Enhancing Atlas Construction via Edge-Assisted 3D Latent Diffusion Models for Brain MRI Generation

Sheng Xu and Yi Hong*
Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
*Corresponding author, Email: yi.hong@sjtu.edu.cn.

***Abstract*—Atlas construction is a cornerstone in medical imaging, facilitating research, diagnosis, and personalized treatment planning. In this paper, we introduce an edge-assisted 3D latent diffusion model, Med3Diffusion, trained on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, to enhance atlas construction by synthesizing high-quality brain MRI scans across ages. These generated images augment the original ADNI dataset, addressing the issue of age-related data imbalance and improving the performance of the current atlas-building methods. We employ two state-of-the-art atlas construction methods, i.e., Aladdin and a VoxelMorph-based model, to construct the atlases with this enriched dataset. Our approach not only compensates for the scarcity of certain age groups in the existing dataset but also improves the diversity and quality of the atlas construction. Additionally, experimental results demonstrate the potential of our Med3Diffusion model in generating synthetic 3D medical images for broader applications in medical fields. Our source code is released at https://github.com/sjtu-media/Med3Diffusion.**

I. Introduction

Medical imaging is fundamental to modern diagnostic procedures, providing essential insights into patient anatomy and pathology. Its critical role spans diagnosis, treatment planning, and disease progression monitoring. Central to these applications is the process of atlas construction, which involves creating a standard reference from a collection of individual scans. This reference, or atlas, is indispensable for several reasons. Firstly, it provides a consistent benchmark for interpreting medical images, enabling the detection of deviations from normal anatomy that may signal disease [1]. Also, atlases enhance the accuracy of automated image analysis techniques, e.g., segmentation [2] and anomaly detection [3], by providing a representation of average anatomy. Beyond individual patient care, atlas construction is vital for research, enabling the study of anatomical variations [4] and disease progression [5].

A straightforward method for atlas construction involves choosing a single representative image from the collection [6]. However, this approach often fails to capture the structural diversity and complexity of the entire image collection [7]–[9]. To construct more representative atlases, various optimization-based atlas methods have been proposed [7], [10]. Generally, they start with an initial atlas, such as the collection's average image, and then iteratively minimize the geodesic distance to each image in the collection using image registration techniques. Learning-based approaches for image registration and atlas construction [8], [9] perform image registration by predicting deformation fields through models, thus bypassing the time-consuming numerical optimization process.

Since the construction of atlases relies on datasets of individual scans, a major challenge is the issue of data imbalance. This occurs when available datasets do not uniformly represent the diversity of the population, including variations across different age groups, genders, or health conditions. Such imbalance can lead to biased representations in the constructed atlases, compromising their reliability and applicability.

Recent advancements in diffusion models [11] have demonstrated their ability to generate high-resolution and remarkably realistic natural images. The introduction of stable diffusion [12] has further enabled the controlled generation of high-quality images under various conditions. Such models offer a solution to the data imbalance problem in medical imaging datasets, facilitating the creation of more unbiased and representative atlases. However, generating high-quality 3D medical images, especially under specific conditions like age and gender, is still in its early stages, with only a few studies exploring this area [13], [14].

A significant challenge faced by current methods arises from the unique structural complexity of human tissues and organs, especially in brain MRI scans. When diffusion models are directly applied to medical image generation tasks, they can sometimes produce images with anomalous topological structures or blurred image appearance. These inaccuracies are critical concerns, as maintaining precise structural integrity is essential for reliable medical analysis and atlas building.

To address this challenge, we introduce a medical edge-assisted 3D diffusion model, Med3Diffusion, which integrates edge detection outputs into the generative process of 3D diffusion models. This integration significantly enhances the model's ability to preserve essential structural details in synthesized images. We apply our Med3Diffusion model to generate synthetic brain MRI scans across various age and gender conditions, specifically targeting the mitigation of the data imbalance in existing datasets. By leveraging diffusion models, our method augments datasets by generating data for underrepresented conditions, promoting a more balanced distribution across different ages and genders. This approach not only addresses the scarcity of data in certain subgroups but also improves the overall quality and fairness of atlas construction. Experiments on augmenting two representative atlas building methods, i.e., Aladdin [9] and VoxelMorph-based model [8], demonstrate a marked improvement in the quality and unbiasedness of atlas construction by using our method to balance the learning datasets.

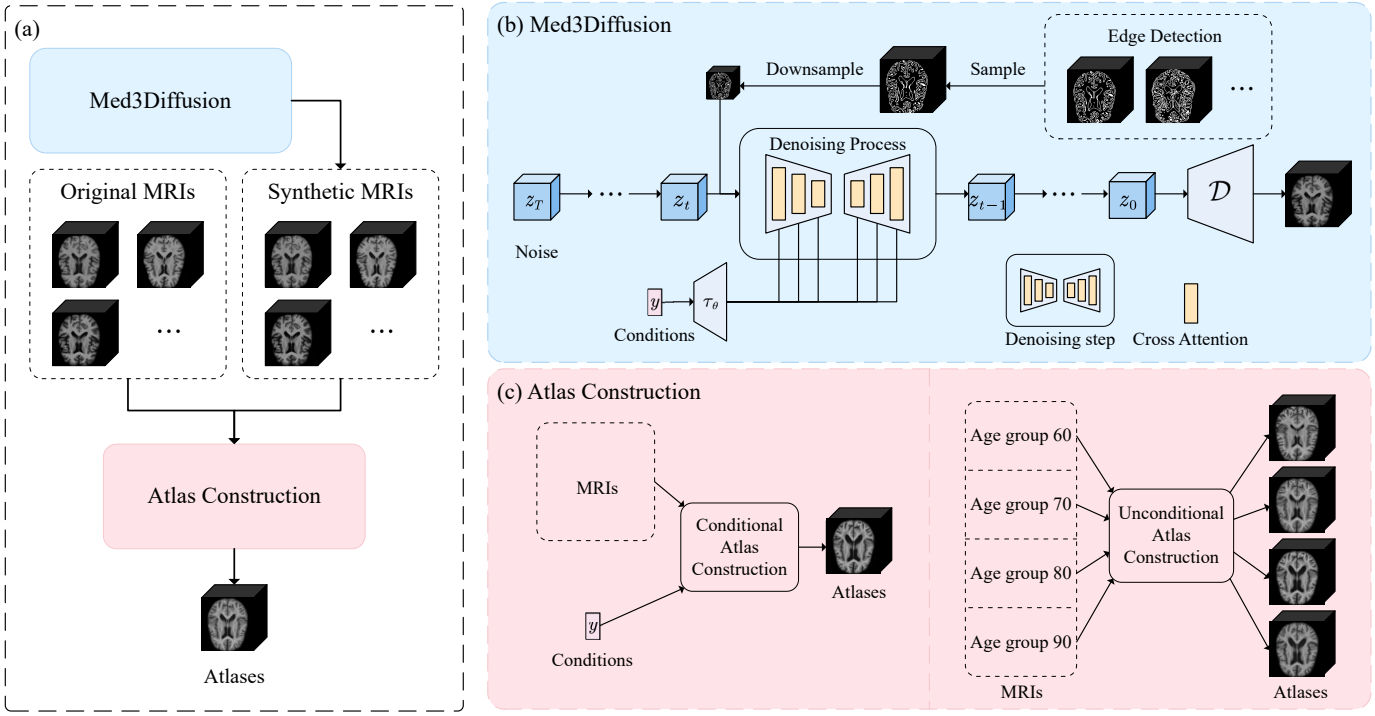

Fig. 1. Overview of our proposed method. (a) Synthetic MRI scans are generated using our Med3Diffusion model, then combined with original scans for atlas construction. (b) The denoising process of our Med3Diffusion for generating MRI scans under given conditions, using edge detection samples from the training set that match these conditions. (c) Two atlas-building strategies using both conditional and unconditional models.

Our contributions in this paper are summarized as follows:

- We propose Med3Diffusion, which integrates edge images of medical volumes into the generative process of the 3D latent diffusion model. This integration significantly enhances the structural integrity of generated images, ensuring a more accurate representation of the intricate anatomical structures essential for atlas construction.
- We apply Med3Diffusion to the brain MRI generation, demonstrating its improved image quality over the original latent diffusion model and its potential to address the data imbalance problem across age and gender conditions.
- We demonstrate a marked improvement in the quality and unbiasedness of atlas construction by augmenting current state-of-the-art (SOTA) atlas building methods.

II. Methodology

In this paper, we focus on the task of building brain atlases across different ages. As illustrated in Fig. 1, we first generate synthetic brain MRIs using our proposed Med3Diffusion, which helps balance the image distribution of a dataset across conditions such as age and gender to build unbiased brain atlases. To enhance the structural integrity of the generated image volumes, we integrate edge images of brain MRIs into the generative process of a 3D condition latent diffusion model (LDM). By combining the original and synthetic MRIs, we create an enriched dataset that is then fed into an atlas construction model, either Aladdin [9] or the model based on VoxelMorph [8], to complete the atlas construction.

A. Med3Diffusion

Considering that medical images, e.g., brain MRIs, are often 3D volumes with millions of voxels, directly using diffusion models to learn the data distribution in this high-dimensional image space would require significant memory usage and computational cost. Following Rombach et al. [12], we employ a 3D Variational Autoencoder (VAE) [15] to map 3D medical scans into latent space. This latent space perceptually corresponds to the original image space but greatly reduces computational complexity. That is, given a 3D scan $\mathbf{x} \in \mathbb{R}^{D\times H\times W}$, the encoder $\mathcal{E}$ of the VAE maps $\mathbf{x}$ to the latent representation $\mathbf{z} = \mathcal{E}(\mathbf{x}) \in \mathbb{R}^{c\times d\times h\times w}$, with downsampling performed at a factor of $f = \frac{D}{d} = \frac{H}{h} = \frac{W}{w}$, and the decoder $\mathcal{D}$ reconstructs the origin scan from $\mathbf{z}$, i.e., $\hat{\mathbf{x}} = \mathcal{D}(\mathbf{z})$.

We pre-train the VAE on our training set collected from the ADNI dataset [16] before training the diffusion model. The loss function of the VAE is defined as

$$\mathcal{L}_{VAE} = \mathcal{L}_{sim}(\mathbf{x}, \hat{\mathbf{x}}) + \lambda \cdot \mathcal{L}_{reg}(\mathbf{z}), \tag{1}$$

where $\mathcal{L}_{sim}$ is the voxel-level similarity term, $\mathcal{L}_{reg}$ is the regularization term, and λ is a constant value to balance these two terms. During the training process, we employ the L_2 norm as the voxel-level similarity between the original image input $\mathbf{x}$ and its reconstruction $\hat{\mathbf{x}}$. Additionally, we incorporate the Kullback-Leibler (KL) divergence between latent $\mathbf{z}$ and the standard normal distribution as a regularization to mitigate the high variance of the data distribution in the latent space.

We extend the diffusion model to 3D and learn the distribution of data within this 3D latent space. During the diffusion process, the diffusion model gradually injects noise into the data through a Markov chain of length T. Conversely, in the reverse denoising process, the model reconstructs the original data from the noise by progressively removing the model-predicted noise. For noise prediction, we utilize a 3D UNet [17] $\epsilon_\theta(\cdot, t)$ as illustrated in Fig. 1. Here, θ represents the learnable parameters and t represents the time step.

To enable our model to generate data under various conditions such as different ages and genders, we introduce the cross-attention mechanism into the UNet architecture, following the approach of Rombach et al. [12]. We encode the conditions of age and gender into numerical vectors $\mathbf{y}$, which are then mapped into conditional embeddings $\tau_\phi(\mathbf{y}) \in \mathbb{R}^{M \times d_\tau}$ using a Multilayer Perceptron (MLP), where M is the sequence length and d_τ is the feature dimension of the embedding. These conditional embeddings $\tau_\phi(\mathbf{y})$ are then interacted with the intermediate layers of UNet through cross-attention layers implemented as follows:

$$\text{Attention}(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d_f}})V \tag{2}$$
$$Q = W_Q^{(i)} \cdot \varphi_i(z_t),\ K = W_K^{(i)} \cdot \tau_\phi(\mathbf{y}),\ V = W_V^{(i)} \cdot \tau_\phi(\mathbf{y})$$

where z_t represents the latent at time step t, $\varphi_i(z_t) \in \mathbb{R}^{N \times d_\epsilon^i}$ denotes the flattened intermediate representation of UNet, i indicates the i-th layer in the UNet, N is the sequence length and d_ϵ^i is its feature dimension, $W_Q^{(i)} \in \mathbb{R}^{d_f \times d_\epsilon^i}$, $W_K^{(i)} \in \mathbb{R}^{d_f \times d_\tau}$, $W_V^{(i)} \in \mathbb{R}^{d_f \times d_\tau}$ are the learnable projection matrices of the cross-attention layer, d_f is the feature dimension used for the cross-attention layer.

As shown in Fig. 2, images generated by LDMs may exhibit structural abnormalities, distortions, or blurred edges. To address this, we consider introducing brain structure as a condition to constrain the image generation of diffusion models. An edge image is a good choice since it contains structural information while being computationally efficient to process. By incorporating edge images derived from Canny edge detection [18] into our diffusion model, we ensure that the generated image maintains crucial structural details.

To work in the latent space, the edge image $e \in \mathbb{R}^{D \times H \times W}$ are downsampled to $e' \in \mathbb{R}^{d \times h \times w}$ and concatenated with noise latent. This combined input is then taken by the UNet, as shown in Fig. 1. During the training process, we have the edge image of a training sample as input. During the inference phase, we leverage edge images from the training set that match or are similar to the target conditions, facilitating the generation of images that better align with specific structural requirements in that age or gender group.

We train Med3Diffusion by optimizing the loss function:

$$\mathcal{L} = \mathbb{E}_{\mathcal{E}(\mathbf{x}), \mathbf{y}, \epsilon \sim \mathcal{N}(0,I), t} \left[\|\epsilon - \epsilon_\theta(z_t, e', \tau_\phi(\mathbf{y}), t)\|_2^2 \right], \tag{3}$$

where I is the identity matrix, t is the time step uniformly sampled from $\{1, \cdots, T\}$, and T is the pre-defined total time step of the diffusion model.

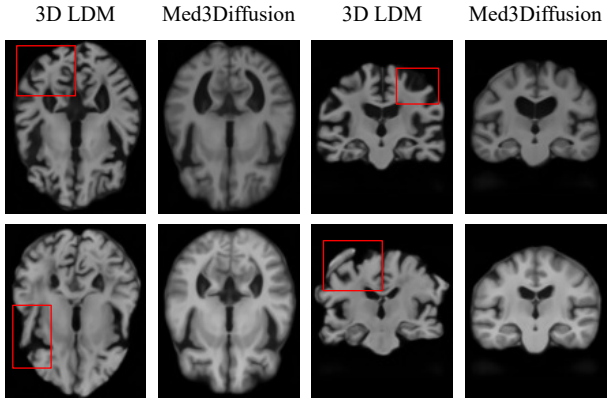

Fig. 2. Visualization of image scans generated by LDM and Med3Diffusion in two views. Images in each row are generated under identical conditions by 3D LDM and Med3Diffusion, respectively. Scans generated by LDM may exhibit structural abnormalities or distortions, as highlighted in the red boxes.

B. Atlas Construction

We utilize a well-trained Med3Diffusion to generate image scans under various conditions, thereby augmenting existing datasets where certain conditions are underrepresented. By incorporating the synthetic 3D scans with the original ones to form an augmented dataset, we can use this enriched dataset for atlas building based on current atlas construction models. In this paper, we employ two distinct approaches for atlas construction: Aladdin [9] and a VoxelMorph-based model [8].

Aladdin [9] is a joint atlas construction and diffeomorphic registration framework that operates without specific conditions, generating a single atlas for the entire image collection. We divide the augmented dataset into several age groups, allowing Aladdin to create a specific atlas for each age group. Hence, we leverage the diversity and balance of the augmented dataset to ensure adequate representation of age groups that initially had sparse data. Without the augmented dataset, using Aladdin on just the original dataset probably results in biased atlases for certain age groups due to insufficient data.

The VoxelMorph-based model in [8] is capable of generating condition-specific atlases based on factors such as age and gender. This innovative approach integrates the atlas generation process with the subsequent registration network [19], training them simultaneously. As a result, it generates an atlas under a given condition and then warps the atlas to images corresponding to that condition. Training this model with an augmented dataset mitigates potential biases arising from disparities in data distribution across different conditions in the original dataset. This approach ensures a more balanced and accurate representation across diverse conditions.

III. Experiments

A. Dataset and Experimental Settings

Dataset. We utilize a dataset from the Alzheimer's Disease Neuroimaging Initiative (ADNI) [16] comprising 1417 brain MRI scans labeled as Cognitively Normal (CN) for our experiments. The dataset is subject-wisely divided into training and

TABLE I
DICE SCORES BETWEEN SEGMENTATION MASKS OF THE SYNTHETIC SCANS AND DEFORMED SCANS IN THE TEST SET. THE BEST RESULT IS IN BOLD.

Model	Total (N=440)	Age 60 (N=8)	Age 70 (N=170)	Age 80 (N=217)	Age 90 (N=45)
3D LDM	58.20±23.99	60.13±19.19	56.99±24.01	59.63±24.03	55.82±23.88
Med3Diffusion (ours)	**70.91±20.79**	**74.77±19.03**	**70.26±21.12**	**71.76±19.96**	**69.04±22.92**

TABLE II
DICE SCORES BETWEEN SEGMENTATION MASKS OF THE SCANS IN THE TEST SET AND THE ATLASES DEFORMED TO THEM. THE BEST RESULT IS IN BOLD.

Method	MRI Generation	Total (N=440)	Age 60 (N=8)	Age 70 (N=170)	Age 80 (N=217)	Age 90 (N=45)
Aladdin [9]	No	76.27±12.25	73.41±18.21	75.13±13.30	77.33±10.94	75.95±12.34
	3D LDM	76.17±12.06	75.16±12.82	75.13±12.99	77.50±11.03	73.88±12.31
	Med3Diffusion (ours)	**77.20±11.83**	**77.44±11.99**	**76.11±13.04**	**77.91±10.79**	**77.85±11.47**
Dalca et al. [8]	No	77.93±10.69	76.59±11.12	76.93±11.87	78.63±9.77	78.55±9.78
	3D LDM	77.57±10.44	76.40±10.87	76.80±11.59	78.20±9.52	77.66±9.73
	Med3Diffusion (ours)	**78.35±10.60**	**77.32±10.92**	**77.47±11.79**	**78.99±9.70**	**78.75±9.68**

test sets, with 977 and 440 scans, respectively. Furthermore, as Aladdin [9] operates without specific conditions, we divide the dataset into four age groups: 60 (i.e., [55, 64)), 70 (i.e., [65, 74)), 80 (i.e., [75, 84)), and 90 (i.e., [85, 94)). That is, for each age group i, it contains brain MRI scans collected from subjects aged between $i-5$ and $i+5$. All scans are of the image shape $138\times176\times138$. During the experimental process, scans are padded and then downsampled to $104\times128\times104$ to fit into a single RTX 3090 GPU.

Image Generation. We train Med3Diffusion using medical scans from the original dataset along with their corresponding edge images detected by the Canny algorithm. After training, we use the trained Med3Diffusion to generate synthetic scans under various age and gender conditions to augment the original dataset. During the generation process, the model samples edge image volumes from the training set that match or closely resemble the specific conditions under which the new images are being generated. This ensures that the augmented dataset contains 100 scans for each age from 55 to 95, with an equal distribution of 50 scans for both males and females.

We then employ the entire augmented dataset to train the VoxelMorph-based model proposed by Dalca et al. [8], enabling it to generate specific atlases under various conditions. For Aladdin [9], we also divide the augmented dataset into four age groups, training Aladdin model separately to construct specific atlases for each age group.

Experimental Settings. We implement Med3Diffusion with PyTorch. We set the downsampling factor $f=4$, the number of latent channels $c=2$, the regularization coefficient $\lambda=10^{-7}$ in the VAE loss function. The results of Canny edge detection are obtained using the Canny image filter implemented in SimpleITK [20], with the low threshold set at 0.05 and the high threshold set at 0.1. The edge detection results are directly downsampled by $f=4$ and then concatenated with noise latent in the latent space. Initially, we train the VAE using AdamW optimizer for 300 epochs. Subsequently, we freeze the VAE and train the UNet in latent space using AdamW for 500 epochs to ensure convergence. Both VAE and UNet are trained on a single RTX 3090 GPU with a learning rate of 10^{-4}, a batch size of 2, and a weight decay of 10^{-2}.

B. Experimental Results

In assessing the quality of scans generated by our Med3Diffusion, we employ the Fréchet Inception Distance (FID) as a metric. The FID is calculated using an approach similar to Pinaya et al. [14], where features are extracted via a pre-trained MedicalNet [21]. The FID for our proposed Med3Diffusion is 2.05e-4, indicating superior image quality due to a lower FID compared to that of the 3D LDM without edge detection integration, which records a FID of 9.34e-4.

To further demonstrate the structural integrity of image scans generated using our Med3Diffusion compared to those generated directly by the 3D LDM, we generate images for the test dataset under specific conditions using both models. We then obtain segmentation masks for these scans using a pre-trained brain MRI segmentation model SynthSeg [22]. Subsequently, we employ a pre-trained registration model [19] to calculate the deformation field from the image scans in the test dataset to the generated ones. By deforming the segmentation masks of the test images to match those of the generated scans, we compute the dice score to measure their structural overlaps. This allows us to quantitatively compare the structural integrity of the scans generated by the two models. The results, reported in Table I, significantly outperform 3D LDM across all age groups, indicating the scans generated by Med3Diffusion exhibit greater structural integrity. This also highlights the superior capability of our Med3Diffusion model in maintaining the detailed and accurate structural features necessary for effective medical analysis and diagnosis. Fig. 2 presents the generated medical volumes using the 3D LDM and our Med3Diffusion model. Some structures in the images generated by the 3D LDM appear incomplete

or less well-defined, especially in the coronal view; while our model exhibits better structural integrity.

C. Atlas Construction Results

We evaluate the quality of the atlases constructed using different approaches by comparing the dice scores between the anatomical segmentation masks of the scans in the test set and the anatomical segmentation masks of the atlases registered to these scans. The segmentation masks are obtained using SynthSeg [22]. We train the Voxelmorph-based model [8] separately on the original dataset and datasets augmented with scans generated either by the LDM or our Med3Diffusion. Additionally, we train twelve specific Aladdin [9] models for each of the four age groups after dividing these three datasets accordingly. The performance of the atlases constructed by these trained models across the entire test set and within each age group is reported in Table II. The results indicate that both VoxelMorph-based [8] and Aladdin [9] atlas construction methods show consistent improvements after utilizing the dataset augmented by our Med3Diffusion model.

More importantly, using the dataset augmented by Med3Diffusion for atlas construction shows more pronounced improvements in age groups with a limited number of scans in the original training set, such as the age groups of $60s$ and $90s$. Take the Age 60 and 90 columns of Table II for example. Using our Med3Diffusion and Aladdin for atlas building, the Dice score improves from 73.41% to 77.44% for age 60 and from 75.95% to 77.85% for age 90, respectively. This advancement is primarily due to the mitigation of biases in these age groups, achieved by adding data generated by Med3Diffusion. Consequently, our method results in the construction of more unbiased and representative atlases for these age groups.

On the other hand, the results of the atlases built using the dataset augmented by 3D LDM closely resemble those obtained using the original dataset, suggesting that the contribution of LDM-generated data to atlas construction is not significantly beneficial and may, in some instances, even be counterproductive. This underscores the necessity for cautious consideration when using generative models to augment realistic datasets and further highlights the superiority of our Med3Diffusion assisted by edge images of medical volumes.

The dataset augmented by Med3Diffusion can also improve the robustness of atlas construction methods, especially in age groups that are underrepresented in the original dataset. As shown in Fig. 3, atlases constructed using the original dataset by Aladdin may exhibit instability in these age groups. When registering these atlases to certain images, abnormal shrinkage and deformation can occur in the brain's pallidum and putamen regions, deviating significantly from the target images. However, training Aladdin with the Med3Diffusion-augmented dataset does not exhibit such instability.

Fig. 4 visualizes the brain atlases for ages 60, 70, 80, and 90 years estimated by Aladdin [9] and VoxelMorph-based model [8]. It can be observed that the atlases constructed using the augmented dataset exhibit clearer textures and less blurry edges compared to those built only with the original dataset.

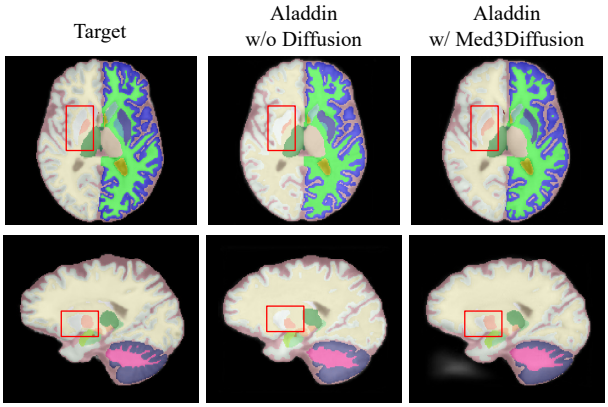

Fig. 3. Segmentation masks of target images, atlases constructed by Aladdin registered to target images, trained on the original and Med3Diffusion-augmented datasets, respectively. Regions in red boxes: pallidum and putamen.

With the aid of synthetic data generated by Med3Diffusion, the estimated atlases display clearer and more defined textures, particularly evident in the fine details of the brain structures.

IV. Conclusion and Discussion

In this study, we introduce a Med3Diffusion model to augment the ADNI dataset, thereby constructing more representative and unbiased atlases. Our experiments have demonstrated that utilizing the augmented dataset leads to improvements in existing atlas construction methods, such as the Aladdin or VoxelMorph-based model, particularly for groups that were previously underrepresented in the original dataset. Our research highlights the potential of using Med3Diffusion for data augmentation in the field of medical image analysis. Due to the GPU memory issue, we have to reduce the 3D image volume size to some extent in this work. In future work, we will explore the potential of generating larger image volumes with limited computational resources.

Acknowledgment

This research work was supported by NSFC 62203303 and Shanghai Municipal Science and Technology Major Project 2021SHZDZX0102.

References

[1] M. Liu, D. Zhang, D. Shen, and ADNI, "View-centralized multi-atlas classification for alzheimer's disease diagnosis," *Human brain mapping*, vol. 36, no. 5, pp. 1847–1865, 2015.

[2] P. Aljabar, R. A. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert, "Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy," *Neuroimage*, vol. 46, no. 3, pp. 726–738, 2009.

[3] A. V. Faria, J. Zhang, K. Oishi, X. Li *et al.*, "Atlas-based analysis of neurodevelopment from infancy to adulthood using diffusion tensor imaging and applications for automated abnormality detection," *Neuroimage*, vol. 52, no. 2, pp. 415–428, 2010.

[4] P. Roland, C. Graufelds, J. Wählin, L. Ingelman, M. Andersson, A. Ledberg, J. Pedersen, S. Åkerman, and A. Dabringhaus, "Human brain atlas: for high-resolution functional and anatomical mapping," *Human brain mapping*, vol. 1, no. 3, pp. 173–184, 1994.

[5] P. M. Thompson, M. S. Mega, R. P. Woods, C. I. Zoumalan, C. J. Lindshield, R. E. Blanton, J. Moussai, C. J. Holmes, J. L. Cummings, and A. W. Toga, "Cortical change in alzheimer's disease detected with a disease-specific population-based brain atlas," *Cerebral Cortex*, vol. 11, no. 1, pp. 1–16, 2001.

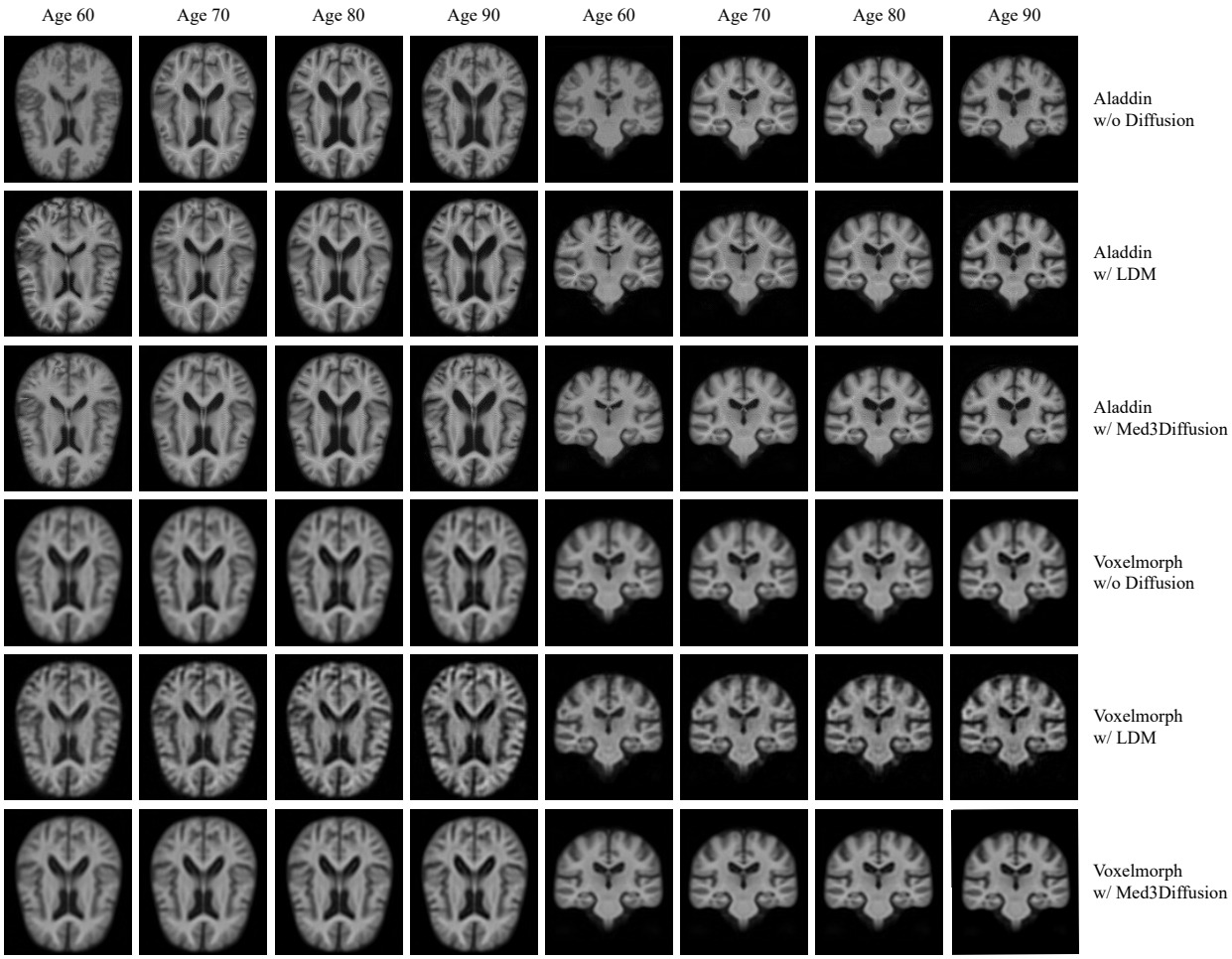

Fig. 4. Visualization of brain atlases generated by Aladdin [9] and VoxelMorph-based model [8], without diffusion and with 3D LDM and our Med3Diffusion.

[6] S. Ho, E. Bullitt, and G. Gerig, "Level-set evolution with region competition: automatic 3-d segmentation of brain tumors," in *International Conference on Pattern Recognition*, vol. 1. IEEE, 2002, pp. 532–535.

[7] S. Joshi, B. Davis, M. Jomier, and G. Gerig, "Unbiased diffeomorphic atlas construction for computational anatomy," *NeuroImage*, vol. 23, pp. S151–S160, 2004.

[8] A. Dalca, M. Rakic, J. Guttag, and M. Sabuncu, "Learning conditional deformable templates with convolutional networks," *Advances in neural information processing systems*, vol. 32, 2019.

[9] Z. Ding and M. Niethammer, "Aladdin: Joint atlas building and diffeomorphic registration learning with pairwise alignment," in *Proceedings of the IEEE/CVF CVPR*, 2022, pp. 20 784–20 793.

[10] P. Lorenzen, B. C. Davis, and S. Joshi, "Unbiased atlas formation via large deformations metric mapping," in *MICCAI Proceedings, Part II 8*. Springer, 2005, pp. 411–418.

[11] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *NeurIPS*, vol. 33, pp. 6840–6851, 2020.

[12] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF CVPR*, 2022, pp. 10 684–10 695.

[13] W. Peng, E. Adeli, T. Bosschieter, S. H. Park, Q. Zhao, and K. M. Pohl, "Generating realistic brain mris via a conditional diffusion probabilistic model," in *MICCAI Proceedings*. Springer, 2023, pp. 14–24.

[14] 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," in *MICCAI Workshop on Deep Generative Models*. Springer, 2022, pp. 117–126.

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

[16] S. G. Mueller, M. W. Weiner, L. J. Thal, R. C. Petersen, C. R. Jack, W. Jagust, J. Q. Trojanowski, A. W. Toga, and L. Beckett, "Ways toward an early diagnosis in alzheimer's disease: the alzheimer's disease neuroimaging initiative (adni)," *Alzheimer's & Dementia*, vol. 1, no. 1, pp. 55–66, 2005.

[17] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI proceedings, part III 18*. Springer, 2015, pp. 234–241.

[18] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 6, pp. 679–698, 1986.

[19] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "Voxelmorph: a learning framework for deformable medical image registration," *IEEE transactions on medical imaging*, vol. 38, no. 8, pp. 1788–1800, 2019.

[20] Z. Yaniv, B. C. Lowekamp, H. J. Johnson, and R. Beare, "Simpleitk image-analysis notebooks: a collaborative environment for education and reproducible research," *Journal of digital imaging*, vol. 31, no. 3, pp. 290–303, 2018.

[21] S. Chen, K. Ma, and Y. Zheng, "Med3d: Transfer learning for 3d medical image analysis," *arXiv preprint arXiv:1904.00625*, 2019.

[22] B. Billot, D. N. Greve, O. Puonti, A. Thielscher, K. Van Leemput, B. Fischl, A. V. Dalca, J. E. Iglesias *et al.*, "Synthseg: Segmentation of brain mri scans of any contrast and resolution without retraining," *Medical image analysis*, vol. 86, p. 102789, 2023.