video data collected by trial participants on mobile devices [2]. This approach is demonstrated

through a series of trials assessing acne cream effectiveness, where convolutional neural networks (CNNs) and linear mixed models are employed to analyze image-based outcomes, showcasing the potential of multimodal trials in personalized healthcare.

Schneider et al. further this exploration by proposing a fully automated approach to analyze multimodal N-of-1 trials without the need for manual labels, offering a solution to the scalability challenges posed by the need for expert labeling in multimodal outcome analysis. [8]. Their methodology combines unsupervised learning models, specifically autoencoders, with statistical inference to identify treatment effects from image data. By creating lower-dimensional embeddings of the images and then applying principal component analysis (PCA), they reduce the data to a single dimension suitable for statistical testing. This unsupervised pipeline is tested against the same series of acne cream trials, validating the effectiveness of the unsupervised approach in replicating the treatment effect identified in one participant through the expert analysis.

2.3 ACNE04

The ACNE04 dataset by Xiaoping Wu et al.[10] significantly advances acne severity grading with 1,457 images and over 18,983 lesions annotated by dermatologists. The labels for each image were derived using a professional grading criterion, considering the relationship between lesion counts and acne severity. This dataset not only highlights the relationship between lesion count and severity but also supports the development and evaluation of machine learning models for dermatological analysis and automatic acne grading.

3 Related Work

Zijia Lu and Sri Krishnamurthy developed SkinGAN, using Generative Adversarial Networks (GANs) to generate facial images for acne diagnosis [6]. The architecture is designed to create images across acne severity levels with small patches positioned through landmark identification with MediaPipe and a triangulation algorithm (see figure 1). The training process involves data preprocessing, learning acne features, and iterative quality improvements. SkinGAN's performance was evaluated through visual comparison and quantitative metrics to ensure accurate representation of acne severities.

Hazem Zein et al. utilized StyleGAN algorithms to generate a synthetic dataset of faces with annotated acne severities, aiming to overcome dataset scarcity and privacy issues [12]. The process included dataset curation, preprocessing, and transfer learning with a pre-trained model on celebrity faces, adjusted to include acne features. This method produced a variety of realistic acne severity images, which were used to train and evaluate a CNN-based classifier, demonstrating GANs' potential in creating medically relevant imaging datasets. Unfortunately,

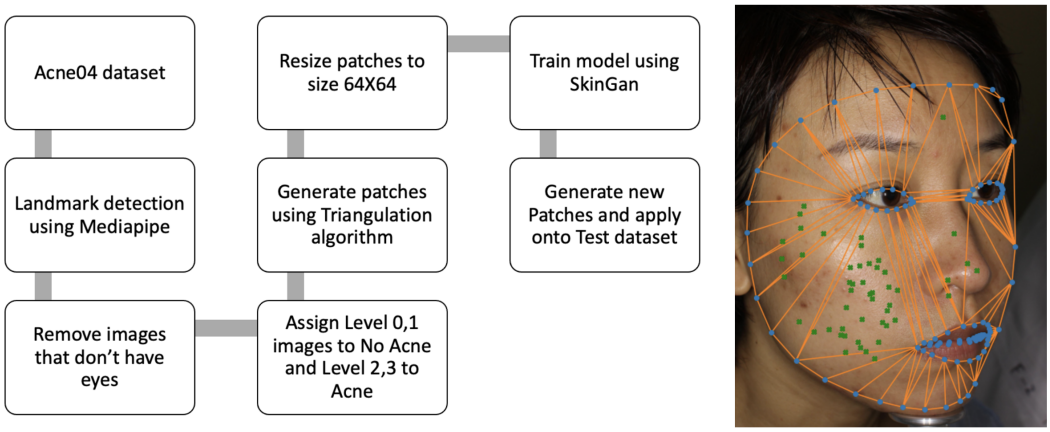

Figure 1: An illustration of the workflow of SkinGAN to generate new patches on an input image using MediaPipe and a triangulation algorithm.

the authors did not publish their model's codebase so it could not be used as part of this project.

4 Problem

Despite significant advancements in generative models for medical imaging, there remain substantial gaps in their application, particularly in dermatology. One primary challenge lies in the generation of images that accurately reflect the wide range of acne severities, a task complicated by the nuanced differences between severity levels and the subjective nature of human assessment. Additionally, ensuring that generated images are identity-preserving across different severity levels, diverse and representative of the global population presents a further challenge, highlighting the need for a generative model that can synthesize facial images in a controlled, yet varied manner. This project aims to address these issues by developing a generative model capable of producing high-quality images annotated with specific acne severities.

5 Solution

The project aimed to generate images portraying different levels of acne severity, employing advanced image generation techniques. Initially, the focus was on adapting the existing SkinGAN model to create images with variable acne severity by generating multiple patches on a single input image.

Given some limitations encountered with SkinGAN, the project shifted towards exploring stable diffusion models (SD) for their potential in generating more realistic and varied images. SD models, particularly known for their ease of finetuning

and well-documented popular codebases, represent the cutting edge in current image generation technology. The decision to pivot to SD models was driven by their superior capability to produce high-quality images and the availability of extensive resources to facilitate model training and adaptation.

6 Implementation

The implementation phase involved two primary efforts: attempting to modify SkinGAN for broader application and experimenting with stable diffusion models to generate acne severity images. The modification of SkinGAN aimed to expand its functionality from generating single acne patches to applying multiple patches on input images to simulate varying acne severity levels [3]. Despite technical adjustments to the model's code, the generated patches lacked cohesion, resulting in images that did not meet the project's realism criteria although the displayed person's identity was well-preserved (see figure 2).

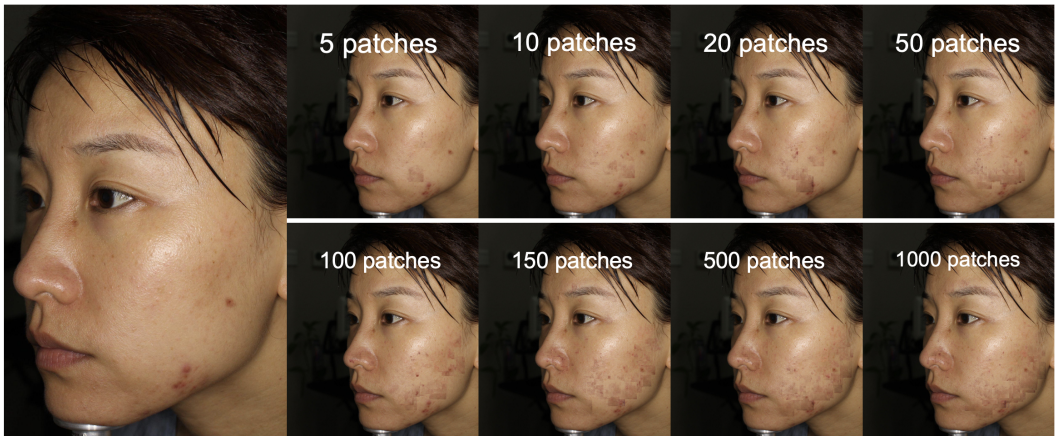

Figure 2: The original SkinGAN codebase was adapted to generate multiple patches on a single input image to simulate different severity levels [3]. The overlapping patches create visible artifacts, especially on faces with lighter skin tones.

Consequently, the focus shifted to stable diffusion models, starting with attempts to prompt existing SD base models to generate images with different acne severity levels [1, 9, 11]. Both direct generation and img2img methods were explored, including using augmented pictures from SkinGAN as input. These initial attempts were largely unsuccessful, with the models either omitting acne features entirely or applying unnatural overlays to the images. Yet, these insights were helpful to select the most promising SD base models for finetuning.

The project then adopted DreamBooth, a technique for fine-tuning text-to-image diffusion models, to create images of a specific subject under varying contexts [7]. A manually filtered version of the ACNE04 dataset removing any images with strongly visible watermarks was used to fine-tune one SD1.5 for all severity levels and four separate SDXL models (each trained on images of only one severity level) [4, 5]. Finally, a program was developed that could ingest a CSV file describing an N-of-1 study protocol with multiple patients and varying acne severity levels over time and generate all required images[5].

Figure 3: Using a text prompt to generate multiple synthetic images with an SDXL model fine-tuned on very severe acne cases.

Figure 4: The SD1.5 model fine-tuned on all acne severity levels at once was not able to differentiate well between mild acne (generated image on the left) and very severe acne (generated image on the right).

7 Evaluation

The evaluation of the generated images focused on assessing their realism and the accuracy of acne severity representation. While the bigger SDXL models - each fine-tuned for one specific severity level - produced realistic images, their image inference times with only text prompts were slow with up to 20s per image using

a single T4 GPU. The SD1.5 model, trained with images across all severity levels and corresponding embedding tokens, yielded faster but unpredictable outcomes, often not aligning with the prompted severity levels (see figure 4. Applying several optimization steps, the img2img inference times for SDXL image generation was reduced to 5s per image.

Img2img results with SD models were also evaluated but did not meet expectations when a random input image of a full-face portrait was used. The generated images either lacked acne features or introduced unnatural overlays, and the subject's appearance sometimes changed significantly with longer inference times. However, these outcomes improved significantly when an image with mild acne was randomly chosen from the ACNE04 dataset and used as an input with only very few inference steps applied (see figure 5).

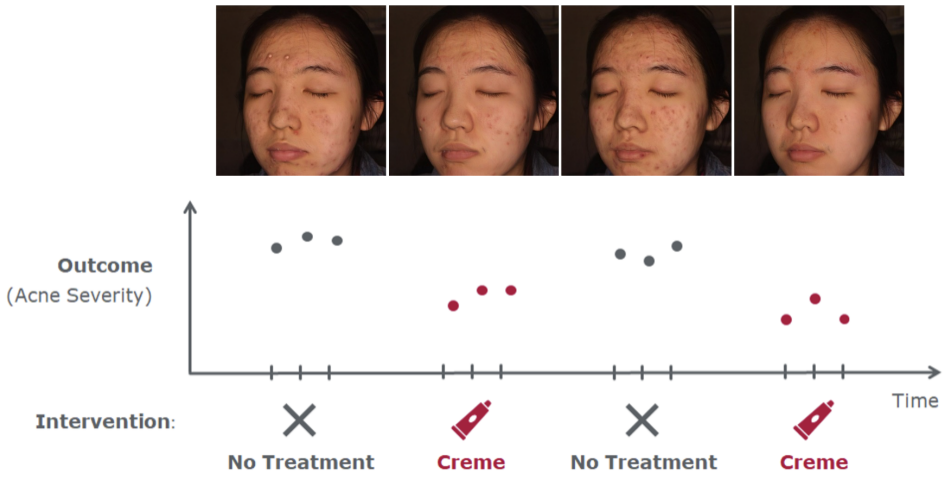

Figure 5: Using an input image with mild acne from the ACNE04 dataset, only few inference steps (strength=3) and a strong focus on the accompanying text prompt (guidance_scale=12) yielded results with high fidelity and accuracy.

7.1 Threats to Validity

Several potential threats to the validity of our evaluation include the representativeness of the training dataset, the model's generalizability to unseen data, and the subjective nature of acne severity assessment. The limited diversity in the initial dataset ACNE04 could bias the model, affecting the variety and realism of the generated images.

Figure 6: Prompting a fine-tuned SDXL model to generate an image with very severe acne using an out-of-distribution portrait as input did not lead to satisfactory results, even with a high guidance_scale (focusing more strongly on the prompt) and low strength (first image on the upper left).

8 Discussion

Despite the success of SDXL models in generating images that realistically depict various acne severity levels, a significant limitation emerged: the models' inability to precisely control the placement of acne patches on the face. This issue underscores a critical gap in the current capabilities of generative models, particularly in their capacity to maintain the identity of individuals while performing detailed, localized edits on specific facial features, such as acne patches, within portrait images. This challenge is not trivial; it directly impacts the utility of generated images for applications in personalized medicine and clinical research, where the accurate simulation of disease progression or treatment effects on the same individual over time is paramount.

The necessity for models that can seamlessly integrate identity preservation with the ability to make subtle, localized modifications to existing acne patches — without introducing or eliminating acne patches randomly — becomes evident. Such precision is crucial for creating a series of images that can faithfully represent the natural progression or regression of acne on a patient's face, providing valuable insights into the efficacy of treatments over time. Furthermore, the ability to accurately simulate temporal changes in acne severity on a consistent facial identity would significantly enhance the potential of generative models in dermatological studies, facilitating a deeper understanding of acne dynamics and contributing to the development of more effective treatment protocols.

Enhancing the training of stable diffusion models by extending training durations across all classes and incorporating broader, more diverse datasets could significantly improve the models' precision in replicating acne severity and lesion placement. These data could possibly be generated with the newly fine-tuned SDXL models if they're carefully evaluated. Additionally, integrating other identity-preserving adapters like IP-Adapters during fine-tuning could better ensure the maintenance of subjects' identity while accurately modifying acne features. Finally,

the generated dataset for 10 patients à 140 images each with varying acne severity levels should be used to evaluate the previously existing unsupervised learning models from Schneider et al [8].

9 Conclusion

The project presents a promising advancement in the generation of synthetic medical images for acne severity assessment, showcasing the potential of Diffusion Models and other generative techniques in overcoming the limitations of traditional data collection methods. The generated images by SDXL exhibit a high degree of realism and fidelity, marking a significant step forward in the simulation of multimodal N-of-1 trials. Future work will focus on further enhancing the model's identity-preserving accuracy while enabling temporal changes of specific acne patches.

References

[1] T. Brooks, A. Holynski, and A. A. Efros. "InstructPix2Pix: Learning to Follow Image Editing Instructions". In: *CVPR*. 2023.

[2] J. Fu, S. Liu, S. Du, S. Ruan, X. Guo, W. Pan, A. Sharma, and S. Konigorski. "Multimodal N-of-1 trials: A Novel Personalized Healthcare Design". In: *arXiv preprint arXiv:2302.07547* (2023).

[3] M. Hettich. *[Fork] SkinGAN: Generating Skin Images for Medical Research*. `https://github.com/ManuelHettich/SkinGAN`. Accessed: 2024-03-22. 2024.

[4] M. Hettich. *A filtered version of the ACNE04 Dataset for Acne Severity Classification*. `https://huggingface.co/datasets/ManuelHettich/acne04`. Accessed: 2024-03-22. 2024.

[5] M. Hettich. *Stable Diffusion Training and Inference Notebooks for Simulating Multimodal N-of-1 Trials*. `https://github.com/ManuelHettich/synthetic_acne_images`. Accessed: 2024-03-22. 2024.

[6] Z. Lu and S. Krishnamurthy. "SkinGAN: Medical Image Synthetic Data Generation using Generative Methods". In: *Journal Name* Volume Number.Issue Number (2021). Available online, Page Numbers.

[7] N. Ruiz, Y. Li, V. Jampani, Y. Pritch, M. Rubinstein, and K. Aberman. "DreamBooth: Fine Tuning Text-to-image Diffusion Models for Subject-Driven Generation". In: (2022).

[8] J. Schneider, T. Gärtner, and S. Konigorski. "Multimodal Outcomes in N-of-1 Trials: Combining Unsupervised Learning and Statistical Inference". In: *arXiv preprint arXiv:2309.06455* (2023).

[9] Q. Wang, X. Bai, H. Wang, Z. Qin, and A. Chen. "InstantID: Zero-shot Identity-Preserving Generation in Seconds". In: *arXiv preprint arXiv:2401.07519* (2024).

[10] X. Wu, N. Wen, J. Liang, Y.-K. Lai, D. She, M.-M. Cheng, and J. Yang. "Joint Acne Image Grading and Counting via Label Distribution Learning". In: *College of Computer Science, Nankai University; Beijing Tsinghua Changgung Hospital; School of Computer Science and Informatics, Cardiff University* (2023). Available at https://github.com/xpwu95/ldl.

[11] H. Ye, J. Zhang, S. Liu, X. Han, and W. Yang. "IP-Adapter: Text Compatible Image Prompt Adapter for Text-to-Image Diffusion Models". In: (2023).

[12] H. Zein, S. Chantaf, R. Fournier, and A. Nait-Ali. *Generative Adversarial Networks for Anonymous Acneic Face Dataset Generation*. arXiv:2211.04214v1 [cs.CV]. 2022. arXiv: 2211.04214 [cs.CV].