Proceedings Paper:
Deo, Y., Lin, F., Dou, H. et al. (4 more authors) (2024) Few-shot learning in diffusion models for generating cerebral aneurysm geometries. In: 2024 IEEE International Symposium on Biomedical Imaging (ISBI). 21st IEEE International Symposium on Biomedical Imaging (ISBI 2024), 27-30 May 2024, Athens, Greece. IEEE ISBN: 979-8-3503-1334-5 ISSN: 1945-7928 EISSN: 1945-8452

https://doi.org/10.1109/ISBI56570.2024.10635313

FEW-SHOT LEARNING IN DIFFUSION MODELS FOR GENERATING CEREBRAL ANEURYSM GEOMETRIES

*Yash Deo[1] ✉, Fengming Lin[1], Haoran Dou[1], Nina Cheng[1], Nishant Ravikumar[1], Alejandro F Frangi[1 2 3 4 5], Toni Lassila[1 *]*

[1] CISTIB,University of Leeds, Leeds, UK
[2] NIHR Leeds Biomedical Research Centre, Leeds, UK
[3] Alan Turing Institute, London, UK
[4] Medical Imaging Research Center (MIRC), Electrical Engineering and Cardiovascular Sciences Departments, KU Leuven, Leuven,
[5] Division of Informatics, Imaging and Data Science, Schools of Computer Science and Health Sciences, University of Manchester, Manchester, UK

ABSTRACT

The study of brain vessel pathologies is critical for the advancement of neurovascular medicine, yet researchers often face significant hurdles due to the scarcity of imaging data for certain uncommon types of aneurysms. Generative deep learning models have been proposed to address the lack of high-quality labeled medical images - however, the shortage of data also presents a unique challenge in training generative models. To address this issue, our work explores the efficacy of training latent diffusion models (LDMs) with few-shot learning, enabling the generation of detailed vessel segmentations from as few as five images per class. By incorporating set-based vision transformers for class embeddings and leveraging signed distance functions (SDFs) as a novel form of conditioning, our method reduces the need for extensive datasets for training. Comparative studies with established generative models, including variational autoencoders (VAEs) and generative adversarial networks (GANs), highlight the robustness of our approach. Our model not only successfully generates high-quality segmentations of brain vessels with aneurysms but also significantly outperforms the standard generative models.

***Index Terms*—** Diffusion Models, Image Synthesis, Brain Vessel Synthesis, Transformers

1. INTRODUCTION

Cerebral aneurysms pose significant neurosurgical and neurological concerns and have the potential to lead to life-threatening conditions, like subarachnoid hemorrhage (SAH). Their prevalence in the general population underscores their contribution to morbidity and mortality. A major challenge in this domain arises from the scarcity of comprehensive data, particularly for less common aneurysm phenotypes. This data limitation poses significant obstacles in developing accurate and robust diagnostic models. Generative models, particularly in medical imaging, present a promising solution to this issue. They hold the potential to synthesize high-quality, detailed images of cerebral aneurysms, even in scarcity. However, the effectiveness of traditional generative models is typically constrained by the availability of data, a notable hurdle in the context of aneurysm imaging, where extensive datasets are often lacking.

In the domain of medical imaging, generative models such as generative adversarial networks (GANs) [1] have emerged as a promising solution, offering the potential to create detailed and accurate representations of anatomical structures [2] . More recently, diffusion models (DDPM) have shown great prowess in generating synthetic data [3] and outperforming GANs in image synthesis [4]. Diffusion models have also been successfully used to generate synthetic brain magnetic resonance images (MRIs) [5, 6, 7] and vascular structures [8]. However, the efficacy of these models is often limited by the requirement for extensive training datasets, which are not always available. Even when labeled datasets are available, medical imaging datasets suffer from data imbalance due to certain anatomical phenotypes being underrepresented.

The concept of few-shot learning, a technique for training models with limited data, has become increasingly relevant in medical imaging domains characterized by data scarcity and class imbalance. This is especially true for rare or underrepresented cerebral aneurysm types. While few-shot learning has been explored in diffusion models in prior research [9, 10], our work is the first to our knowledge to apply this concept to the generation of cerebral aneurysms in brain vessel imaging.

* t.lassila@leeds.ac.uk

Our study addresses this gap by introducing an innovative approach using latent diffusion models (LDMs) with few-shot learning, allowing for the generation of high-fidelity models of brain vessels with aneurysms from a very limited number of samples in each class. Through the integration of transformer-based class embeddings, we reduce the reliance on having a large number of samples from each class to conditionally generate images. We also leverage signed distance functions (SDF) as a conditioning variable to further enhance the quality of the generated vessels and maintain vessel continuity. We compare the performance of our model against other generative models such as a 3D GAN, 3D variational auto-encoders (VAEs), and also against vanilla diffusion models. To assess the quality of the generated aneurysms, we use metrics such as multi-scale structural similarity (MS-SSIM), Fréchet inception distance (FID), and 4GR SSIM. To our knowledge, this is the first study to use generative/diffusion models to generate synthetic brain vessels with aneurysms.

2. METHODS

2.1. Data and Preprocessing

For training our model, we utilized the @neurIST dataset encompassing 225 3D Rotational Angiography (3DRA) scans of the brain, each with at least one cerebral aneurysm. Out of these, detailed information regarding aneurysm location and other conditional attributes was available for 105 cases. Within these 105 labeled cases, there were more than 15 different classes of aneurysms based on their location with each class having about 7 sample cases on average. In the initial phase of preprocessing, we extracted vessel segmentations from the 3DRA volumes. This extraction was facilitated by the application of VASeg, a segmentation tool designed for vascular imaging [11]. Post-segmentation, the 3DRA volumes underwent a process of centerline cropping, ensuring a focus on the most relevant vascular structures. These cropped segments were then resized to uniform dimensions of $128 \times 128 \times 100$, optimizing them for subsequent processing and analysis. The final step involved the categorization of aneurysms based on their location attributes and saving them as class variables to act as a conditioning vector to the diffusion model. In this study, we mainly focus on basilar tip, medial wall carotid, and ophthalmic segment carotid aneurysms. Each class contains around 5 samples.

2.2. Latent Diffusion Model

Diffusion models have demonstrated remarkable success in synthesizing high-quality medical images and vascular structures. Central to the operation of diffusion models is the concept of a Markov chain, which is employed to methodically introduce Gaussian noise into the observed data through a sequence of diffusion steps. The crux of these models lies in their ability to reverse this diffusion process, thereby enabling the generation of new samples from the noise-infused data.

Despite their effectiveness, a notable challenge with conventional diffusion models arises when dealing with high-dimensional data such as the images of size $128 \times 128 \times 100$ used in our study. To circumvent this computational complexity, we have opted to utilize a latent diffusion model (LDM). The architecture of LDM comprises two pivotal components: a pre-trained autoencoder and a diffusion model. The autoencoder is tasked with learning a lower-dimensional latent representation of the brain vasculature from $128 \times 128 \times 100$ to $128 \times 128 \times 1$. This reduction in dimensionality is crucial as it allows for a more manageable and efficient manipulation of data. Concurrently, the diffusion model is designed to focus on modeling the high-level semantic representations within this latent space. By operating in a space of reduced dimension, the LDM alleviates the computational burden but retains the capacity to capture and model the intricate details and nuances of the brain vascular structures.

Like in [4], the diffusion process can be defined through forward and reverse Markov chains, where the forward process iteratively transforms the data x_0 into a standard Gaussian X_T as follows:

$$\begin{aligned} q\left(\mathbf{x}_{1:T}|\mathbf{x}_0\right) &= \prod_{t=1}^{T} q\left(\mathbf{x}_t|\mathbf{x}_{t-1}\right), q\left(\mathbf{x}_t|\mathbf{x}_{t-1}\right) \\ &:= \mathcal{N}\left(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I}\right) \end{aligned}$$

where $q\left(\mathbf{x}_t|\mathbf{x}_{t-1}\right)$ is the transition probability at the time step t based on the noise schedule β_t. Therefore, the noisy data $\mathbf{x}_t$ can be formulated as $q\left(\mathbf{x}_t|\mathbf{x}_0\right) = \mathcal{N}\left(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I}\right)$, where $\alpha_t := 1-\beta_t, \bar{\alpha}_t := \prod_{s=1}^{t} \alpha_s$.

Consecutively, the reverse process parameterised by θ can then be defined as:

$$\begin{aligned} p_\theta\left(\mathbf{x}_0|\mathbf{x}_T\right) &= p\left(\mathbf{x}_T\right) \prod_{t=1}^{T} p_\theta\left(\mathbf{x}_{t-1}|\mathbf{x}_t\right), p_\theta\left(\mathbf{x}_{t-1}|\mathbf{x}_t\right) \\ &:= \mathcal{N}\left(\mathbf{x}_{t-1}; \mu_\theta\left(\mathbf{x}_t, t\right), \mathbf{\Sigma}_\theta\left(\mathbf{x}_t, t\right)\right) \end{aligned}$$

The simplified evidence lower bound (ELBO) [4] loss can be formulated as a score-matching task, where the neural network predicts the actual noise ϵ added to the observed data:

$$\mathcal{L}_\theta := \mathbb{E}_{\mathbf{x}_0, t, C, \epsilon \sim \mathcal{N}(0,1)}\left[\left\|\epsilon - \epsilon_\theta\left(x_t, t, C\right)\right\|^2\right]$$

where C is the conditioning vector in conditional generation. In our study, the conditioning vector encodes the location of the aneurysm.

The 3D binary masks of the vessels generated from 3DRA volumes in our dataset are passed through the encoder of the pre-trained autoencoder to obtain a dimensionally reduced latent space. This latent space serves as the input for our diffusion model. Consequently, the diffusion model's output is

also in this latent space, which is then processed through the decoder of the pre-trained autoencoder to reconstruct the 3D binary masks.

We first train our latent diffusion model unconditionally with no additional condition features on the unlabeled samples in the dataset so that it can learn to generalize the structure of the vessels. After pre training on the unlabeled data, we train the model over limited labeled cases from the three selected classes (basilar tip, medial wall carotid and ophthalmic segment carotid aneurysms) along with class-wise conditioning from a transformer and signed distance fields (SDF) based features.

2.3. Transformer based class conditioning

An inherent fault with generative models (especially diffusion) is their intrinsic reliance on substantial data volumes to train effectively and produce convincing outputs. This issue is particularly pronounced in our study, given the limited availability of data, with some classes containing as few as five samples. Such a sparse dataset poses significant difficulties for generative models, as they struggle to accurately approximate the distribution of the data.

To address this challenge, we introduce an innovative approach that integrates transformer [12] -based class features to guide the diffusion process. We employed a set based vision transformers (ViT) [13] model, designed to ingest the entire 3D volume and function as a classifier, determining the specific location of an aneurysm within the brain. Following the successful training of the ViT, we removed its final classification layer. Subsequently, we processed the images from each class through this transformer to extract class-wise encoded features. These features, in conjunction with the class conditioning variables, were then incorporated into the conditioning vector of the diffusion model, enhancing its ability to generate data representative of each class.

2.4. Signed Distance Field (SDF) based Conditioning

Although diffusion models show great success in generating medical images, generating vascular structures is challenging as vessels have structural features that need to be maintained, most importantly vessel continuity. Also, aneurysms are small compared to the total size of cerebral vasculature, which makes them hard to track and generate. Studies have shown that adding shape based features to the generative process can improve performance in these tasks [14, 15, 8].

To this end, we incorporate signed distance fields (SDF) as an additional input to the diffusion process. The primary idea behind SDFs is to associate each point in space with a distance value, and the sign of this distance value indicates whether the point is inside or outside of the shape which makes them particularly useful for tasks like shape analysis and 3D rendering. We first convert the segmentation masks for each class into corresponding SDFs. These SDFs act as an input to a 3D ResNet, which similar to the set-based ViT described in the previous subsection is trained to act as a classifier. After successful training, the final output layer is removed and the class-wise features are extracted and incorporated as conditions in the diffusion process. The introduction of these features enhances the quality of the generated vessels by promoting the generation of more continuous vessels as can be seen in Panel B in Fig. 2. The overall architecture of the model is shown below in Fig. 1

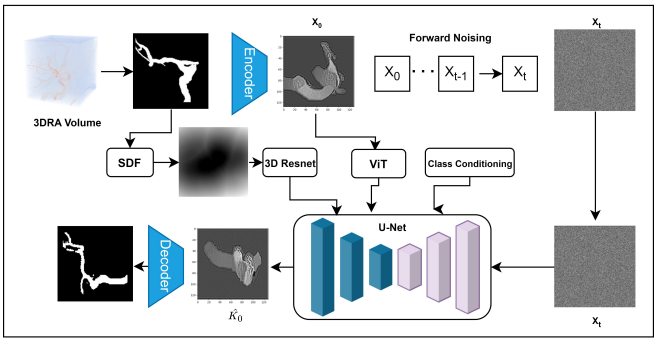

Fig. 1. Overview of the architecture of the model.

3. EXPERIMENTS AND RESULTS

3.1. Implementation Details

All models were implemented in TensorFlow 2.8 and Python 3. For the forward diffusion process we use a linear noise schedule with 1000 time steps. The model was trained for 2000 epochs with a learning rate of 0.0005 on a Nvidia Tesla T4 GPU and 38 Gb of RAM with Adam optimiser. The vision transformer was trained on a Nvidia V100 GPU with 38Gb of RAM.

3.2. Results and Discussion

The performance of our proposed model was compared against established generative models serving as baselines. These include a 3D convolutional variational autoencoder (3D C-VAE)[16], a 3D-α-Wasserstein generative adversarial network (3D-α-WGAN)[17], and a conventional diffusion model (Vanilla DDPM). The purpose of this comparison was to ascertain the efficacy of our approach relative to these well-established models in generating high-quality cerebral vascular images. To quantitatively assess the realism of the generated vasculature by each model, we employed the Fréchet inception distance (FID) score. The FID-score was computed using a pre-trained InceptionV3 network as a feature extractor. It is important to note that a lower FID score is indicative of higher perceptual image quality, reflecting

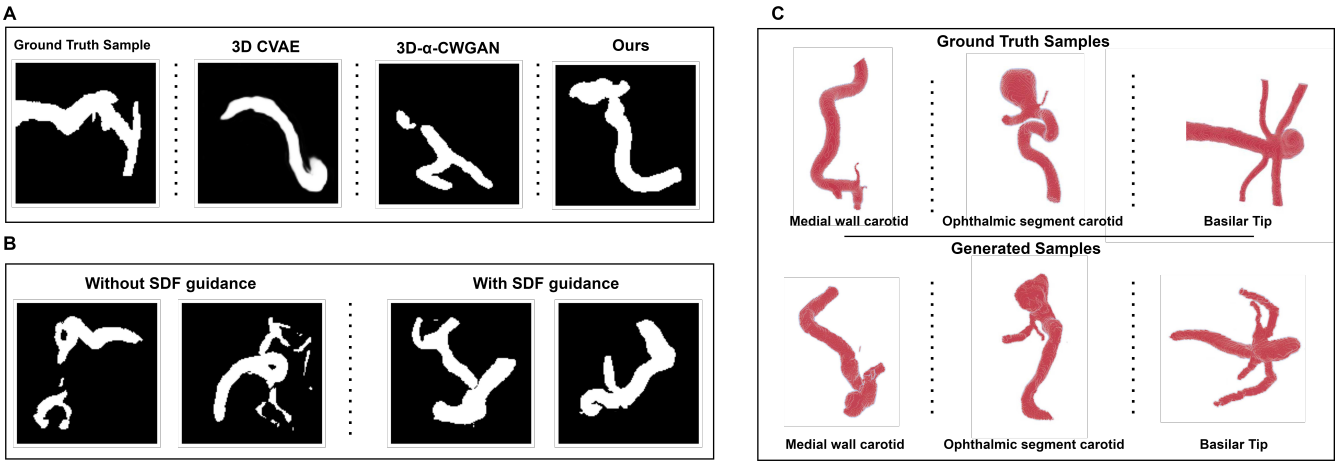

Fig. 2. Panel A compares the MIPs from the generated cases from different models. Panel B showcases the effect of adding SDF based conditioning to the diffusion process. Panel C compares the volumetric meshes generated from generated and ground truth cases from each class

greater realism in the generated images. Additionally, to provide a comprehensive evaluation of image quality, we utilized the multi-scale structural similarity index (MS-SSIM) and 4-G-R SSIM metrics, as outlined in references [18, 19]. These metrics are extensively used in the field to assess the quality of synthesized images. A higher score in both MS-SSIM and 4-G-R SSIM typically signifies superior image quality, implying a closer resemblance to the actual ground truth images. An extremely high score from MS-SSIM and 4-G-R SSIM however could indicate very high levels of similarity between the synthesised cases and the ground truth indicating low variablilty. The MS-SSIM and 4-G-R SSIM scores were calculated over six synthesized cases for each model.

Table 1 encapsulates the evaluation scores achieved by our model, 3D C-VAE, 3D-α-WGAN, and Vanilla DDPM, based on the aforementioned metrics. This comparative analysis enables us to elucidate the strengths and limitations of our approach in the context of existing generative models.

Table 1. Quantitative evaluation of Synthetic vessels

Model	FID ↓	MS-SSIM ↑	4-G-R SSIM ↑
3D CVAE	8.78	0.36	0.31
3D-α-WGAN	3.55	0.67	0.56
DDPM	4.41	0.69	0.55
Ours	2.56	0.71	0.61

Table 1 showcases that our model outperforms the other baselines in terms of FID, indicating that the distribution of the synthesized variants by our model more closely aligns with the real data distribution compared to other evaluated models. Furthermore, our approach outperforms the others in terms of MS-SSIM and 4-G-R SSIM scores, reflecting higher image quality and a closer resemblance of the generated vessels to the real ones.

Figure 2 provides a qualitative evaluation through a visual comparison of the synthesized samples from each model. Panels A and B employ maximum intensity projection (MIP) to render 3D binary masks of the vessels onto a 2D plane for analysis. In Panel A, we present the comparisons based on the MIP of the cases generated by each respective model. The convolutional variational autoencoder (VAE) primarily reproduces the fundamental structure of the vessels, achieving continuous vessel formation but lacking in variability and branching features. The generative adversarial network (GAN) introduces greater variability and detail in the vessel structures; however, it encounters challenges in maintaining vessel continuity. In contrast, our model excels in generating realistic and continuous vascular structures, closely mirroring the intricacies of actual vessels. Panel B delineates the differential impact of employing SDF-based conditioning in our diffusion model, underscoring its essential role in preserving vessel continuity, a feature that is notably compromised in its absence.

Recognizing the limitations of MIPs in accurately representing the complex three-dimensional nature of vascular structures, we further conducted a comparison using volumetric meshes which are showcased in panel C in Figure 2. These meshes were generated from binary masks for each class and compared against their corresponding ground truth samples. This analysis revealed that the cases synthesized by our model not only bear key characteristics akin to the ground truth but also exhibit discernible variability, demonstrating the model's efficacy in replicating both the fidelity and diversity of real-world vascular formations.

While the quality of the generated vessels from our study seems promising, it is important to acknowledge the limitations posed by the lack of extensive training data. This

scarcity potentially restricts the variability of the generated cases, as the model's capacity to learn diverse vessel structures is directly tied to the dataset's breadth. Additionally, it is crucial to consider anatomical accuracy in the context of variability. Excessive variability in the generated structures might not accurately reflect the true anatomical complexity of cerebral vessels. Therefore, while our model demonstrates proficiency in replicating realistic vessel structures, the balance between variability and anatomical fidelity remains a key consideration for the authenticity and applicability of the generated outputs.

4. CONCLUSION

This study introduced a novel approach for generating brain vessel segmentations with aneurysms, particularly under the constraint of having classes with limited data. By employing latent diffusion models enhanced with transformer-based class embeddings and signed distance functions, our model demonstrated superior performance over traditional generative models like 3D C-VAE and 3D-α-WGAN in terms of image quality and realism.

5. ACKNOWLEDGEMENT

This research was partially supported by the National Institute for Health and Care Research (NIHR) Leeds Biomedical Research Centre (BRC), UKRI Frontier Research Guarantee INSILICO (EP/Y030494/1) ,EC Sixth Framework Programme @neurIST (FP6-2004-IST-4-027703) and the Royal Academy of Engineering Chair in Emerging Technologies (CiET1919/19).

6. REFERENCES

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Commun. ACM*, vol. 63(11), pp. 139–144, 2020.

[2] Yash Deo, Rodrigo Bonazzola, Haoran Dou, Yan Xia, Tianyou Wei, Nishant Ravikumar, Alejandro F Frangi, and Toni Lassila, "Learned local attention maps for synthesising vessel segmentations from t2 mri," in *International Workshop on Simulation and Synthesis in Medical Imaging*. Springer, 2023, pp. 32–41.

[3] F. Khader, G. Mueller-Franzes, S.T. Arasteh, T. Han, C. Haarburger, M. Schulze-Hagen, P. Schad, S. Engelhardt, B. Baessler, S. Foersch, and J. Stegmaier, "Medical diffusion–denoising diffusion probabilistic models for 3D medical image generation," *arXiv preprint arXiv:2211.03364*, 2022.

[4] G. Müller-Franzes, J.M. Niehues, F. Khader, S.T. Arasteh, C. Haarburger, C. Kuhl, T. Wang, T. Han, S. Nebelung, J.N. Kather, and D. Truhn, "Diffusion probabilistic models beat GANs on medical images," *arXiv preprint arXiv:2212.07501*, 2022.

[5] L. Jiang, Y. Mao, X. Chen, X. Wang, and C. Li, "CoLa-Diff: Conditional latent diffusion model for multi-modal MRI synthesis," *arXiv preprint arXiv:2303.14081*, 2022.

[6] W. Peng, E. Adeli, Q. Zhao, and K.M. Pohl, "Generating realistic 3D brain MRIs using a conditional diffusion probabilistic model," *arXiv preprint arXiv:2212.08034*, 2022.

[7] W.H. Pinaya, P.D. Tudosiu, J. Dafflon, P.F. Da Costa, V. Fernandez, P. Nachev, S. Ourselin, and M.J. Cardoso, "Brain imaging generation with latent diffusion models," *Deep Generative Models: Second MICCAI Workshop, DGM4MICCAI 2022, Held in Conjunction with MICCAI 2022*, pp. 117–126, 2022.

[8] Yash Deo, Haoran Dou, Nishant Ravikumar, Alejandro F Frangi, and Toni Lassila, "Shape-guided conditional latent diffusion models for synthesising brain vasculature," *arXiv preprint arXiv:2308.06781*, 2023.

[9] Abhishek Sinha, Jiaming Song, Chenlin Meng, and Stefano Ermon, "D2c: Diffusion-decoding models for few-shot conditional generation," *Advances in Neural Information Processing Systems*, vol. 34, pp. 12533–12548, 2021.

[10] Giorgio Giannone, Didrik Nielsen, and Ole Winther, "Few-shot diffusion models," *arXiv preprint arXiv:2205.15463*, 2022.

[11] Fengming Lin, Yan Xia, Shuang Song, Nishant Ravikumar, and Alejandro F Frangi, "High-throughput 3dra segmentation of brain vasculature and aneurysms using deep learning," *Computer Methods and Programs in Biomedicine*, vol. 230, pp. 107355, 2023.

[12] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[13] Seung Hoon Lee, Seunghyun Lee, and Byung Cheol Song, "Vision transformer for small-size datasets," *arXiv preprint arXiv:2112.13492*, 2021.

[14] B.A. Brooksby, H. Dehghani, B.W. Pogue, and K.D. Paulsen, "Near-infrared (NIR) tomography breast image reconstruction with a priori structural information from

MRI: algorithm development for reconstructing heterogeneities," *IEEE J. Sel. Top. Quantum. Electron.*, vol. 9, no. 2, pp. 199–209, 2003.

[15] B. Yu, L. Zhou, L. Wang, Y. Shi, J. Fripp, and P. Bourgeat, "Ea-GANs: edge-aware generative adversarial networks for cross-modality MR image synthesis," *IEEE Trans. Med. Imaging*, vol. 38, no. 7, pp. 1750–1762, 2019.

[16] D.P. Kingma and M. Welling, "Auto-encoding variational Bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[17] G. Kwon, C. Han, and D.S Kim, "Generation of 3D brain MRI using auto-encoding generative adversarial networks," *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019*, vol. 22, no. 3, pp. 118–126, 2019.

[18] C. Li and A.C. Bovik, "Content-partitioned structural similarity index for image quality assessment.," *Signal Processing: Image Communication*, 2010.

[19] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and Ommer, "Analyzing the role of visual structure in the recognition of natural image content with multi-scale SSIM," *Human Vision and Electronic Imaging XIII. vol. 6806, pp. 410–423. SPIE*, 2008.