Enhancing Fitness Visualization Using Stable Diffusion-Based Inpainting Techniques

Sanjay Sampathkumar
Department of Artificial Intelligence and Robotics
Master's Student of Hochschule Hof
Hof, Germany

Sentoor Kumar Panjanathan Muruga Prakash
Department of Artificial Intelligence and Robotics
Master's Student of Hochschule Hof
Hof, Germany

Abstract This paper presents a Stable Diffusion-based inpainting system developed to generate realistic fitness transformations. The project focuses on modifying user-uploaded images to visualize a more fit version of the individual while maintaining anatomical realism. Various inpainting techniques were tested, and performance was assessed using image quality metrics such as CLIP, FID, DBCNN, BRISQUE, and NIQE. The study faced challenges related to inpainting accuracy, segmentation limitations, and visual inconsistencies in body transformations. Future improvements aim to enhance segmentation precision, refine inpainting techniques, and optimize model performance for better realism and user experience.

I. INTRODUCTION

A. BACKGROUND

With increasing interest in **fitness visualization**, AI-driven techniques have emerged to help individuals visualize potential body transformations. Traditional methods often lack **realism and personalization**, making them less effective in motivating users. Advances in **generative AI and diffusion models** offer solutions that can generate realistic fitness transformations.

B. Objectives

This project aims to:

1. Implement a Stable Diffusion-based inpainting approach for body transformation.
2. Use image segmentation to accurately select body areas for modification.
3. Evaluate visual quality and realism using standard image assessment metrics.
4. Identify limitations in AI-generated fitness visualizations and suggest improvements.

II. WORK DONE SO FAR

B. Implemented features

The system was designed using a zero-shot Stable Diffusion-based inpainting approach, eliminating the need for retraining. The key steps implemented include:

1. **Image Upload & Preprocessing**

- Implemented a Gradio-based user interface (UI) to allow users to upload images.
- Developed preprocessing steps to handle different image resolutions and formats.

2. **Automated Image Segmentation**

- Used text-to-mask segmentation (CLIPSeg model) to identify body areas for transformation.
- Integrated a Unprompted extension for automatic mask generation [1].

3. **Inpainting-Based Body Transformation**

- Applied Stable Diffusion inpainting using Realistic Vision v6.0 B1 [2].
- Implemented different transformation intensities: light, medium, and full body modifications.

III. METHODOLOGY

Despite implementing a functioning **AI-based fitness visualization system**, several challenges arose:

A. Segmentation Issues

1. **Mask generation errors:** The text-to-mask model sometimes failed to correctly identify

muscle regions, leading to incomplete transformations.

2. **Inaccurate region selection:** The segmentation model struggled with **different poses, lighting, and clothing styles**, causing inconsistent results.

B. Inpainting Limitations

1. **Distortions in body reshaping:** The diffusion model occasionally produced unnatural body proportions.
2. **Loss of anatomical consistency**: Some transformations introduced artifacts, particularly in areas with complex textures (e.g., arms and abdominal muscles).
3. **Over-smoothing of details:** In some cases, muscle definition was blurred or lost after transformation.

C. Biases in AI Outputs

1. **Gender disparities:** Female images often had lower perceptual realism, despite achieving better FID scores.
2. **Body type inconsistencies:** The model performed well on slim body types, but struggled with heavier body types, leading to unnatural results.

D. Performance & Computation Challenges

1. **Slow processing times:** Inpainting took longer than expected, affecting real-time usability.
2. **Hardware limitations**: Running diffusion models with high-quality settings required significant GPU resources.

E. Dataset overview

The model was trained using a dataset of 155 images located in the folder **"30_make a fit, strong version of me person"**. These images primarily depict **fitness-enhanced versions of individuals**, aligning with our project goal of generating **realistic fitness transformations**.

1. **Dataset Structure:**

- The dataset consists of **preprocessed fitness images** suitable for Stable Diffusion fine-tuning.

- Images were categorized based on **desired fitness attributes**, including **muscle definition, body symmetry, and posture improvements**.

- These images were utilized in **LoRA fine-tuning** to help the model learn more **precise inpainting techniques** for fitness visualization.

2. **Dataset Limitations:**

- The dataset lacked **sufficient diversity in body types**, which **may have contributed to bias** in transformation results.

- Some **images required additional captions or metadata**, which **could have improved training accuracy**.

- The limited dataset size meant that the model **was trained on fewer variations**, affecting generalization.

IV. Results and Discussion

A. Model comparison results

To evaluate the effectiveness of the FitMe fine-tuned model compared to the SD v1.5 model, we performed an in-depth analysis using several image quality assessment metrics. Below are the averaged results across all generated images:

Table 1 Average performance metrics by both models

Metric	FitMe Model	SD V1.5
CLIP Score	0.2505	0.2681
FID Score	100.27	98.23
DBCNN Score	0.5602	0.5646
BRISQUE Score	55.48	60.09
NIQE Score	15.78	15.68

B. Key Observations

1. **CLIP Score**: SD v1.5 performs slightly better, meaning it aligns more closely with the input prompts.
2. **FID Score**: SD v1.5 generates slightly more realistic images, though FitMe is close.
3. **DBCNN Score**: Both models show similar feature-based quality scores.
4. **BRISQUE Score**: FitMe produces **less perceptual distortion**, meaning the images appear smoother.

V. Visual Analysis of Model Outputs

To further examine the effectiveness of the **FitMe** and **SD v1.5** models, a **comparative visual analysis** was conducted. The original input images, along with their respective outputs from both models, were collaged into a set of three comparative grids:

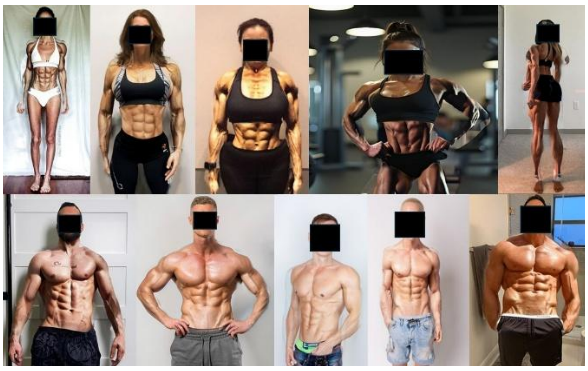

Figure 1 Input Image Collage – A collage of the 10 original input images used for inpainting and transformation.

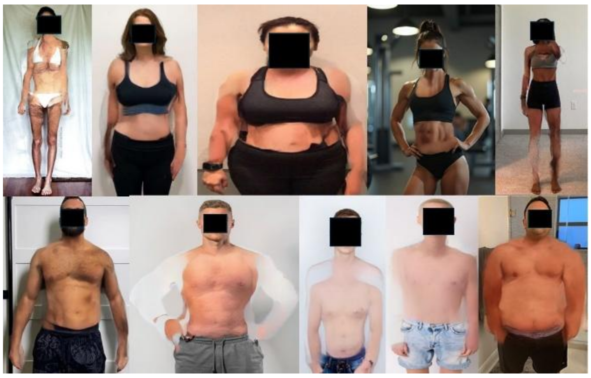

Figure 2 FitMe Model Output Collage – A collage of the 10 images generated by the FitMe fine-tuned model.

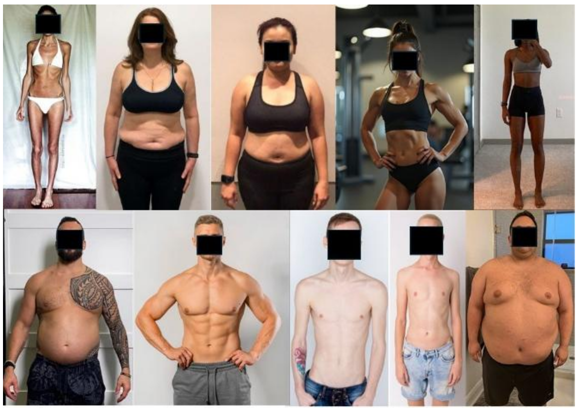

Figure 3 SD v1.5 Model Output Collage – A collage of the 10 images generated by the SD v1.5 model.

This structured approach allows for a **direct comparison** of how each model performs across **varied subjects, body types, and fitness transformations**.

VI. Key Observations from Visual Analysis

A. Anatomy & structural consistency

1. SD v1.5: The model maintains better anatomical consistency overall. Muscular structures, body proportions, and overall limb placements are more aligned with realistic human anatomy.
2. FitMe: The model, in some cases, introduces distortions, particularly in muscle definition and proportions. Certain outputs display unnatural abdominal placements, uneven shoulders, or excessive smoothing of body parts.

B. Adherence to Prompts

1. SD v1.5: This model adheres more closely to the input prompts, particularly in lighting conditions, muscular structure, and gender-specific characteristics. However, it sometimes over-enhances muscle details, making the results appear slightly exaggerated.
2. FitMe: FitMe struggled with strict adherence to the input prompts in some images. Some outputs show inconsistent lighting effects, altered skin tones, and occasional mismatches with the gender specification in prompts.

C. Texture, Detail, and Image Quality

1. SD v1.5: The textures and details, particularly for skin, clothing, and lighting, appear sharper and well-defined. The model maintains better contrast, making features like abs, arms, and legs more distinct.
2. FitMe: The model softens textures, making some areas appear blurry or over-smoothed. In some cases, this results in loss of finer details, particularly in areas with clothing folds, muscle contours, or lighting-based shadows.

D. Clothing fit & realism

1. SD v1.5: The clothing appears well-fitted and follows the contours of the body realistically. However, in some images, artifacts appear around the waist or shoulder areas, likely due to inconsistencies in inpainting.
2. FitMe: The model sometimes generates ill-fitting clothing, where outfits appear stretched, distorted, or out of alignment with the body.

E. Lighting & skin tone consistency

1. SD v1.5: The model maintains consistent lighting and skin tones. Shadows and highlights align logically with the input images.
2. FitMe: Some outputs exhibit inconsistent lighting, with overexposed or underexposed areas, particularly on faces and abs. Additionally, skin tones sometimes appear uneven or too synthetic.

VII. Final Observations from the Visual Comparison

After closely analyzing the collaged images, several conclusions emerge:

1. FitMe is better at preserving natural facial and skin tone consistency but tends to under-enhance muscle transformations.
2. SD v1.5 is better at defining fitness attributes but sometimes introduces artificial-looking muscles and exaggerated lighting effects.
3. Both models struggle with anatomical proportions, requiring further refinement in body part segmentation and fine-tuning techniques.
4. Neither model perfectly follows the given fitness prompts, highlighting a need for more training data and better prompt engineering techniques.

A. Potential improvements based on visual analysis

1. Better Prompt Engineering: Refining prompt structures to guide specific muscle enhancements and avoid distortions.
2. Longer Training Cycles: Increasing LoRA fine-tuning beyond 10 epochs to improve muscle refinement and body shape consistency.
3. More Diverse Training Data: Including varied fitness levels, ethnicities, and poses to improve generalization across body types.
4. Enhanced Segmentation & Masking: Implementing better inpainting region selection to improve body modification precision.

VIII. Additional Challenges Faced

During the development and training of the FitMe model, we encountered several unexpected challenges that affected the final results and performance.

A. **Update Wiped Model**: A major issue was a system update that erased our previously optimized model, forcing us to switch to a different training method.

B. **Switch to Kohyaa Trainer**: Due to compatibility issues with our previous training setup, we migrated our fine-tuning process to **Kohyaa**, requiring adjustments to our training strategy.

C. **Limited Training Resources**: Our model was trained for **only 10 epochs**, which significantly limited its ability to generalize to diverse body types and poses.

D. **Dataset Limitations**: The dataset we used required **more captions and diversity** to improve realism and fairness in body transformations. More images, higher-quality annotations, and better captioning would have improved overall results.

IX. Fine-Tuning Process & Lessons Learned

A. Kohyaa fine-tuning approach

After switching to **Kohyaa** for fine-tuning, we adjusted our training process to optimize the performance of the inpainting model. The fine-tuning involved:

1. **Training LoRA weights** to enhance image consistency and transformation accuracy.

2. **Optimizing embeddings** to better learn fitness-related transformations.

3. **Adjusting prompts** to improve control over output generation.

B. Fine-tuning hyperparameters used

1. Learning rate = 1e-4
2. Training epochs = 10
3. Batch size = 2
4. Optimizer = AdamW

C. LESSONS LEARNED

1. More Training Steps Needed: Future models should be trained for at least 20–30 epochs to achieve better generalization.
2. Larger Dataset Required: A more diverse dataset with better captions and variations in fitness levels would significantly enhance accuracy.
3. Improved Inpainting Strategy: Some images showed artifacts in body parts, which could be improved using better segmentation and inpainting masks.

4. Refined Model Architecture: Experimenting with different LoRA architectures could improve transformation accuracy.

X. Potential Real-World Applications

1. Fitness Progress Tracking: Users can upload images to visualize potential fitness transformations and set realistic fitness goals.
2. Health and Rehabilitation: Could be used for AI-assisted physiotherapy to model expected body recovery post-injury.
3. AI-Powered Virtual Training Programs: This model could help fitness coaches and nutritionists personalize transformation plans.
4. Virtual Try-Ons: Used in fashion and fitness clothing industries to predict how different body types will look in specific outfits.

XI. Conclusion and future work

This study demonstrated the potential of Stable Diffusion and inpainting techniques for fitness visualization but also highlighted significant challenges.

A. Key takeaways

1. AI-driven fitness transformations are feasible but require precise segmentation and inpainting techniques.

2. Gender and body type biases exist and must be addressed to improve fairness.

3. Performance optimization is necessary to make the system more user-friendly.

B. Future improvements

1. Enhance segmentation accuracy to refine inpainting region selection.

2. Improve inpainting algorithms to handle diverse body transformations.

3. Introduce a hybrid AI-human correction system where users can refine masks before inpainting.

4. Optimize computational efficiency to make real-time transformations feasible.

5. Gather user feedback to improve transformation realism based on subjective perception.

By addressing these challenges, the project aims to develop a more robust and realistic fitness visualization system.

XII. References

[1] T. Lüdecke et al., “Image segmentation using text and image prompts,” *arXiv preprint*, 2022.
[2] SG161222, “Realistic Vision v6.0,” *CivitAI*, 2024.
[3] R. Rombach et al., “High-resolution image synthesis with latent diffusion models,” *arXiv preprint*, 2022.
[4] Y. Wang et al., “Towards context-stable image inpainting,” *arXiv preprint*, 2023.