Synchronous Image-Label Diffusion with Anisotropic Noise for Stroke Lesion Segmentation on Non-contrast CT

Jianhai Zhang[1], Tonghua Wan[2], M. Ethan MacDonald[3], Bijoy K Menon[1], Wu Qiu[2⋆], and Aravind Ganesh[1]

[1] Department of Clinical Neurosciences, Cumming School of Medicine, University of Calgary, Alberta, Canada
[2] School of Life Science and Technology, Huazhong Science of Science and Technology, Wuhan, China
[3] Department of Radiology, Cumming School of Medicine, University of Calgary, Alberta, Canada

Abstract. Automated segmentation of stroke lesions on non-contrast CT (NCCT) images is essential for efficient diagnosis of stroke patients. Although diffusion probabilistic models have shown promising advancements across various fields, their application to medical imaging exposes limitations due to the use of conventional isotropic Gaussian noise. Isotropic Gaussian noise overlooks the structural information and strong voxel dependencies in medical images. In this paper, a novel framework employing synchronous diffusion processes on image-labels is introduced, combined with a sampling strategy for anisotropic noise, to improve stroke lesion segmentation performance on NCCT. Our method acknowledges the significance of anatomical information during diffusion, contrasting with the traditional diffusion processes that assume isotropic Gaussian noise added to voxels independently. By integrating correlations among image voxels within specific anatomical regions into the denoising process, our approach enhances the robustness of neural networks, resulting in improved accuracy in stroke lesion segmentation. The proposed method has been evaluated on two datasets where experimental results demonstrate the capability of the proposed method to accurately segment ischemic infarcts on NCCT images. Furthermore, comparative analysis against state-of-the-art models, including U-net, transformer, and DPM-based segmentation methods, highlights the advantages of our method in terms of segmentation metrics. The code is publicly available at https://github.com/zhangjianhai/SADPM.

1 Introduction

The primary objective of automated computer-aided analysis is to enable algorithms to deliver swift, accurate, and reliable results to support clinical diagnoses. In medical image segmentation for stroke lesions, the conventional

⋆ Corresponding Author

methods involve manual contouring of images for volume measurements. This process is time-consuming and sensitive to observer dependence. In response, computer-aided techniques have emerged to enhance efficiency and eliminate observer bias, showing substantial potential for automated stroke lesion segmentation [1][14][23].

Recently, the variants of denoising diffusion probabilistic models (DDPMs) have shown an improvement and future potential for segmentation of medical images for a broad range of applications [22][27]. The success of DDPMs is attributed to an efficient recursive denoising mechanism that is robust against a high degree of noise disturbances [30]. However, most of the applications are demonstrated in the domain of natural images, such as: portraits, paysage, and camera surveillance, where isotropic Gaussian noise is usually assumed and used in the denoising process. By contrast, medical images possess unique characteristics, stable image structure, immobilized intensities and texture etc., that make isotropic noise unsuitable for segmentation tasks. To this end, we propose a specific kind of anisotropic noise, allowing for the correlations among coordinates in the cerebral images, potentially improving segmentation performance. Furthermore, there is rare research on such specific type of noise.

Specifically, cerebral images typically have voxels with a relatively standard atlas structural information [18][7] that can be used as a prior, whereas natural images are more unconstrained. For instance, the facial recognition will deal with the portraits with a variety of poses, illuminations, zoom and sizes, etc. A limitation of current models based on DDPMs is using isotropic Gaussian noise, which assume that the voxels are independent so that some useful prior information regrettably fails to be utilized.

In this paper, a fully probabilistic inference framework is proposed for stroke lesion segmentation on non-contrast CT scans based on a synchronous image-label denoising diffusion model with anisotropic Gaussian noise [28][8]. A variational inference is developed to efficiently train the model. A set of related strategies on synchronous diffusion process to image-labels is proposed to eventually infer the lesion labels. For sampling anisotropic Gaussian noises, since each noise sampling is an element located on a high-dimensional symmetric positive definite manifold (Riemannian manifold), it is impossible to get the sampling instances immediately from such a matrix distribution [3]. To implement the efficient sampling, an approximate sampling algorithm using Random Fourier Features with time effectiveness is developed to obtain the anisotropic noise for our segmentation tasks. The segmentation framework is illustrated in Fig.1.

2 Methodology

2.1 Semantic Segmentation of Medical Images

We extend the diffusion model to segmentation tasks with the form $p_\theta(y_0|x_0) = p_\theta(y_0, x_0)/p_\theta(x_0) \propto p_\theta(y_0, x_0)$, where $p_\theta(y_0, x_0) \triangleq \int p_\theta(x_{0:T}, y_{0:T}) dx_{1:T}, y_{1:T}$, and $y_{1:T}$ are the latent variables. The joint distribution $p_\theta(x_{0:T}, y_{0:T})$ is defined as the reverse process, and it can be factorized as an original DDPM [11]

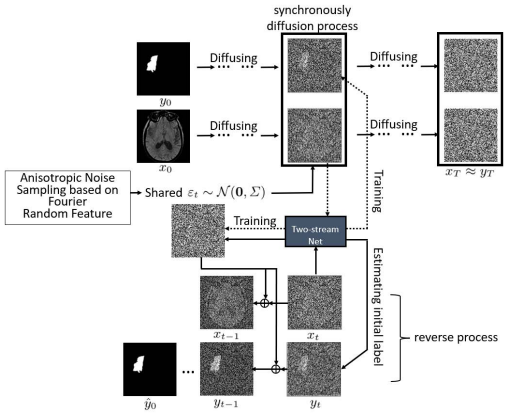

Fig. 1. Synchronous diffusion process and reverse process with anisotropic noise for image segmentation tasks

part $p_\theta(x_{0:T})$ and a conditional reverse process $p_\theta(y_{0:T}|x_{0:T})$, which is a series of terms by Markov chains starting at $p_\theta(y_T|x_T)$:

$$p_\theta(y_{0:T}|x_{0:T}) = p_\theta(y_T|x_T)\prod_{t=1}^{T} p_\theta(y_{t-1}|y_t, x_{t-1})$$

The diffusion process is an approximate conditional posterior $q(y_{1:T}|y_0, x_{0:T})$:

$$q_(y_{1:T}|y_0, x_{0:T}) = \prod_{t=1}^{T} q(y_t|y_{t-1}, x_t)$$

A KL Divergence is defined between the diffusion process $q_(y_{1:T}|y_0, x_{0:T})$ and the reverse process $p_\theta(y_{0:T}|x_{0:T})$ to obtain the variational upper bound on the negative log likelihood because of non-negative property of KL divergence. After simplifying, the loss is represented as:

$$\mathcal{L} \triangleq \mathbb{E}_q\Big[\log\frac{p_\theta(y_0|x_0)}{p_\theta(y_T|x_T)} + \sum_{t=1}^{T}\log\frac{q(y_t|y_{t-1}, x_t)}{p_\theta(y_{t-1}|y_t, x_{t-1})}\Big] \tag{1}$$

To optimize the loss, an additional sub-network for estimating the label with added noise y_T from the image x_T is introduced into the backbone of the network.

In (1), however, using y_T as the start point of inference could produce a poor result and degrade the segmentation performance. Thus, we obtain a suitable initial y_t for inferring y_0. To this end, we add a time window of length T_p to train the model at each time, which is the loss $\mathcal{L}_p$ in (2), guaranteeing that the label y_0 could be efficiently restored. With the initial y_t is predicted in the model, the loss function is further simplified as:

$$\mathcal{L} = \underbrace{\mathbb{E}_q\Big[\sum_{t=2}^{T}\log\frac{q(y_{t-1}|y_t, y_0, x_t)}{p_\theta(y_{t-1}|y_t)}\Big]}_{\mathcal{L}_d} - \underbrace{\sum_{t=0}^{T_p}\log p_\theta(y_t|x_t)}_{\mathcal{L}_p} - \underbrace{\log p_\theta(y_0|y_1)}_{\mathcal{L}_{d_0}} \tag{2}$$

Algorithm 1 Label Inference for Semantic Segmentation

Input: image x_0, **Parameters**: $d_i \in \mathbb{R}^{(0,1)}$, $T_i < T$
Initialization: noised image $\hat{x}_{T_i} = x_{T_i}$ and noised label $\hat{y}_{T_i}$ using (3)
estimate noise $\hat{\varepsilon}_{T_i}$ using the trained neural network
for $t = T_i, \cdots, 1$ **do**
    estimate $\hat{x}_0$, $\hat{y}_0$ using Eq.(3)
    if $t > 1$ **then**
        sampling $\varepsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$: estimate $\hat{x}_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\hat{x}_0 + \frac{\sqrt{\alpha_t}\gamma_{t-1}}{\sqrt{\gamma_t}}\hat{\varepsilon}_t + \frac{\beta_t\gamma_{t-1}}{\gamma_t}\varepsilon$
          and $\hat{y}_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\hat{y}_0 + \frac{\sqrt{\alpha_t}\gamma_{t-1}}{\sqrt{\gamma_t}}\hat{\varepsilon}_t + \frac{d_i\beta_t\gamma_{t-1}}{\gamma_t}\varepsilon$
        update next loop noise $\hat{\varepsilon}_{t-1}$ from $\hat{x}_{t-1}$ using the trained neural network
    else if $t = 1$ **then**
        estimate $\hat{y}_0$ using Eq.(3), $\hat{y}_0^{\mathbf{final}} = \sqrt{\bar{\alpha}_{t-1}}\hat{y}_0 + \frac{\sqrt{\alpha_t}\gamma_{t-1}}{\sqrt{\gamma_t}}\hat{\varepsilon}_t$
Output: estimate label $\hat{y}_0^{\mathbf{final}}$

For the loss $\mathcal{L}_d$, the term $p_\theta(y_{t-1}|y_t)$ in the reverse process is still compared using KL divergence by the conditional posterior $q(y_{t-1}|y_t, y_0, x_t)$, as introduced in [11]. To make the distribution $q(y_{t-1}|y_t, y_0, x_t)$ tractable, a further assumption is made that images and labels are both overlain with the same Gaussian noises during the diffusion process. Therefore, the posterior $q(y_t|y_{t-1}, y_0, x_t)$ is degenerated as $q(y_{t-1}|y_0, \varepsilon_t)$, which only relies on the noise at time t and y_0:

$$q(y_{t-1}|y_0, \varepsilon_t) = \mathcal{N}(y_{t-1}|\tilde{\mu}_t, \tilde{\beta}_t\mathbf{\Sigma})$$
$$\texttt{where } \tilde{\mu}_t = \sqrt{\bar{\alpha}_{t-1}}y_0 + \sqrt{\alpha_t}(\gamma_{t-1}\gamma_t^{-\frac{1}{2}})\varepsilon_t, \;\; \tilde{\beta}_t = \beta_t\gamma_{t-1}\gamma_t^{-1}$$

Since the diffusion process for images and labels share the same noise $\varepsilon_t \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$, the images and labels at time t can thus be sampled as:

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{\gamma_t}\varepsilon_t, \;\; y_t = \sqrt{\bar{\alpha}_t}y_0 + \sqrt{\gamma_t}\varepsilon_t \tag{3}$$

The loss $\mathcal{L}_{d_0}$ is used for optimizing the last layer of generating y_0. During the inference, there may be many cumbersome discrete segmentation points that degrade the performance to some degree. To limit the influence of noise to this effect, in addition to applying the same strategy to stop adding uncertainty inference of the variance $\tilde{\beta}_1$ in [11], a convolutional layer is also appended with frozen parameters of all ones to get rid of the spurious points with a threshold. Finally, the segmentation label can be inferred by **Algorithm 1**.

2.2 Anisotropic Noise and Approximate Sampling on High-dimensional Manifold

The acquisition of isotropic Gaussian noise is simple and straightforward, i.e., $\varepsilon_t \sim \mathcal{N}(\mathbf{0}, \Sigma)$ and Σ is set to be $\mathbf{I}$. We argue that a kind of customized noise is more suitable for medical images, where the voxels are holding a fixed structure information, such as the brain atlas for stroke lesion. Recently work [6][24]

has demonstrated that isotropic noise can work well in natural images. Unfortunately, the isotropic noise assumes that the voxels are independent so that the useful structural prior information regrettably fails to be utilized. Considering the correlation among voxels into the denoising process is able to efficiently help optimize more robust neural networks, yielding a potential improvement to some degree on the specific stoke lesion segmentation tasks on non-contrast CT scans.

2.3 Sampling Anisotropic Noise on High-dimensional Manifold

Since anisotropic covariance Σ is of the manifold [13] with the dimension $d_\Sigma = |\mathcal{SPD}_{++}^{\Sigma}| = \frac{d(d-1)}{2}$, where $d = L \times H \times W$. It is impossible to conduct sampling immediately from $\mathcal{N}(\mathbf{0}, \Sigma)$ on a such high-dimensional distribution. To make this tractable, an approximate algorithm using Random Fourier Features [21] (RFF) is developed to obtain anisotropic noise samplings.

Bochner Theorem[19]: A continuous shift-invariant kernel $k(x^n, x^m) = k(x^n - x^m)$ on $\mathbb{R}^d$ is positive definite if and only if $k(\delta)$ is the Fourier transform of a non-negative measure. □

The Bochner theorem states that the Fourier transform $p(\omega)$ of a non-negative measure $k(\delta)$ is a proper probabilistic density distribution. Thus, a shift-invariant kernel $k(\nabla)$ can be represented with $\zeta_\omega(x^n) = e^{j\omega^T x^n}$:

$$k(x^n - x^m) = \int_{\mathbb{R}^d} p(\omega) e^{j\omega^T (x^n - x^m)} d\omega = \mathbb{E}[\zeta_\omega(x^n)\zeta_\omega(x^m)^*] \tag{4}$$

$\zeta_\omega(x^n)\zeta_\omega(x^m)^*$ is an unbiased expectation of $k(x, y)$ with distribution $p(\omega)$. For the term $e^{j\omega^T(x^n - x^m)}$, a real-valued function $z_\omega(x^n) = \sqrt{2}\texttt{cos}(\omega^T x^n + b)$ is suggested to satisfy the expectation condition $k(x^n, x^m) = \mathbb{E}[z_\omega(x^n)^T z_\omega(x^m)]$, where ω is i.i.d. samples from $p(\omega) \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and b is sampled uniformly from $[0, 2\pi]$. As such, with $\{\omega_k\}_{k=1}^D$ and $\{b_k\}_{k=1}^D$, the D-dimensional mapping is:

$$z_{RFF}(x^n) = \sqrt{\frac{2}{D}}\big[\texttt{cos}(\omega_1^T x^n + b_1); \cdots; \texttt{cos}(\omega_D^T x^n + b_D)\big] \in \mathbb{R}^D \tag{5}$$

Thus, with the mapping $z_{RFF} : \mathbb{R}^d \to \mathbb{R}^D$, the covariance estimate is $\hat{\Sigma}$:

$$\hat{\Sigma}_D = Z_\mathcal{D} J J^T Z_\mathcal{D}^T \in \mathcal{SPD}_{++}^D \ \texttt{where} \ Z_\mathcal{D} \in \mathbb{R}^{D \times N} \tag{6}$$

where $J = N^{-\frac{3}{2}}(N\mathbf{I} - \mathbf{1}_N \otimes \mathbf{1}_N)$ is a centering matrix[4], $\mathbf{1}_N$ is a column vector of N ones, $\otimes$ is matrix outer-product, $Z_\mathcal{D}$ is a matrix where the columns are stacked by random Fourier features (5) over image samples. Therefore, one can easily obtain anisotropic noise/samples from $\hat{\Sigma}_D$ with the Cholesky decomposition[25]:

$$\varepsilon_A = Lu, \ \ \hat{\Sigma}_D = LL^T, \ \ u \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \tag{7}$$

where L is a lower triangular matrix.

If mapping the low-dimensional sample ε_A back to the original high-dimensional space, the scope values of $\omega^T x + b$ needs to be restricted because the function

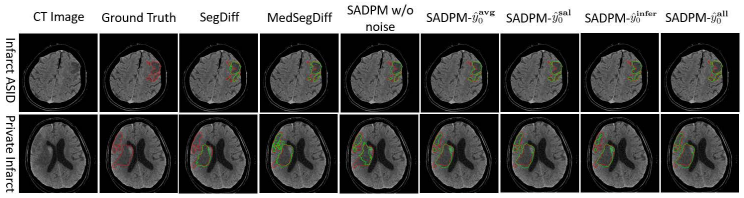

Fig. 2. Visual results from infarct lesion segmentations on ASID and Private Infarct

mapping $\texttt{arccos}(\cdot)$ is not invertible or unique in terms of $\texttt{cos}(\cdot)$. The proposed strategy is to map $\omega^T x + b$ into $\mathbb{R}^{[0,\pi]}$ as well as it can be. Thus, the sampling distribution for ω and b are adjusted as $\mathcal{N}(\mathbf{0}, \frac{\pi}{4D}\mathbf{I})$ and $[\frac{\pi}{4}, \frac{3\pi}{4}]$ in Eq.(5). The reverse mapping is represented as:

$$\hat{x}(\varepsilon_A) \approx W^T \Big[\texttt{arccos}\Big(\sqrt{\frac{D}{2}}\varepsilon_{A,1}\Big) - b_1; \cdots; \texttt{arccos}\Big(\sqrt{\frac{D}{2}}\varepsilon_{A,D}\Big) - b_D\Big] \in \mathbb{R}^d \quad (8)$$

where $W \in \mathbb{R}^{D\times d}$ is the matrix with entries sampled from $\mathcal{N}(\mathbf{0}, \frac{\pi}{4D}\mathbf{I})$. Since the matrix W is not square and there is no strict inverse matrix, the property of $W^T W \approx \mathbf{I_d}$ is used to approximate fit the estimate $\hat{x}$ as the anisotropic noise.

Table 1. Quantitative Performance on **Infarct AISD** with Spatial Overlap Metrics

Datasets	**Infarct AISD**				**Private Infarct**			
Methods	Dice	Recall	Pre.	AUC	Dice	Recall	Pre.	AUC
SegResNet[15]	0.425	0.408	0.513	0.750	0.343	0.330	0.443	0.706
UNETR[10]	0.388	0.402	0.435	0.714	0.328	0.327	0.392	0.684
SwinUNETR[9]	0.420	0.394	0.549	0.761	0.366	**0.511**	0.357	0.668
nnUNet[12]	0.434	0.423	0.508	0.750	0.337	0.360	0.465	0.692
nnUNet++[31]	0.457	0.462	0.534	0.752	0.371	0.398	0.420	0.699
SegDiff[2]	0.462	0.413	0.661	0.705	0.396	0.367	0.649	0.681
MedSegDiff[26]	0.495	0.457	0.656	0.727	0.404	0.360	0.606	0.678
SDPM[11]	0.495	0.448	0.647	0.722	0.420	0.367	0.703	0.681
Our SADPM	**0.524**	**0.530**	**0.686**	**0.763**	**0.453**	0.450	**0.706**	**0.725**

3 Experiments

3.1 Dataset, Pre-processing and Hyper-parameters Settings

Two datasets were used: 1) A public **Infarct AISD** segmentation dataset is used, where 273 Non-Contrast-enhanced CT (NCCT) scans (5 mm slice thickness) of acute ischemic stroke with the interval from symptom onset to CT less than 24 hours are used for training and validating the model, and the 80 participants

are used for testing. 2) A private dataset, named **Private Infarct**, containing 195 AIS patient NCCT scans (5 mm) were included. Of 195 patients, 123 images were used for training while the remained 72 for testing.

The skulls were removed for NCCT images using the network[16]. Then, the NCCT images were trimmed to the identical size of 448-by-448 by center crop along z-axis. Online data augmentations was performed, including adding noise, rotations, scalings, inplane flipping, etc. The variance schedule β_t is the sigmoid curve[17]. P2 Weighting coefficients during the training is suggested in [5]. The Adam optimizer is used with the learning rate of 1e-4.

Table 2. Quantitative Performance on **Infarct AISD** and **Private Infarct** based on Segmentation Metrics

Datasets	**Infarct AISD**				**Private Infarct**			
Methods	HD/HD95	IoU	VC	VDP	HD/HD95	IoU	VC	VDP
SegResNet[15]	114.8/56.8	0.313	0.451	0.591	177.2/102.5	0.252	0.430	0.670
UNETR[10]	138.6/71.5	0.274	0.516	0.597	182.5/110.5	0.237	0.428	0.672
SwinUNETR[9]	124.8/57.1	0.304	0.585	0.605	185.5/108.1	0.261	0.451	0.488
nnUNet[12]	133.2/65.9	0.316	0.532	0.576	166.5/106.6	0.243	0.427	0.639
nnUNet++[31]	111.1/59.8	0.336	0.531	0.538	184.1/104.6	0.277	0.457	0.602
SegDiff[2]	96.0/47.9	0.343	0.575	0.586	99.6/53.8	0.312	0.542	0.632
MedSegDiff[26]	93.6/45.2	0.373	0.581	0.542	111.8/68.5	0.303	0.450	0.639
SDPM[11]	163.2/46.2	0.370	0.544	0.551	144.9/**42.0**	0.312	0.557	0.632
Our SADPM	**91.5**/**42.9**	**0.393**	**0.603**	**0.470**	**89.2**/42.3	**0.337**	**0.547**	**0.543**

3.2 Evaluation Metrics

Two types of metrics [20] are included based on spatial overlap and segmentation index. The spatial overlap metrics include Dice, Recall, Precision (Pre.) and the area under the ROC curve (AUC) at the pixel-wise level. The segmentation metrics include Hausdorff distance, Intersection-Over-Union (IoU), Volume Correlation (VC) based on the Pearson product-moment correlation coefficient, and Volume Difference Percentages (VDP), were used to quantitatively assess the model prediction performance compared to manual contouring.

3.3 Methods for Performance Comparison

Seven state-of-the-art segmentation methods were used as benchmarks for comparisons to verify the efficacy from our proposed method, denoted as SADPM. The SegResNet method [15] is based on the residual network for segmentation tasks and it is specially optimized for the medical imaging analysis. The strength of SegResNet is that the parametric volume of the model is small and can be trained fast. The UNETR [10] and SwinUNETR methods [9] are based on transformer architectures. The nnuNet related methods are the basic U-shaped networks with deep supervision implemented in [12]. The nnuNet++ method [31]

is a nested nnuNet architecture for learning the features for the convolutional kernels. SegDiff [2] and MedSegDiff [26] are similar with our segmentation model based on DDPM, but was implemented in totally different ways where the images are used as conditions during reverse process. SDPM is implemented with the same segmentation framework as SADPM, but with isotropic noise. In our experiments, the nnuNet++, SegResNet, UNETR and SwinUNETR methods were implemented by the framework of MONAI (monai.io). SegDiff and MedSegDiff were implemented by the released code from public. For fair comparisons, all the models in our experiments were adjusted with the optimal configurations.

4 Results

Visual illustrations on lesion segmentation are shown in Fig.2, which illustrates that SADPM is able to pick up lesion segmentations accurately, and had superior performance when compared with the other methods. Quantitative results in Table.1, 2 show that the proposed SADPM obtained the best performance overall with respect to the metrics in 3.2 outperforming the models of nnuNet, nnuNet++, SegRegNet, UNETR, SwinUNETR, SegDiff and MedSegDiff.

Our SADPM obtained the best performance with the key metrics of a mean dice of 0.524 and 0.453, AUC of 0.763 and 0.725, IoU of 0.393 and 0.337, VC of 0.603 and 0.547 and VDP of 0.470 and 0.543 in AISD and Private Infarct datasets, which was better than the other methods for the infarct of stroke lesion segmentation task.

5 Discussion and Conclusion

The experiments show that the advantages of the proposed SADPM by introducing an fully probabilistic framework of latent variant model with anisotropic noises rather than the isotropic noises and some efficient inference mechanism, making the results more explainable and robust to the segmentation tasks.

It is worth emphasizing that SADPM outperforms not only traditional segmentation methods but also transformer-based approaches and other diffusion probabilistic models (DDPM) based methods. The mean Dice coefficient, a pivotal measure of spatial agreement, demonstrates the superior ability of SADPM to accurately delineate stroke lesions in non-contrast CT images. The AUC, IoU, VC, and VDP further affirm the robustness and effectiveness of SADPM, surpassing the performance of competing models.

This study has several limitations. Firstly, the datasets employed in this research are limited. A larger pool of training samples could potentially enhance both the segmentation accuracy and its generalizability to diverse datasets. Secondly, it's important to note that the final label inference is subtly influenced by stochastic factors. Obtaining an average prediction based on multiple inferences is a time-consuming process, presenting a challenge for real-time applications. Thirdly, generating anisotropic noise within our model is also time-consuming.

Investigating more efficient methods for noise generation could significantly contribute to optimizing the overall computational efficiency. Lastly, exploring more advanced label fusion techniques may offer opportunities for further performance enhancement in segmentation tasks. Addressing these limitations in future iterations of the study could contribute to refining the proposed methodology and extending its applicability in medical image segmentation [29].

In conclusion, this study introduces a novel approach employing a synchronous image-label diffusion model facilitated by a latent variable model with anisotropic Gaussian noise for the segmentation of stroke lesions on non-contrast CT images. The proposed method exhibits a high level of efficacy through extensive experiments on two datasets. These findings not only underscore the effectiveness of the model but also hint at its potential applicability in the crucial task of stroke lesion volume measurement. The proposed methodology, with the demonstrated results, holds promise for contributing valuable insights and advancements in the field of medical image segmentation for stroke diagnosis and prognosis.

Acknowledgements This work was supported in part by the National Key Research and Development Program of China (2023YFC2410802), the Hubei Provincial Key Research and Development Program (2023BCB007), the High-Performance Computing platform of Huazhong University of Science and Technology and computer power at Wuhan Seekmore Intelligent Imaging Inc.

Disclosure of Interests The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Abbasi, H., Orouskhani, M., Asgari, S., Zadeh, S.S.: Automatic brain ischemic stroke segmentation with deep learning: A review. Neuroscience Informatics p. 100145 (2023)
2. Amit, T., Nachmani, E., Shaharbany, T., Wolf, L.: Segdiff: Image segmentation with diffusion probabilistic models. arXiv:2112.00390 (2021)
3. Calandriello, D., Lazaric, A., Valko, M.: Distributed adaptive sampling for kernel matrix approximation. In: Artificial Intelligence and Statistics. pp. 1421–1429. PMLR (2017)
4. Cavazza, J., Zunino, A., San Biagio, M., Murino, V.: Kernelized covariance for action recognition. In: 2016 23rd International Conference on Pattern Recognition (ICPR). pp. 408–413. IEEE (2016)
5. Choi, J., Lee, J., Shin, C., Kim, S., Kim, H., Yoon, S.: Perception prioritized training of diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11472–11481 (2022)
6. Croitoru, F.A., Hondru, V., Ionescu, R.T., Shah, M.: Diffusion models in vision: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence (2023)
7. Fidon, L., Aertsen, M., Kofler, F., Bink, A., David, A.L., Deprest, T., Emam, D., Guffens, F., Jakab, A., Kasprian, G., et al.: A dempster-shafer approach to trustworthy ai with application to fetal brain mri segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence (2024)

8. Frangakis, A.S., Hegerl, R.: Noise reduction in electron tomographic reconstructions using nonlinear anisotropic diffusion. Journal of structural biology **135**(3), 239–250 (2001)
9. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 7th International Workshop, BrainLes 2021, Held in Conjunction with MICCAI 2021, Virtual Event, September 27, 2021, Revised Selected Papers, Part I. pp. 272–284. Springer (2022)
10. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H.R., Xu, D.: Unetr: Transformers for 3d medical image segmentation. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 574–584 (2022)
11. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems **33**, 6840–6851 (2020)
12. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)
13. Jayasumana, S., Hartley, R., Salzmann, M., Li, H., Harandi, M.: Kernel methods on the riemannian manifold of symmetric positive definite matrices. In: proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 73–80 (2013)
14. Mazurowski, M.A., Dong, H., Gu, H., Yang, J., Konz, N., Zhang, Y.: Segment anything model for medical image analysis: an experimental study. Medical Image Analysis **89**, 102918 (2023)
15. Myronenko, A.: 3d mri brain tumor segmentation using autoencoder regularization. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part II 4. pp. 311–320. Springer (2019)
16. Najm, M., Kuang, H., Federico, A., Jogiat, U., Goyal, M., Hill, M.D., Demchuk, A., Menon, B.K., Qiu, W.: Automated brain extraction from head ct and cta images using convex optimization with shape propagation. Computer Methods and Programs in Biomedicine **176**, 1–8 (2019)
17. Nichol, A.Q., Dhariwal, P.: Improved denoising diffusion probabilistic models. In: International Conference on Machine Learning. pp. 8162–8171. PMLR (2021)
18. Nowinski, W.L.: Usefulness of brain atlases in neuroradiology: Current status and future potential. The Neuroradiology Journal **29**(4), 260–268 (2016)
19. Shawe-Taylor, J., Cristianini, N.: Kernel methods for pattern analysis. Cambridge university press (2004)
20. Taha, A.A., Hanbury, A.: Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. BMC medical imaging **15**(1), 1–28 (2015)
21. Tancik, M., Srinivasan, P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., Ramamoorthi, R., Barron, J., Ng, R.: Fourier features let networks learn high frequency functions in low dimensional domains. Advances in Neural Information Processing Systems **33**, 7537–7547 (2020)
22. Wang, H., Li, X.: Towards generic semi-supervised framework for volumetric medical image segmentation. Advances in Neural Information Processing Systems **36** (2024)

23. Wang, R., Lei, T., Cui, R., Zhang, B., Meng, H., Nandi, A.K.: Medical image segmentation using deep learning: A survey. IET Image Processing **16**(5), 1243–1267 (2022)
24. Whang, J., Delbracio, M., Talebi, H., Saharia, C., Dimakis, A.G., Milanfar, P.: Deblurring via stochastic refinement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16293–16303 (2022)
25. Williams, C.K., Rasmussen, C.E.: Gaussian processes for machine learning, vol. 2. MIT press Cambridge, MA (2006)
26. Wu, J., Fang, H., Zhang, Y., Yang, Y., Xu, Y.: Medsegdiff: Medical image segmentation with diffusion probabilistic model. arXiv:2211.00611 (2022)
27. Xing, Z., Ye, T., Yang, Y., Liu, G., Zhu, L.: Segmamba: Long-range sequential modeling mamba for 3d medical image segmentation. arXiv preprint arXiv:2401.13560 (2024)
28. Yang, R., Yang, Y., Zhou, F., Sun, Q.: Directional diffusion models for graph representation learning. Advances in Neural Information Processing Systems **36** (2024)
29. You, C., Dai, W., Min, Y., Liu, F., Clifton, D., Zhou, S.K., Staib, L., Duncan, J.: Rethinking semi-supervised medical image segmentation: A variance-reduction perspective. Advances in Neural Information Processing Systems **36** (2024)
30. Yue, Z., Wang, J., Loy, C.C.: Resshift: Efficient diffusion model for image super-resolution by residual shifting. In: Thirty-seventh Conference on Neural Information Processing Systems (2023)
31. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4. pp. 3–11. Springer (2018)