Synthetic MRI Image Generation Using a Combination of VQ-GAN and Flow Matching

Nora Schaba

Supervisors: Prof. Dr. Elke Hergenröther, Prof. Dr. Andreas Weinmann, Institut supervisor : Dr.-Ing. Cristina Oyarzun Laura

Hochschule Darmstadt, University of Applied Sciences, Fachbereich Mathematik und Naturwissenschaften Informatik

Motivation

Deep learning has revolutionized medical imaging but remains constrained by the challenges of data scarcity, variability in pathologies, and ethical concerns around patient privacy. These limitations hinder the development of robust and generalizable models [3].

Flow Matching (FM) models, while successful in non-medical applications, remain underexplored for medical imaging, especially for three-dimensional (3D) data. This work investigates their potential for generating synthetic 3D MRI images, specifically focusing on brain scans associated with Alzheimer's disease. A novel framework combining FM with a Variational Autoencoder-based Generative Adversarial Network (VQ-GAN) is proposed to achieve efficient and high-quality generation

Research Questions

This work is guided by the following research questions:

1. How effective is the flow matching framework for generating realistic and high-quality 3D medical images, specifically brain MRI scans?
2. What are the benefits and challenges of combining VQ-GAN and flow matching models for latent-space representation and synthetic image generation?
3. Can a flow matching model operating in the latent space improve computational efficiency without compromising on generation quality?
4. What are the challenges and limitations of applying flow matching models to medical imaging, and how can they be addressed?

Methodology

The proposed framework is a two-stage model designed for efficient and high-quality 3D medical image generation:

The first stage employs a VQ-GAN [1] to compress high-dimensional MRI data into a compact latent space while preserving structural integrity.

- **Encoder:** Converts 3D MRI images into a low-dimensional latent representation, retaining essential anatomical details.
- **Decoder:** Reconstructs the original MRI images from the latent space representation.
- **Codebook:** A set of discrete embeddings ensures efficient vector quantization between the encoder and decoder.
- **Discriminator:** Enhances realism by distinguishing real images from reconstructed ones during training.

This stage reduces data complexity, optimizing it for use in the second stage while maintaining high reconstruction quality.

The second stage employs a Flow Matching (FM) [2] model to generate synthetic images by operating within the latent space produced by the VQ-GAN.

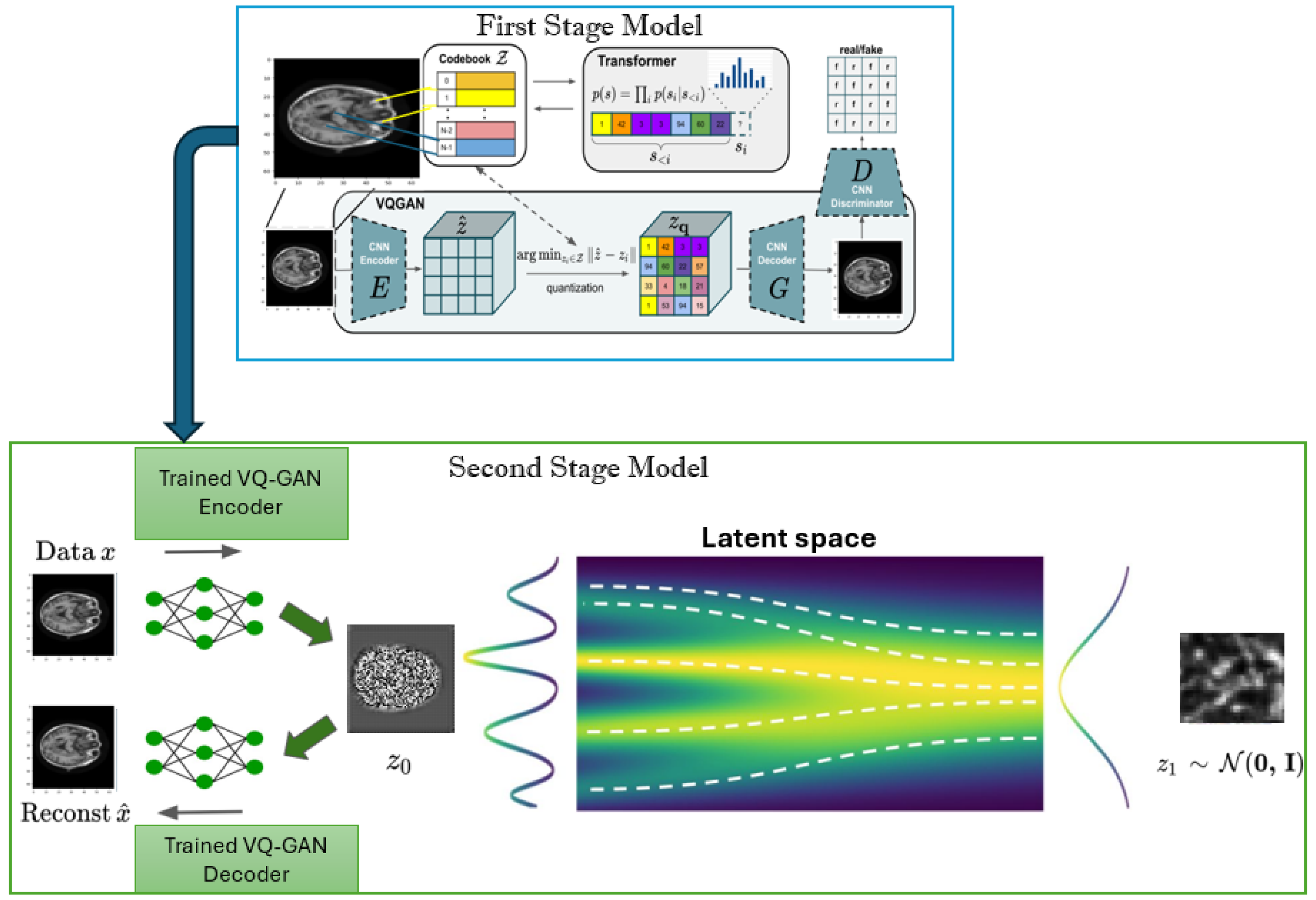

Figure 1. Proposed two-stage framework for 3D synthetic MRI image generation.

Result

The experiments demonstrate the potential of Flow Matching (FM) combined with VQ-GAN for generating synthetic 3D MRI images. The evaluation spans both quantitative and qualitative metrics to assess model performance.

The FM model significantly outperformed traditional denoising diffusion models in terms of training time, reducing it from 10 days to 13 hours, while maintaining competitive image quality. Key metrics include:

- **Peak Signal-to-Noise Ratio (PSNR):** 21 dB
- **Structural Similarity Index (SSIM):** 0.73
- **Multi-Scale Structural Similarity Index (MS-SSIM):** 0.65

These results indicate the FM model's ability to capture macro-level anatomical features but highlight challenges in preserving finer structural details.

The generated images effectively capture major anatomical structures, including general brain regions. However, some limitations were noted:

- **Blurriness:** High-frequency details are not fully captured, leading to slightly blurred outputs.

A highlighted block containing some math

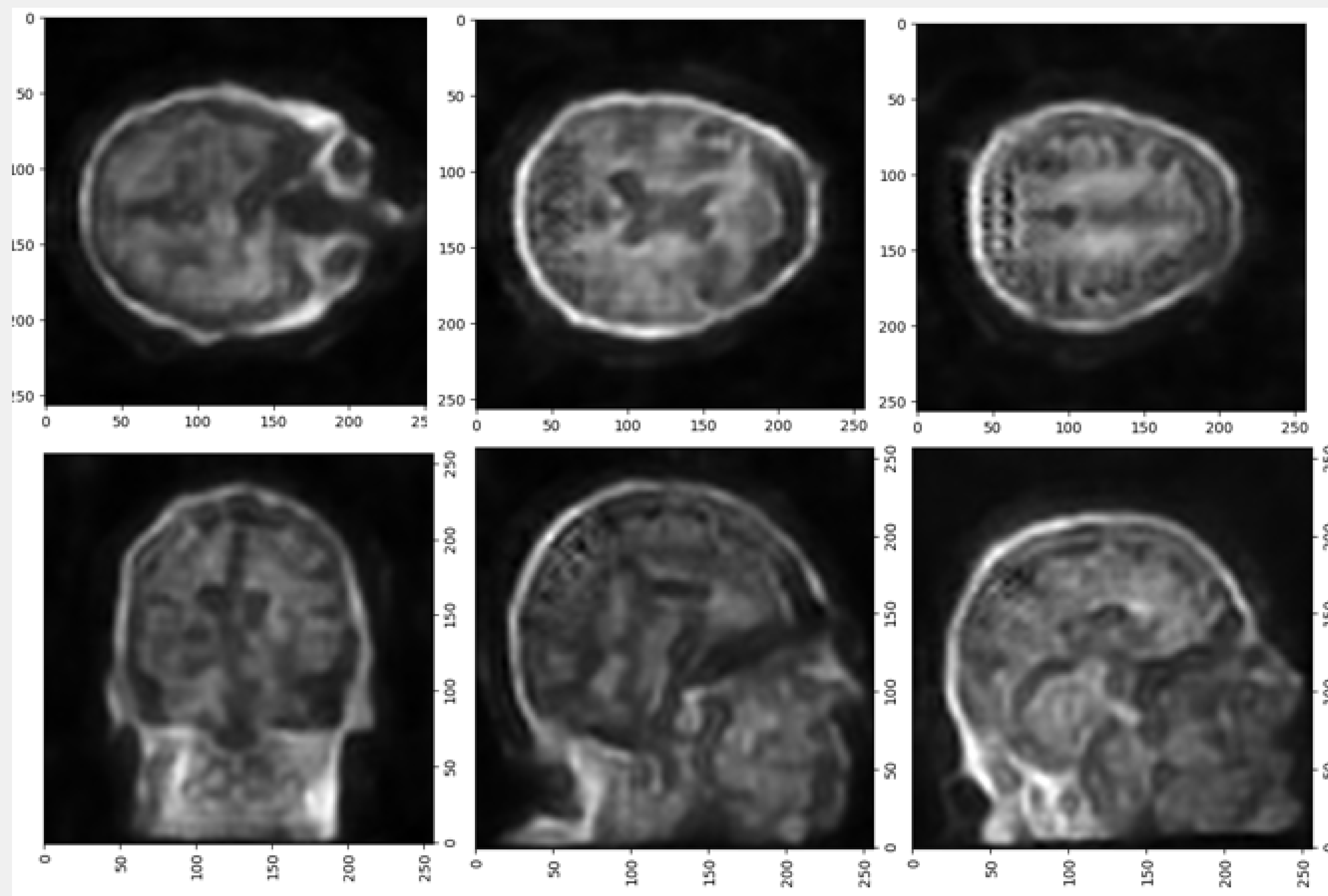

Figure 2. Results of the flow matching model. Improved preprocessing and reconstruction have led to clearer brain structures and better overall image quality.

Conclusion

This research investigated Flow Matching (FM) models combined with VQ-GAN for generating realistic 3D brain MRI images. The FM model demonstrated its effectiveness by reducing training time from 10 days (diffusion models) to 13 hours while maintaining strong macro-level anatomical fidelity. The combination of VQ-GAN and FM allowed efficient latent-space transformations but faced challenges such as artifacts, loss of fine details, and dataset bias.

While FM models improved computational efficiency and generation quality, key limitations like resolution constraints and preprocessing decisions highlight areas for further refinement. This work establishes a foundation for advancing FM models in medical imaging, with future efforts focusing on high-resolution datasets, diverse imaging modalities, and improved architectures to enhance quality and reliability.

References

[1] Patrick Esser, Robin Rombach, and Björn Ommer.
Taming transformers for high-resolution image synthesis, 2021.

nora.schaba@stud.h-da.de