Accelerating Denoising Diffusion Probabilistic Model via Truncated Inverse Processes for Medical Image Segmentation

Xutao Guo[†a,b], Yang Xiang[†b], Yanwu Yang[a,b], Chenfei Ye[d], Yue Yu[*b] and Ting Ma[*a,b,c,d]

[a]School of Electronics and Information Engineering, Harbin Institute of Technology at Shenzhen, Shenzhen, China.
[b]The Peng Cheng Laboratory, Shenzhen, Guangdong, China.
[c]Guangdong Provincial Key Laboratory of Aerospace Communication and Networking Technology, Harbin Institute of Technology at Shenzhen, China.
[d]The International Research Institute for Artifcial Intelligence, Harbin Institute of Technology at Shenzhen, Shenzhen, China.

ARTICLE INFO

Keywords:
Denoising diffusion probabilistic models
accelerating
medical image segmentation
uncertainty

ABSTRACT

Due to the impressive advancements achieved by Denoising Diffusion Probability Models (DDPMs) in image generation, researchers have explored the possibility of utilizing DDPMs in discriminative tasks to achieve superior performance. However, the inference process of DDPMs is highly inefficient since it requires thousands of iterative denoising steps. In this study, we propose an accelerated denoising diffusion probabilistic model via truncated inverse processes (ADDPM) that is specifically designed for medical image segmentation. The inverse process of ADDPM starts from a non-Gaussian distribution and terminates early once a prediction with relatively low noise is obtained after multiple iterations of denoising. We employ a separate powerful segmentation network to obtain pre-segmentation and construct the non-Gaussian distribution of the segmentation based on the forward diffusion rule. By further adopting a separate denoising network, the final segmentation can be obtained with just one denoising step from the predictions with low noise. ADDPM greatly reduces the number of denoising steps to approximately one-tenth of that in vanilla DDPMs. Our experiments on three segmentation tasks demonstrate that ADDPM outperforms both vanilla DDPMs and existing representative accelerating DDPMs methods. Moreover, ADDPM can be easily integrated with existing advanced segmentation models to improve segmentation performance and provide uncertainty estimation.

1. Introduction

Denoising Diffusion Probabilistic Models (DDPMs) (Sohl-Dickstein et al., 2015; Ho et al., 2020) have become a popular research topic in computer vision due to their impressive performance in both unconditional and conditional generation tasks (Song et al.; Wolleb et al., 2022). DDPMs can be trained on ground truth and use images as priors during sampling to generate image-specific segmentations, as illustrated in Figure 1 (Wolleb et al., 2022; Guo et al., 2022b). In medical image segmentation, annotations are often subject to variability among annotators due to differences in expertise and inherent ambiguity of medical images (Kohl et al., 2018; Guo et al., 2022a). Algorithms that only provide the most likely hypotheses can lead to misdiagnosis and suboptimal treatment (Begoli et al., 2019), especially when image segmentation is crucial for the subsequent diagnosis or treatment. To address this issue, medical images are often annotated by multiple experts to reduce subjective biases(Liao et al., 2022). DDPMs offer a solution to this problem as their inference process is stochastic and can generate several segmentation masks for the same input image. This enabes the computation of pixel-wise uncertainty maps and an ensemble of segmentations, which can improve segmentation performance (Wolleb et al., 2022; Guo et al., 2022b). This feature is particularly beneficial in clinical settings as it allows for the interpretation of multiple possible semantic segmentation hypotheses, providing potential diagnoses and suggesting further actions to resolve current ambiguity.

[*]This study is supported by grants from the National Natural Science Foundation of China (62106113, 62276081), the Innovation Team and Talents Cultivation Program of National Administration of Traditional Chinese Medicine (NO:ZYYCXTD-C-202004), Basic Research Foundation of Shenzhen Science and Technology Stable Support Program (GXWD20201230155427003-20200822115709001), Shenzhen Longgang District Science and Technology Development Fund Project (LGKCXGZX2020002), and The Major Key Project of PCL (PCL2021A06). Thanks for the support provided by the OpenI Community: https://git.openi.org.cn.

[*]Xutao Guo and Yang Xiang are co-first authors
[*]Corresponding author: Yue Yu, Ting Ma.
ORCID(s):

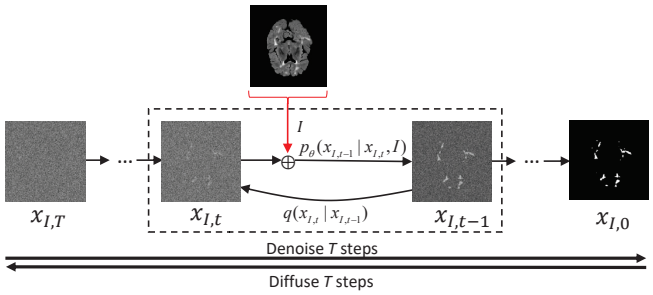

Figure 1: The diffusion and inverse processes of the DDPMs for medical image segmentation. In each step t of the inverse denoising process, the conditional information is induced by concatenating the medical images I with the noisy segmentation mask $x_{I,t}$.

The diffusion process and the inverse process of DDPMs correspond to two distinct Markov chains. The diffusion process involves gradually introducing Gaussian noise to data samples until the data distribution becomes Gaussian, while the generative process is the inverse of the diffusion process (Sohl-Dickstein et al., 2015; Ho et al., 2020). DDPMs generate samples by iteratively removing noise from Gaussian noise samples using a trained deep neural network for denoising. In medical image segmentation, the generative objective of DDPMs is to predict segmentations (Wolleb et al., 2022; Guo et al., 2022b). However, a significant challenge with DDPMs is the inefficiency of inference. Obtaining segmentations from DDPMs typically requires thousands of denoising steps, and each step of which involves a forward prediction of the denoising neural network.

Several approaches have been proposed to accelerate the inverse process of DDPMs, including using non-Markovian inverse process (Begoli et al., 2019), introducing knowledge distillation (Salimans and Ho), diffusing in a lower-dimensional latent space (Rombach et al., 2022), and using adaptive noise scheduling (Kingma et al., 2021). However, these methods have limitations and cannot significantly accelerate sampling without compromising the quality of generation. Other methods (Zheng et al., 2022; Lyu et al., 2022) improve sampling efficiency by truncating the diffusion processes, boosting the performance at the same time. However, these methods require the combination of GAN(Goodfellow et al., 2020) or VAE(Kingma and Welling, 2014) models, which are difficult to train, and their performance is limited by the quality of the generated images from the pre-trained generative model. In summary, none of the above methods are specifically designed to accelerate sampling for segmentation tasks. Our preliminary work, PD-DDPM, was accepted in the 20th IEEE International Symposium on Biomedical Imaging (ISBI2023) conference paper (Guo et al., 2022b), and it is the first accelerated DDPMs model developed specifically for medical image segmentation. The key idea behind PD-DDPM is to obtain pre-segmentation results using a separate segmentation network and construct noise predictions based on the forward diffusion rule. By starting from these noisy predictions, clean segmentation results can be generated with fewer inverse denoising steps. PD-DDPM only considers truncating some of the initial iterative steps in the inverse process. ADDPM is an extension of PD-DDPM proposed in this paper that truncate both the initial and final stages of the inverse process. The number of denoising steps in ADDPM is approximately one-third of that in PD-DDPM.

In this paper, we propose an accelerating denoising diffusion probabilistic model via truncated inverse processes (ADDPM) that is specifically designed for medical image segmentation. The core idea behind ADDPM is to truncate both the initial and final stages of the inverse process, resulting in a more efficient inference process that only considers a smaller number of steps in the middle. To achieve this, we first obtain pre-segmentation results using a separate segmentation network and then generate noise predictions (non-Gaussian distribution) based on the forward diffusion rule. Next, starting from these noisy predictions, we can use fewer inverse denoising steps to generate clean segmentation results. Further, we terminate the inverse processes early when a low noisy segmentation result is obtained, and then apply a separate denoising network to denoise the low noisy segmentation and obtain the final segmentation result. Our experiments demonstrate that ADDPM significantly reduces the number of denoising steps required (from 1000 steps to 100 steps in the best case of our experiments), without sacrificing segmentation performance. When integrated with a strong pre-segmentation model, ADDPM outperforms both vanilla DDPMs and the pre-segmentation model alone. Moreover, ADDPM can be easily integrated with existing advanced segmentation models to improve segmentation performance and provide uncertainty estimation. We evaluate ADDPM across different datasets and demonstrate that

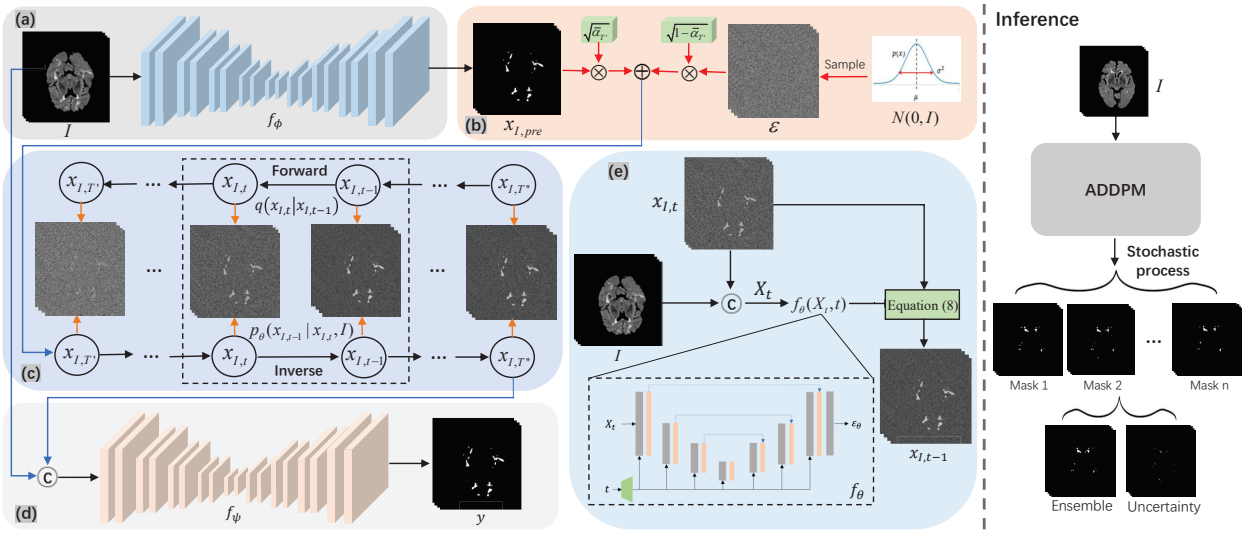

Figure 2: Overview of ADDPM for medical image segmentation. (a) The image I is input into the segmentation network f_ϕ to obtain the pre-segmentation result $x_{I,pre}$. (b) Based on equation (10), the noise segmentation prediction $x_{I,T'}$ for step T' can be obtained by performing a single diffusion operation on $x_{I,pre}$. (c) Starting from $x_{I,T'}$, iterative denoising is performed to obtain the noise prediction $x_{I,T''}$. (d) The denoising network f_ψ is used to denoise the noise prediction $x_{I,T''}$ to obtain the final segmentation result y. (e) In each step of the inverse denoising process, $x_{I,t}$ is obtained based on $x_{I,t-1}$ according to equation (8). During inference, ADDPM samples the same test sample n times to generate n different segmentation masks. The ensemble segmentation and uncertainty map are generated based on the n different predictions.

it achieves optimal performance compared to existing representative methods. Since both the pre-segmentation and denoising networks can be easily implemented in a single step, this imposes only a small computational overhead on ADDPM.

In summary, we have made the following three contributions:

1. We propose ADDPM, an accelerated denoising diffusion probabilistic model via truncated inverse processes, specifically designed for medical image segmentation. ADDPM significantly reduces the number of denoising steps required for generate segmentation results while improving segmentation performance.
2. ADDPM retains the key properties of vanilla DDPMs, such as uncertainty estimation and ensemble, and can be easily integrated with existing advanced segmentation models to further enhance segmentation performance and provide uncertainty estimation capabilities.
3. Our experiments on three different datasets demonstrate that ADDPM achieves optimal performance compared to existing representative methods.

This paper is organized as follows. Section 2 describes the related work. Section 3 presents in detail each component of our method. Section 4 introduces the experimental setup and experimental data in detail. Section 5 evaluates the proposed methods and reports results. Sections 6 conclude the work of this paper.

2. Related Work

2.1. Medical Image Segmentation

CNN for Segmentation: In recent years, convolutional neural networks have made great development in medical image segmentation (Ronneberger et al., 2015; Song et al., 2022; Zhou et al., 2019). With superior performance and elegant structure, U-Net has become a common basic model in medical image segmentation (Ronneberger et al., 2015). And many later works also extended this architecture to achieve more accurate segmentation, such as ResUNet (Xiao et al., 2018), DenseUNet (Cao et al., 2020), AttUNet (Oktay et al.) and UNet++ (Zhou et al., 2019). Inspired by residual connections and dense connections in computer vision tasks, ResUNet (Xiao et al., 2018) and DenseUNet (Cao et al., 2020) replace the encoder backbone of UNet with residual connections and dense connections, respectively.

Some other technologies such as attention mechanism (Vaswani et al., 2017), atrous convolution (Chen et al., 2017), pyramid structure (Zhao et al., 2017), etc. have also been successfully introduced to improve the performance of the UNet. U-Net++ (Zhou et al., 2019) proposed nested and dense skip connections which can reduce the semantic gap between the encoder and the decoder.

Transformer for Segmentation: Inspired by the great success of Transformer in NLP (Vaswani et al., 2017), more and more Transformer-based methods appear in CV task (Dalmaz et al., 2022; He et al., 2021). Dosovitskiy et al. (Dosovitskiy et al.) introduced transformers to vision tasks for the first time, demonstrating state-of-the-art performance on image classification tasks. Recently researchers have also introduced transformers for medical image segmentation. TransUNet (Chen et al., 2021) introduces transformers into encoders to capture global dependencies, and is the first to study using Transformers to solve medical image segmentation problems. UNETR (Hatamizadeh et al., 2022) utilizes pure Transformers as encoders to efficiently capture global multi-scale information. Swin-Unet (Cao et al., 2023) uses a layered Swin transformer with a shift window as an encoder to extract contextual features.

2.2. Denoising Diffusion Probabilistic Model

DDPMs are a class of generative models that has received increasing attention due to their remarkable achievements in unconditional and conditional generative tasks (Sohl-Dickstein et al., 2015; Ho et al., 2020). To date, it has been widely used in a variety of applications, ranging from generative tasks such as image generation (Esser et al., 2021), image super-resolution (Li et al., 2022), and image inpainting (Liu et al., 2022) to discriminative tasks such as image segmentation (Wolleb et al., 2022; Guo et al., 2022b), anomaly detection (Wolleb et al., 2022). Recently, DDPM-based researchs has also emerged in the medical image segmentation. Based on the DDPMs, medical image segmentation can be described as a conditional image generation task, which allows to compute pixel-wise uncertainty maps of the segmentation and allows an ensemble of segmentations to boost the segmentation performance (Wolleb et al., 2022; Guo et al., 2022b). Wolleb et al. (Wolleb et al., 2022) proposed a weakly supervised learning method based on DDIM for medical anomaly detection. Hu et al. (Hu et al., 2022) utilizes DDPM to denoise optical coherence tomography (OCT) retinal data in an unsupervised manner.

2.3. Accelerating DDPMs

Recently there has been some works focused on speeding up the sampling process for DDPMs. Song et al. (Song et al.) attempted to reduce the number of diffusion steps by using non-Markovian inverse processes. Watson et al. (Watson et al., 2022) propose a dynamic programming algorithm that finds the optimal denoising time-step schedule for DDPMs. San-Roman et al. (San-Roman et al., 2021) present a adaptive noise scheduling to estimate the noise parameters given the current input at inference time, which requiring less steps. Salimans & Ho (Salimans and Ho) propose to progressively distill a trained DDPM for fast sampling. Some methods (Vahdat et al., 2021; Rombach et al., 2022) shifted the diffusion process to the latent space using pre-trained autoencoders. However, the above methods cannot achieve significant acceleration without sacrificing the quality of generation. Some other methods (Zheng et al., 2022; Lyu et al., 2022) improve sampling efficiency by truncating the forward diffusion processes, boosting the performance at the same time. But those methods need to combine GAN (Goodfellow et al., 2020) or VAE (Kingma and Welling, 2014) models that are difficult to train. And all of the above methods do not implement accelerated sampling specifically for segmentation tasks. PD-DDPM (Guo et al., 2022b) is our previously proposed accelerated DDPM model specifically for medical image segmentation, which only considers truncating some of the initial iterative steps in the inverse process.

3. Method

Figure 2 illustrates the overview of our proposed ADDPM. Firstly, a separate segmentation network f_ϕ is trained to obtain the pre-segmentation x_{pre} from the image I. According to the forward diffusion rule, the pre-segmentation result x_{pre} is diffused to the T'-th step to obtain $x_{T'}$. Then, based on $x_{T'}$ and following the rules of the inverse diffusion process, we iteratively denoise for $T' - T''$ steps to obtain the segmentation prediction $x_{T''}$ at step T''. Finally, a separate denoising network f_ψ is used to obtain the clean segmentation prediction y from $x_{T''}$ in one step. The number of iteration steps of the entire inverse denoising process is reduced from T steps of vanilla DDPMs to $T' - T''$ steps of ADDPM.

3.1. Preliminaries on DDPMs

In DDPMs, the forward diffusion process is a first-order Markov chain that perturbs the data distribution $q(x_0)$ by gradually adding Gaussian noise with variance $\beta_t \in (0, 1)$ at time t, until the data distribution converges to a standard Gaussian distribution. The form of the forward process can be summarized as follows:

$$\begin{aligned} q(x_{1:T}|x_0) &= \prod_{t=1}^{T} q(x_t|x_{t-1}), \\ q(x_t|x_{t-1}) &= \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t\mathbf{I}). \end{aligned} \tag{1}$$

where $x_{1:T}$ denotes the set of variables $x_1, x_2, ..., x_T$. The value of T typically ranges from 1000 to 4000 in most works. By sequentially applying $q(x_t|x_{t-1})$ for t steps, we can write:

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\overline{\alpha}_t}x_0, (1-\overline{\alpha}_t)\mathbf{I}), \tag{2}$$

with $\alpha_t = 1 - \beta_t$ and $\overline{\alpha}_t = \prod_{s=1}^{t} \alpha_s$. Using the reparameterization trick, we can express x_t directly as a function of x_0:

$$x_t = \sqrt{\overline{\alpha}_t}x_0 + \sqrt{1-\overline{\alpha}_t}\varepsilon, \tag{3}$$

with $\varepsilon \in \mathcal{N}(0, \mathbf{I})$. Since the diffusion rate is small (i.e., β_t is kept sufficiently small), the inverse process distribution $p_\theta(x_{t-1}|x_t)$ also follows a Gaussian distribution. Therefore, we can use a neural network f_θ to parameterize the Gaussian distribution p_θ to approximate the inverse process. Starting from $p(x_T) = \mathcal{N}(x_T; 0, I)$, the inverse process can be expressed as follows:

$$\begin{aligned} p_\theta(x_{0:(T-1)}|x_T) &= \prod_{t=1}^{T} p_\theta(x_{t-1}|x_t), \\ p_\theta(x_{t-1}|x_t) &= \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)). \end{aligned} \tag{4}$$

To generate an image from the inverse process, we first sample x_T from the underlying data distribution by drawing a latent variable (of the same size as the training data point x_0) from $p(x_T)$, which is chosen to be an isotropic Gaussian distribution. Then, we iteratively draw sample x_{t-1} from $p_\theta(x_{t-1}|x_t)$ for $t = T, T-1, \ldots, 1$ until we obtain a new data point x_0. The generation process of DDPMs is computationally expensive since it requires iterative sampling from the transition distribution $p_\theta(x_{t-1}|x_t)$, which involves many evaluations of the output of f_θ. In (Ho et al., 2020), it was shown that during inference, starting from Gaussian noise x_t, x_{t-1} can be obtained by iteratively denoising x_t as follows:

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{1-\alpha_t}{\sqrt{1-\overline{\alpha}_t}}f_\theta(x_t, t)) + \overline{\alpha}_t z, \tag{5}$$

with $z \sim \mathcal{N}(0, \mathbf{I})$. Following the standard process of DDPMs (Ho et al., 2020), the training objective can be further simplified as:

$$L_{simple} = E_{t,x_0,\varepsilon}[\| \varepsilon - f_\theta(\sqrt{\overline{\alpha}_t}x_0 + \sqrt{1-\overline{\alpha}_t}\varepsilon, t) \|^2]. \tag{6}$$

3.2. DDPMs for Medical Image Segmentation

Figure 1 depicts the modification of DDPMs for medical image segmentation. Let I be the input medical image, and x_I be its corresponding ground truth segmentation with the same dimensions as I. We incorporate the anatomical information of I by appending it as an image prior to x_I, resulting in $X = I \oplus x_I$. During the forward diffusion process q, we only add noise to the ground truth segmentation x_I as follows:

$$x_{I,t} = \sqrt{\overline{\alpha}_t}x_I + \sqrt{1-\overline{\alpha}_t}\epsilon, \tag{7}$$

and we define $X_t = I \oplus x_{I,t}$. Equation 5 is then modified to:

$$x_{I,t-1} = \frac{1}{\sqrt{\alpha_t}}(x_{I,t} - \frac{1-\alpha_t}{\sqrt{1-\overline{\alpha}_t}}f_\theta(X_t, t)) + \overline{\alpha}_t z. \tag{8}$$

3.3. Truncating The Initial Stages of The Inverse Process

One natural question raised from the vanilla DDPMs could be: can we truncate the inverse process to T' ($< T$) steps? If we can obtain the noise segmentation prediction $x_{I,T'}$ in advance, then it can be achieved. As illustrated in Figure 1, to generate an image-specific segmentation, we train the DDPMs on the ground truth segmentation and use the image as a prior during training and in each step of the sampling process (Wolleb et al., 2022; Guo et al., 2022b). Additionally, we can use image priors to train a separate segmentation network f_ϕ to obtain pre-segmentation results $x_{I,pre}$, as shown in Figure 2.

$$x_{I,pre} = f(I;\phi). \tag{9}$$

In the sampling process, we first obtain the pre-segmentation result $x_{I,pre}$ through the pre-segmentation network f_ϕ. Then, according to equation (7), the pre-segmentation result is diffused to T' step to obtain the noise segmentation prediction $x_{I,T'}$.

$$x_{I,T'} = \sqrt{\overline{\alpha}_{T'}} x_{I,pre} + \sqrt{1-\overline{\alpha}_{T'}}\varepsilon. \tag{10}$$

By following the rules of the inverse diffusion process, the sampling process can begin iterative denoising based on $x_{I,T'}$. In this way, the inverse process can be newly defined as:

$$\begin{aligned} p_\theta(x_{I,0:(T'-1)}|x_{I,T'}) &= \prod_{t=1}^{T'} p_\theta(x_{I,t-1}|x_{I,t}), \\ p_\theta(x_{I,t-1}|x_{I,t}) &= \mathcal{N}(x_{I,t-1};\mu_\theta(x_{I,t},t),\Sigma_\theta(x_{I,t},t)). \end{aligned} \tag{11}$$

This is also equivalent to truncating the diffusion process in the middle at $t = T' < T$. So we can use the DDPMs to denoise the non Gaussian distribution $x_{I,T'}$ to a clean segmentation x_0 in fewer steps than the vanilla sampling process according to equation (11). In this way, the number of T step iterations of the vanilla DDPMs is reduced by $T - T'$ steps. Existing advanced segmentation models can all be used as pre-segmentation networks, so this design has good scalability. At the same time, since the pre-segmentation can be completed in one step, it brings only minor computational overhead to ADDPM. Through subsequent experiments, we empirically found that truncating the initial stages of the inverse process can enhance the model's segmentation performance compared to vanilla DDPMs.

3.4. Truncating The Final Stages of The Inverse Process

Based on equation (11), the inverse process starts from step T' to iteratively denoise. With a similar idea, we can further stop the inverse process early, thereby reducing the number of inference iterations even more. As shown in Figure 2, the prediction results in the final stages of the inverse process demonstrate minimal noise. While the segmentation prediction in the final stages of the inverse process may exhibit some noise, it tends to align with the final segmentation prediction result and does not introduce excessive prediction randomness. Therefore, we can early stop the inverse process based on truncating the initial stages of the inverse process. The step at which the inverse process stops early is represented by $T''(T'' < T')$. Based on $x_{I,T''}$, we can denoise it in one step through a separate denoising network f_ψ to obtain the final segmentation result y. In detail, the inverse diffusion process from step $x_{I,T'}$ to step $x_{I,T''}$ can be described as follows:

$$\begin{aligned} p_\theta(x_{I,(T_{\prime\prime}-1):(T'-1)}|x_{I,T'}) &= \prod_{t=T''-1}^{T'} p_\theta(x_{I,t-1}|x_{I,t}), \\ p_\theta(x_{I,t-1}|x_{I,t}) &= \mathcal{N}(x_{I,t-1};\mu_\theta(x_{I,t},t),\Sigma_\theta(x_{I,t},t)). \end{aligned} \tag{12}$$

Based on equation (12), we can iteratively denoise to obtain the noise segmentation prediction $x_{I,T''}$ from $x_{I,T'}$. Then, we concatenate the noise prediction $x_{I,T''}$ and the image I as input to the denoising network f_ψ. The output of f_ψ is the clean segmentation result y.

$$y = f(I, x_{I,T''};\psi) \tag{13}$$

In this paper, we choose the UNet as the denoising network f_ψ. There are several different ways to get the final segmentation result y from $x_{I,T''}$. For example, we can use the denoising network f_ψ to directly denoise $x_{I,T''}$ to

Algorithm 1 Sampling Procedure

Input: I, the original image
Output: y, the segmentation

$x_{I,pre} = f_{\phi}(I)$
$x_{I,T'} = \sqrt{\overline{\alpha}_{T'}} x_{I,pre} + \sqrt{1-\overline{\alpha}_{T'}}\epsilon$
for $t \leftarrow T'$ to T'' **do**
    $X_t = I \oplus x_{I,t}$
    $x_{I,t-1} = \frac{1}{\sqrt{\alpha_t}}(x_{I,t} - \frac{1-\alpha_t}{\sqrt{1-\overline{\alpha}_t}} f_{\theta}(X_t, t)) + \overline{\alpha}_t z$
end for
$X_{T''} = I \oplus x_{I,T''}$
$y = f_{\psi}(X_{T''})$
return y

obtain the segmentation result y. In the following experiments, we will compare these methods in detail. Since the denoising can be completed in one step, the computational overhead brought to ADDPM is very small. In the above way, the number of denoising steps required for the sampling process is greatly reduced.

3.5. Training And Testing

Based on the above description, ADDPM requires training three networks. The training objectives of the three networks are detailed as follows:

Training f_{ϕ}: For f_{ϕ}, the input is the image I, and the output is the pre-segmentation result $x_{I,pre}$. Correspondingly, we use the standard cross-entropy loss function for optimization.

$$L(P, f(I;\phi)) = -\sum_{i} P_i log f_i(I;\phi), \tag{14}$$

where $f_i(I;\phi)$ can be seen as the likelihood function of the real category i, and P is the ground truth probability distribution.

Training f_{ψ}: For f_{ψ}, the input is the noise prediction $x_{I,T''}$ concatenated with the image I, and the output of the denoising network f_{ψ} is the final clean segmentation result y. During training, $x_{I,T''}$ is obtained by equation (7) based on x_I.

$$x_{I,T''} = \sqrt{\overline{\alpha}_{T''}} x_I + \sqrt{1-\overline{\alpha}_{T''}}\epsilon, \tag{15}$$

Where x_I is the ground truth segmentation annotation and y is the predicted segmentation result from f_{ψ}. The cross-entropy loss function is used for optimization.

$$L(P, f(x_{I,T''}, I;\psi)) = -\sum_{i} P_i log f_i(x_{I,T''}, I;\psi), \tag{16}$$

where $f_i(I;\psi)$ can be seen as the likelihood function of the real category i, and P is the ground truth probability distribution.

Training f_{θ}: The parameters for f_{θ} are obtained by minimizing the KL-divergence between the forward and inverse distributions for all timesteps. This can be further simplified by using a posterior distribution $q(x_{I,t}|x_{I,t-1}; x_{I,0})$ (Sohl-Dickstein et al., 2015). And the posterior distribution can be derived using equation (1) and (3) under the Markovian assumption,

$$q(x_{I,t-1}|x_{I,t}, x_{I,0}) = \mathcal{N}(x_{I,t-1}, \mu, \sigma^2\mathbf{I}), \tag{17}$$

where, $\mu(x_t, x_0) = \frac{\sqrt{\overline{\alpha}_{t-1}}(1-\alpha_t)}{1-\overline{\alpha}_t} x_0 + \frac{\sqrt{\alpha_t}(1-\overline{\alpha}_{t-1})}{1-\overline{\alpha}_t} x_t$ and $\sigma^2 = \frac{(1-\overline{\alpha}_{t-1})(1-\alpha_t)}{1-\overline{\alpha}_t}$. During optimization, we assume that both distributions $q(x_{I,t-1}|x_{I,t}, x_{I,0})$ and $p_{\theta}(x_{I,t-1}|x_{I,t})$ are the same. The combination of p and q can be seen as a variational autoencoder (Kingma and Welling, 2014), and the variational lower bound (VLB) can be expressed as follows:

$$L_{vlb} = L_{T''} + L_{T''+1} + ... + L_{T'-1} + L_{T'} \tag{18}$$

$$L_{t-1} = D_{KL}(q(x_{I,t-1}|x_{I,t}, x_{I,0}) \parallel p_\theta(x_{I,t-1}|x_{I,t})) \tag{19}$$

However, the training objective can be further simplified as equation (6) (Ho et al., 2020). In addition, Log-likelihood is a widely used metric in generative modeling, and it is generally believed that optimizing log-likelihood forces generative models to capture all of the modes of the data distribution (Nichol and Dhariwal, 2021). Hence, we obtain the hybrid loss (Nichol and Dhariwal, 2021) by combining equation (6) and (18):

$$L_{hybrid} = L_{simple} + \lambda L_{vlb} \tag{20}$$

Where λ is a regularization parameter. For all experiments in this paper, $\lambda = 0.001$ is set to regularize L_{vlb}.

Testing: During inference, we follow the procedure presented in Algorithm 1, which is a stochastic process. Therefore, sampling twice for the same image I does not result in the same segmentation mask prediction y. Exploiting this property, we can implicitly generate an ensemble of segmentation masks without having to train a new model. This ensemble can be used to improve the segmentation performance.

4. EXPERIMENTS

4.1. Datasets

WMH: The dataset was provided by the WMH segmentation challenge in MICCAI 2017(Kuijf et al., 2019). It consists of 60 cases of brain MRI images, including 3D T1-weighted images and 2D multi-slice FLAIR images, along with manual annotations of white matter hyperintensity in binary masks.

Hip: The dataset is provided by the MRI Hippocampus Segmentation challenge in Kaggle (Malekzadeh, 2019). It contains MRI segmentation of Hippocampus gland (binary masks), which consists of 135 cases MRI images.

Brats13: The dataset is provided by BraTS2013 (Menze et al., 2014), which includeing 20 High-grade and 10 Low-grade with Flair, T1, T1c, and T2 scans of MRI. The ground truth have four different labels: 1-necrosis, 2-edema, 3-non-enhancing tumor, and 4-enhancing tumor. Following the BraTS2013, three different categories such as complete tumor (1-necrosis, 2-edema, 3-non-enhancing tumor, 4-enhance tumor), tumor core (3-non-enhance tumor, 4-enhance tumor) and enhanced tumor (4-enhacne tumor) are considered for the evaluation.

4.2. Implementation Details

Our experiments were implemented using Pytorch and performed on NVIDIA TESLA V-100 (Pascal) GPUs with 32 GB memory. We used the Adam optimizer with a weight decay of 1e-5 to optimize all configurations. For WMH, we set the batch size to 12 and patch size to 128×192. For Hip, we set the batch size to 8 and patch size to 224×192. For Brats13, we set the batch size to 12 and patch size to 160×160. During testing, the segmentation probability maps were predicted by the sliding windows technique with 50% overlaps. The denoising network used in DDPMs had a architecture similar to the UNet used in (Nichol and Dhariwal, 2021), and its attention layer used a multi-head attention (Vaswani et al., 2017). We employed a cosine noise schedule for $T = 1000$ steps for DDPMs. All cases for each task were randomly assigned to a training set (3-fold), a validation set (1-fold), and a test set (1-fold). And we chose the best-performing model on the validation set for testing.

5. RESULTS

5.1. Comparison of Segmentation Performance

The inference process of ADDPM is a stochastic process, meaning that multiple samples of the same image may yield different segmentation mask predictions. In this paper, we sample 5 different segmentation masks for each image in the test set. Then, we average the predicted probability maps (the softmax output of f_ψ) corresponding to the five different segmentation masks to obtain the fused predicted probability map. Finally, we transform the fused prediction probability map into segmentation results. Table 1 presents the Dice score, HD score, Jaccard index, and Precision.

In Table 1, we present the results of quantitative experiments that compare our method with a range of representative methods. The UNet (Ronneberger et al., 2015), AttUnet (Oktay et al.), and UNet++ (Zhou et al., 2019) are the most representative convolutional structured deep learning models for medical image segmentation. The UNETR (Hatamizadeh et al., 2022) and SwinUNet (Cao et al., 2023) are the most representative deep learning models for

Table 1
Comparison of segmentation results of existing advanced models and our proposed method. The performance is measured by Dice, HD, Jaccard, and, Precision.

	Dice ↑ : (mean±std)(%)					Jaccard ↑ : (mean±std)(%)				
	WMH	Hip	Brats13			WMH	Hip	Brats13		
Methods			WT	TC	ET			WT	TC	ET
UNet	$78.71_{\pm 08.13}$	$82.47_{\pm 11.90}$	$83.02_{\pm 13.42}$	$66.04_{\pm 18.92}$	$59.69_{\pm 30.27}$	$65.65_{\pm 11.18}$	$71.56_{\pm 13.72}$	$72.79_{\pm 16.03}$	$52.05_{\pm 19.61}$	$48.26_{\pm 26.81}$
AttUNet	$79.97_{\pm 07.76}$	$82.96_{\pm 11.98}$	$82.18_{\pm 15.08}$	$66.09_{\pm 18.47}$	$59.12_{\pm 29.53}$	$67.34_{\pm 11.01}$	$72.31_{\pm 13.96}$	$71.93_{\pm 17.15}$	$52.09_{\pm 20.00}$	$47.35_{\pm 26.03}$
UNet++	$79.87_{\pm 07.84}$	$82.73_{\pm 12.06}$	$83.19_{\pm 15.72}$	$62.24_{\pm 19.31}$	$54.44_{\pm 32.88}$	$67.19_{\pm 10.93}$	$71.98_{\pm 14.01}$	$73.63_{\pm 18.15}$	$48.00_{\pm 20.22}$	$33.07_{\pm 28.90}$
UNETR	$78.10_{\pm 08.11}$	$81.24_{\pm 12.07}$	$77.20_{\pm 21.30}$	$61.67_{\pm 22.59}$	$52.95_{\pm 30.22}$	$64.82_{\pm 11.13}$	$69.82_{\pm 13.93}$	$66.70_{\pm 22.44}$	$48.33_{\pm 23.20}$	$41.57_{\pm 27.37}$
SwinUNet	$77.03_{\pm 09.53}$	$80.95_{\pm 12.16}$	$76.82_{\pm 21.87}$	$57.70_{\pm 24.20}$	$51.36_{\pm 30.48}$	$63.62_{\pm 12.65}$	$69.41_{\pm 13.92}$	$66.38_{\pm 22.93}$	$44.69_{\pm 24.60}$	$40.25_{\pm 28.07}$
Bayesian UNet	$79.81_{\pm 07.91}$	$82.58_{\pm 12.21}$	$80.53_{\pm 16.95}$	$63.26_{\pm 20.25}$	$57.48_{\pm 30.94}$	$67.16_{\pm 06.57}$	$71.80_{\pm 14.15}$	$70.18_{\pm 19.73}$	$49.61_{\pm 22.75}$	$46.31_{\pm 27.87}$
Probabilistic UNet	$79.22_{\pm 08.02}$	$82.22_{\pm 12.22}$	$81.50_{\pm 19.35}$	$62.10_{\pm 22.74}$	$56.63_{\pm 30.35}$	$66.32_{\pm 11.14}$	$71.27_{\pm 13.81}$	$72.07_{\pm 20.33}$	$48.82_{\pm 23.23}$	$45.26_{\pm 27.59}$
DDPM	$79.62_{\pm 08.35}$	$82.08_{\pm 12.02}$	$81.91_{\pm 20.60}$	$64.47_{\pm 23.09}$	$57.64_{\pm 31.17}$	$66.96_{\pm 11.79}$	$71.01_{\pm 13.77}$	$73.02_{\pm 21.23}$	$51.65_{\pm 24.17}$	$46.74_{\pm 28.91}$
TDPM	$80.13_{\pm 08.08}$	$82.87_{\pm 12.10}$	$82.28_{\pm 19.42}$	$66.38_{\pm 20.48}$	$58.49_{\pm 30.58}$	$67.62_{\pm 11.44}$	$72.19_{\pm 13.98}$	$73.22_{\pm 20.36}$	$52.98_{\pm 21.76}$	$47.34_{\pm 28.18}$
ES-DDPM	$80.31_{\pm 07.66}$	$82.90_{\pm 12.34}$	$82.40_{\pm 19.02}$	$64.92_{\pm 22.63}$	$59.32_{\pm 30.65}$	$67.80_{\pm 10.94}$	$72.30_{\pm 14.25}$	$73.29_{\pm 20.14}$	$52.02_{\pm 23.84}$	$48.30_{\pm 28.54}$
PD-DDPM	$81.22_{\pm 07.29}$	$83.60_{\pm 12.25}$	$84.17_{\pm 14.96}$	$67.60_{\pm 20.08}$	$\mathbf{61.66}_{\pm 30.92}$	$69.02_{\pm 10.51}$	$73.32_{\pm 14.30}$	$74.89_{\pm 17.27}$	$54.34_{\pm 21.77}$	$50.77_{\pm 28.09}$
ADDPM$_{linear}$	$80.51_{\pm 07.64}$	$82.83_{\pm 12.09}$	$82.36_{\pm 19.83}$	$65.23_{\pm 22.82}$	$58.49_{\pm 31.01}$	$68.08_{\pm 10.95}$	$72.13_{\pm 13.98}$	$73.48_{\pm 20.73}$	$52.42_{\pm 24.01}$	$47.54_{\pm 28.71}$
ADDPM$_{cosine}$	$\mathbf{81.45}_{\pm 07.23}$	$\mathbf{83.65}_{\pm 12.28}$	$\mathbf{84.25}_{\pm 14.66}$	$\mathbf{68.00}_{\pm 19.70}$	$61.65_{\pm 31.07}$	$\mathbf{69.34}_{\pm 10.44}$	$\mathbf{73.40}_{\pm 14.34}$	$\mathbf{74.93}_{\pm 17.08}$	$\mathbf{54.71}_{\pm 21.54}$	$\mathbf{50.83}_{\pm 28.29}$
	HD ↓ : (mean±std)					Precision ↑ : (mean±std)(%)				
	WMH	Hip	Brats13			WMH	Hip	Brats13		
Methods			WT	TC	ET			WT	TC	ET
UNet	$3.935_{\pm 03.13}$	$1.819_{\pm 0.885}$	$37.85_{\pm 32.41}$	$34.24_{\pm 28.61}$	$49.51_{\pm 72.74}$	$82.27_{\pm 08.97}$	$82.15_{\pm 12.57}$	$90.21_{\pm 07.50}$	$74.23_{\pm 23.31}$	$73.34_{\pm 33.49}$
AttUNet	$4.190_{\pm 02.86}$	$1.648_{\pm 0.893}$	$44.92_{\pm 27.34}$	$27.33_{\pm 23.94}$	$48.74_{\pm 72.52}$	$80.90_{\pm 09.57}$	$83.70_{\pm 12.68}$	$91.88_{\pm 06.00}$	$79.50_{\pm 20.26}$	$73.43_{\pm 33.63}$
UNet++	$3.915_{\pm 03.04}$	$3.949_{\pm 13.21}$	$36.01_{\pm 25.35}$	$24.76_{\pm 18.58}$	$48.56_{\pm 71.55}$	$83.02_{\pm 09.17}$	$83.93_{\pm 12.80}$	$91.50_{\pm 05.65}$	$73.84_{\pm 22.15}$	$74.80_{\pm 34.50}$
UNETR	$3.619_{\pm 02.46}$	$6.191_{\pm 14.24}$	$52.00_{\pm 30.23}$	$40.61_{\pm 30.18}$	$58.41_{\pm 71.97}$	$82.80_{\pm 08.60}$	$81.45_{\pm 11.90}$	$86.39_{\pm 10.98}$	$80.65_{\pm 18.30}$	$74.28_{\pm 33.71}$
SwinUNet	$5.789_{\pm 04.76}$	$8.602_{\pm 16.70}$	$58.92_{\pm 29.44}$	$44.33_{\pm 32.29}$	$58.47_{\pm 71.85}$	$83.13_{\pm 08.81}$	$81.68_{\pm 13.28}$	$87.65_{\pm 10.12}$	$82.30_{\pm 19.56}$	$76.95_{\pm 33.96}$
Bayesian UNet	$3.828_{\pm 02.90}$	$2.315_{\pm 3.786}$	$32.30_{\pm 28.07}$	$19.90_{\pm 17.99}$	$41.38_{\pm 70.47}$	$79.77_{\pm 10.50}$	$\mathbf{84.76}_{\pm 13.25}$	$92.64_{\pm 05.11}$	$\mathbf{83.52}_{\pm 19.35}$	$75.82_{\pm 34.30}$
Probabilistic UNet	$3.777_{\pm 03.10}$	$1.792_{\pm 1.007}$	$36.28_{\pm 31.44}$	$25.51_{\pm 24.13}$	$45.28_{\pm 72.84}$	$82.40_{\pm 08.95}$	$84.04_{\pm 13.14}$	$90.47_{\pm 07.57}$	$76.22_{\pm 24.65}$	$75.67_{\pm 34.11}$
DDPM	$4.179_{\pm 02.67}$	$1.671_{\pm 0.898}$	$29.68_{\pm 27.78}$	$18.18_{\pm 13.87}$	$44.37_{\pm 72.09}$	$82.04_{\pm 09.81}$	$83.76_{\pm 12.88}$	$92.47_{\pm 05.00}$	$82.46_{\pm 18.77}$	$76.63_{\pm 34.08}$
TDPM	$3.608_{\pm 02.35}$	$1.687_{\pm 0.868}$	$30.07_{\pm 30.97}$	$16.25_{\pm 12.90}$	$40.73_{\pm 71.72}$	$82.53_{\pm 09.16}$	$84.18_{\pm 12.88}$	$92.50_{\pm 04.99}$	$81.96_{\pm 22.45}$	$76.58_{\pm 34.39}$
ES-DDPM	$3.524_{\pm 02.68}$	$1.665_{\pm 0.946}$	$31.38_{\pm 31.23}$	$15.44_{\pm 12.70}$	$39.74_{\pm 72.99}$	$82.77_{\pm 09.03}$	$84.70_{\pm 13.18}$	$92.02_{\pm 06.07}$	$79.18_{\pm 24.17}$	$76.31_{\pm 34.12}$
PD-DDPM	$\mathbf{3.432}_{\pm 02.64}$	$1.609_{\pm 0.916}$	$32.75_{\pm 31.68}$	$22.50_{\pm 22.33}$	$44.37_{\pm 71.60}$	$83.77_{\pm 08.67}$	$84.67_{\pm 13.02}$	$91.73_{\pm 06.67}$	$76.44_{\pm 23.05}$	$75.36_{\pm 33.99}$
ADDPM$_{linear}$	$3.801_{\pm 02.86}$	$1.649_{\pm 0.895}$	$\mathbf{29.27}_{\pm 31.28}$	$15.63_{\pm 15.85}$	$38.28_{\pm 72.47}$	$82.22_{\pm 09.10}$	$84.71_{\pm 13.04}$	$\mathbf{93.47}_{\pm 05.12}$	$79.87_{\pm 23.71}$	$\mathbf{77.79}_{\pm 34.84}$
ADDPM$_{cosine}$	$3.453_{\pm 02.68}$	$\mathbf{1.589}_{\pm 0.939}$	$30.47_{\pm 31.49}$	$\mathbf{15.33}_{\pm 16.33}$	$\mathbf{38.16}_{\pm 73.24}$	$\mathbf{83.93}_{\pm 08.58}$	$84.65_{\pm 13.03}$	$93.44_{\pm 04.77}$	$79.62_{\pm 23.07}$	$77.20_{\pm 34.59}$

medical image segmentation based on the transformer structure. However, none of the above models can estimate the uncertainty of segmentation results. Bayesian U-Net (Gal and Ghahramani, 2016) and Probabilistic U-Net (Baumgartner et al., 2019) are two representative methods that can estimate the uncertainty of the segmentation prediction. Furthermore, we also compare ADDPM with other existing accelerated DDPM models, including TDPM (Zheng et al., 2022), ES-DDPM (Lyu et al., 2022), and PD-DDPM (Guo et al., 2022b). It should be emphasized that the ensemble size in the comparison methods is also set to 5. According to the results in Table 1, for WMH, Hip, and Brats13 segmentation tasks, the pre-segmentation models of ADDPM are selected as AttUNet, AttUNet, and UNet, respectively, as these models achieved the best performance in these three different segmentation tasks compared with other models.

Table 1 demonstrates that ADDPM ($T' = 300$, $T'' = 200$) achieves the best overall performance compared to other methods. Although ADDPM has less improvement in segmentation performance compared to PD-DDPM, ADDPM further reduces the number of iterations required for sampling. Additionally, we found that the truncation-based accelerated DDPMs models (TDPM, ES-DDPM, PD-DDPM, and ADDPM) outperformed the vanilla DDPMs. Moreover, we compared the effect of linear and cosine noise strategies on the model and found that ADDPM performed better under the cosine strategy. Figure 3 illustrates some cases on three segmentation datasets segmented by five individual base masks and their ensemble and uncertainty maps. We observed that five different base segmentation masks generated significantly different results, and the ensemble avoids the worst segmentation result. Additionally, we can clearly identify the areas where the model was uncertain through the uncertainty map.

5.2. Determining Optimal T' and T''

In this analysis, we investigated the impact of the hyperparameter T' on ADDPM. For the WMH segmentation task, we only truncated the initial stages of the inverse process by varying T' among the values of {50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000}. The results, presented in Figure 4, show that the model achieved the best Dice score when $T' = 300$. Furthermore, we observed that truncating the initial stages of the inverse process significantly improves the segmentation performance compared to vanilla DDPMs. We attribute this to the fact that when $t > 300$,

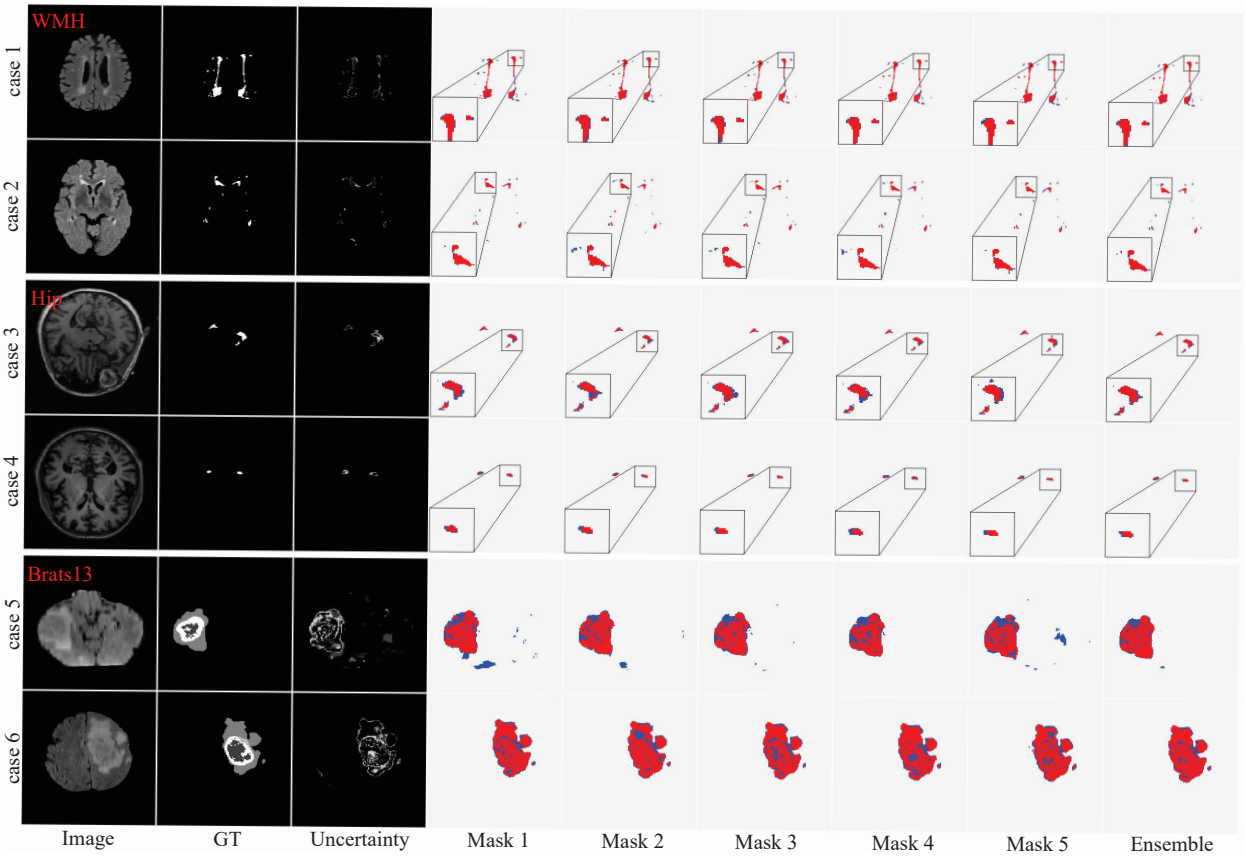

Figure 3: Detailed segmentation results of five base masks, the ensemble mask and uncertainty maps. (a) Image; (b) Ground truth; (c) Uncertainty maps; (d) Base mask 1; (e) Base mask 2; (f) Base mask 3; (g) Base mask 4; (h) Base mask 5; (i) Ensemble mask. The red areas indicate the overlap between the foreground in the predicted segmentation result and the foreground in the Ground truth. The blue ones are the prediction errors. For better visualization, the regions inside the smaller yellow bounding box are zoomed into the larger bounding box.

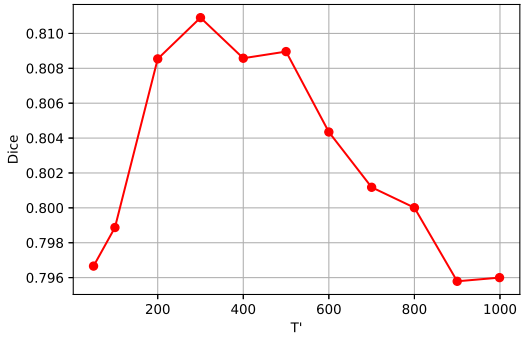

Figure 4: The Dice on testing set with respect to T' (only truncating the initial stages of the inverse process, $T'' = 0$).

the corresponding diffusion rate is relatively high, and the Gaussian assumption may not hold well, which can make it challenging for the model to effectively capture the true distribution of the data (Sohl-Dickstein et al., 2015).

Then, we analyzed the impact of the hyperparameter T'' on ADDPM under the condition of truncating the initial stages of the inverse process ($T' = 300$). By varying T'' among {50, 100, 150, 200, 250, 260, 270, 280, 290, 295}, we trained ADDPM on the WMH segmentation task, which simultaneously truncates the initial and final stages of the

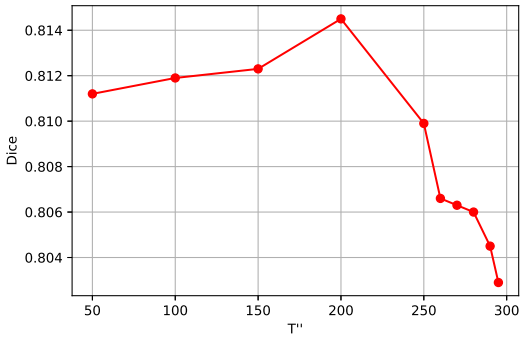

Figure 5: The Dice on testing set with respect to T'' (truncating the initial and final stages of the inverse process, T' is fixed at 300).

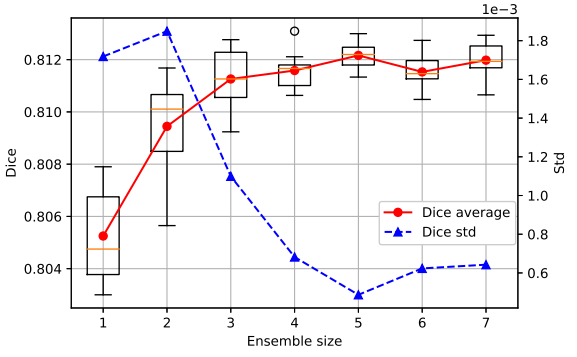

Figure 6: The average and standard deviation of the dice on testing set with respect to ensemble size (T'=300, T''=200).

inverse process. As shown in Figure 5, ADDPM achieves the best Dice score when $T'' = 200$. Compared to truncating the initial of the inverse process, truncating the final of the inverse process has a small improvement in segmentation accuracy but further reduces the number of iterations required for inference.

In conclusion, based on the above analysis, we determined that the optimal truncation positions for the initial and final stages of the inverse process in ADDPM are 300 and 200 respectively. Since the experimental analysis mentioned above is time-consuming, we directly used the optimal T' and T'' settings obtained from the WMH analysis for the Brats13 and Hip segmentation tasks. Table 1 shows that this setting can effectively improve the segmentation performance of the model on the three segmentation tasks.

5.3. Effect of the size of ensembles

Ensemble learning aims at aggregating different base predicitons to boost the model performance. The optimal size of an ensemble, i.e., how many base predicitons in the ensemble are needed, remains an open issue and, as in many related ensemble learning tasks, a task specific parameter that needs to be optimized (Beluch et al., 2018). So we test on WMH with different ensemble sizes and the training process is repeated 3 times.

Figure 6 shows that (1) the ensemble with multiple masks outperformed the only one mask. (2) when ensemble sizes increased, performance tended to saturate. We set the ensemble size to 5 in ADDPM. Figure 6 also shows standard deviation of the Dice score with respect to different ensemble sizes. As the ensemble size increases, the variation of segmentation performance was reduced on the Dice score. The above results show the ensemble of segmentation masks of ADDPM not only boost the segmentation performance, but also ensures a robust segmentation result.

5.4. Effect of pre-segmentation model performance

Here, we analyze the impact of pre-segmentation model performance on ADDPM through the WMH segmentation task. We take different segmentation models in Table 1 as pre-segmentation models, including UNet, AttUnet, UNet++, UNETR, and SwinUNet, which correspond to different pre-segmentation performances. Table 2 lists four

Table 2
The segmentation performance of ADDPM (T'=300, T''=200) based on different pre-segmentation networks. And the performance of different pre-segmentation models.

Methods	Dice	Jaccard	HD	Precision
UNet	$78.71_{\pm 08.13}$	$65.65_{\pm 11.18}$	$3.935_{\pm 3.130}$	$82.27_{\pm 08.97}$
ADDPM	$80.70_{\pm 07.54}$	$68.31_{\pm 10.71}$	$3.396_{\pm 2.242}$	$85.14_{\pm 08.10}$
AttUNet	$79.97_{\pm 07.76}$	$67.34_{\pm 11.01}$	$4.190_{\pm 2.860}$	$80.90_{\pm 09.57}$
ADDPM	$81.06_{\pm 07.61}$	$68.86_{\pm 10.85}$	$3.730_{\pm 2.538}$	$84.00_{\pm 08.94}$
UNet++	$79.87_{\pm 07.84}$	$67.19_{\pm 10.93}$	$3.915_{\pm 3.040}$	$83.02_{\pm 09.17}$
ADDPM	$80.70_{\pm 07.52}$	$68.31_{\pm 10.70}$	$3.608_{\pm 2.365}$	$84.54_{\pm 08.32}$
UNETR	$78.10_{\pm 08.11}$	$64.82_{\pm 11.13}$	$3.619_{\pm 2.460}$	$82.80_{\pm 08.60}$
ADDPM	$80.10_{\pm 07.52}$	$67.47_{\pm 10.56}$	$3.477_{\pm 2.118}$	$85.86_{\pm 06.83}$
SwinUNet	$77.03_{\pm 09.53}$	$63.62_{\pm 11.14}$	$5.790_{\pm 4.760}$	$83.13_{\pm 08.81}$
ADDPM	$79.99_{\pm 07.56}$	$67.33_{\pm 10.63}$	$3.482_{\pm 2.085}$	$85.82_{\pm 06.69}$

Table 3
The segmentation performance of ADDPM (T'=300, T''=200) based on AttUNet with different segmentation performance. And the performance of AttUNet with different segmentation performance.

Pre-segmentation	Dice	Jaccard	HD	Precision
Pre-seg (Dice=59.57)	$59.57_{\pm 19.07}$	$45.05_{\pm 19.51}$	$14.47_{\pm 08.36}$	$66.90_{\pm 18.23}$
ADDPM	$75.86_{\pm 10.53}$	$62.28_{\pm 13.74}$	$06.88_{\pm 05.72}$	$84.64_{\pm 06.90}$
Pre-seg (Dice=67.53)	$67.54_{\pm 16.97}$	$53.33_{\pm 18.50}$	$08.72_{\pm 06.07}$	$76.74_{\pm 19.02}$
ADDPM	$76.82_{\pm 09.70}$	$63.39_{\pm 13.04}$	$06.61_{\pm 05.62}$	$85.55_{\pm 08.47}$
Pre-seg (Dice=71.82)	$71.83_{\pm 11.61}$	$57.34_{\pm 14.45}$	$08.26_{\pm 06.33}$	$82.56_{\pm 09.81}$
ADDPM	$78.22_{\pm 08.04}$	$64.96_{\pm 11.06}$	$05.51_{\pm 03.66}$	$87.79_{\pm 06.11}$
Pre-seg (Dice=74.13)	$74.14_{\pm 12.16}$	$60.34_{\pm 15.01}$	$06.67_{\pm 05.94}$	$79.31_{\pm 15.37}$
ADDPM	$78.12_{\pm 09.12}$	$65.03_{\pm 12.49}$	$04.87_{\pm 03.28}$	$83.99_{\pm 10.48}$
Pre-seg (Dice=75.43)	$75.43_{\pm 12.68}$	$62.13_{\pm 15.60}$	$06.59_{\pm 05.96}$	$74.98_{\pm 16.74}$
ADDPM	$79.12_{\pm 09.35}$	$66.43_{\pm 12.70}$	$04.23_{\pm 03.58}$	$80.70_{\pm 12.82}$
Pre-seg (Dice=77.04)	$77.05_{\pm 09.34}$	$63.61_{\pm 12.52}$	$05.65_{\pm 05.47}$	$81.70_{\pm 10.97}$
ADDPM	$79.47_{\pm 08.06}$	$66.69_{\pm 11.28}$	$04.18_{\pm 02.88}$	$84.94_{\pm 09.07}$
Pre-seg (Dice=78.54)	$78.54_{\pm 08.77}$	$65.54_{\pm 12.09}$	$05.26_{\pm 05.07}$	$81.68_{\pm 09.66}$
ADDPM	$80.36_{\pm 07.99}$	$67.93_{\pm 11.26}$	$03.87_{\pm 02.48}$	$84.50_{\pm 08.87}$
Pre-seg (Dice=79.97)	$79.97_{\pm 07.76}$	$67.34_{\pm 11.01}$	$04.19_{\pm 04.13}$	$82.27_{\pm 08.97}$
ADDPM	$81.40_{\pm 06.53}$	$68.93_{\pm 10.53}$	$03.49_{\pm 02.53}$	$83.80_{\pm 08.54}$

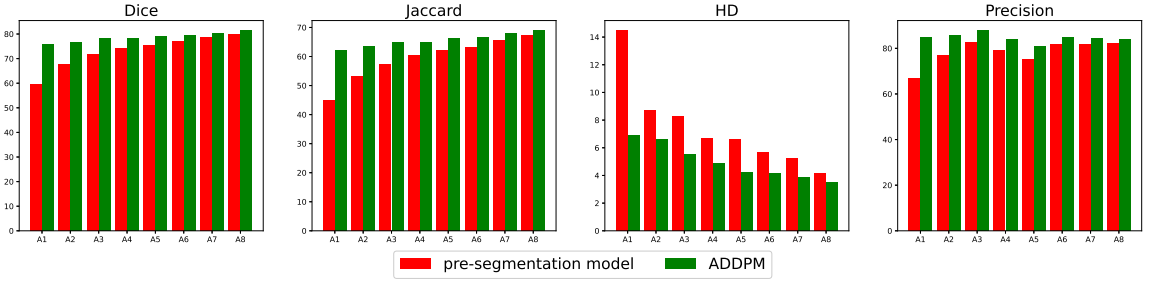

Figure 7: Visualize the segmentation performance of ADDPM (T'=300, T''=200) based on AttUNet with different segmentation performance and the performance of AttUNet with different segmentation performance. $A1$-8 represent the pre-segmented AttUNet with different performances respectively, corresponding to the 8 pre-segmented models in Table 3.

different metrics of ADDPM under different pre-segmentation models. We observed that ADDPM outperforms all the corresponding pre-segmentation models. Therefore, ADDPM can be combined with existing advanced segmentation networks to further improve performance and obtain uncertainty estimates.

We also analyzed the segmentation performance of ADDPM based on the AttUNet network with different pre-segmentation performance, as shown in the Table 3. Here we use AttUNet with different segmentation performance during training as the pre-segmentation model. Table 3 shows that ADDPM all outperforms the corresponding pre-segmentation models. Figure 7 also intuitively visualizes the performance of ADDPM and corresponding pre-segmentation. We can easily find that the lower the performance of the pre-segmentation model, the greater the performance improvement brought by ADDPM (especially in the three metrics of Dice, HD and Jaccard).

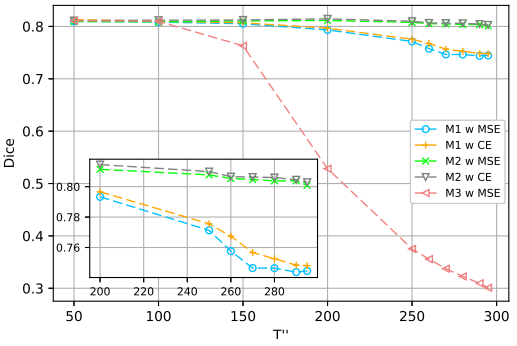

Figure 8: The Dice score for different denoising methods with respect to different T''. And T' is fixed at 300.

Table 4
The best segmentation performance of the different denoising methods and the corresponding optimal steps T''.

Method	CE	MSE	T''	Dice	Jaccard	HD	Precision
M1	√	×	150	$81.31_{\pm 07.20}$	$69.13_{\pm 10.37}$	$\mathbf{3.327}_{\pm 2.613}$	$83.50_{\pm 08.68}$
M1	×	√	150	$80.97_{\pm 07.36}$	$68.68_{\pm 10.56}$	$3.705_{\pm 2.688}$	$\mathbf{84.03}_{\pm 08.70}$
M2	√	×	200	$\mathbf{81.45}_{\pm 07.23}$	$\mathbf{69.34}_{\pm 10.44}$	$3.453_{\pm 2.676}$	$83.93_{\pm 08.58}$
M2	×	√	200	$81.14_{\pm 07.27}$	$68.91_{\pm 10.50}$	$3.476_{\pm 2.716}$	$83.78_{\pm 08.62}$
M3	×	√	100	$81.01_{\pm 07.38}$	$68.74_{\pm 10.63}$	$3.329_{\pm 2.637}$	$83.29_{\pm 08.86}$

5.5. Ablation study of different denoising methods with truncating the final stages of the inverse process

The truncating the final stages of the inverse process is achieved by denoising the noise prediction $x_{T''}$ through a separately denoising network f_ψ. There are several ways to get the final segmentation result y from $x_{T''}$. Here we compare these methods in detail by WMH segmentation task.

Method 1: The y is obtained directly from the $x_{T''}$. This method takes the $x_{T''}$ as the input of the denoising network, and the denoising network outputs the y.

Method 2: The y is obtained by the $x_{T''}$ and the image I. The I is concatenated with the $x_{T''}$ as the input of the denoising network, and the denoising network outputs the y.

Method 3: Based on the $x_{T''}$, the denoising network first learns the noise $n_{x_{T''}}$ contained in the $x_{T''}$. The estimated noise $n_{x_{T''}}$ is then subtracted from the $x_{T''}$ to obtain the y.

Based on the above three implementation methods, we also analyzed the impact of two different loss functions (Cross entropy and MSE) on the denoising network f_ψ. Table 4 shows the best segmentation performance of the above methods and the corresponding optimal steps T''. Figure 8 shows the Dice score for different denoising methods with respect to different T''. By varying the T'' among $\{50, 100, 150, 200, 250, 260, 270, 280, 290, 295\}$, we train ADDPM. Compared with the other two denoising methods, **Method 2** achieve better performance, and its performance is the best when T'' is equal to 200. And the performance of the Cross-entropy on **Method 2** is also better than MSE.

6. CONCLUSION

This paper proposes an accelerated denoising diffusion probabilistic model via truncated inverse processes (ADDPM) that is specifically designed for medical image segmentation. The key idea of ADDPM is to truncate the inverse processes to consider only a small number of steps in the middle of the inference process. Experiments demonstrate that ADDPM achieves superior performance compared to vanilla DDPMs, even with a significantly reduced number of inference iteration steps (from 1000 steps to 100 steps in the best case of our experiments). And ADDPM outperforms existing acceleration methods for DDPMs. Further, ADDPM can be integrated with existing advanced segmentation models to further enhance segmentation performance and obtain uncertainty estimation.

References

Baumgartner, C.F., Tezcan, K.C., Chaitanya, K., Hötker, A.M., Muehlematter, U.J., Schawkat, K., Becker, A.S., Donati, O., Konukoglu, E., 2019. Phiseg: Capturing uncertainty in medical image segmentation, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22, Springer. pp. 119–127.

Begoli, E., Bhattacharya, T., Kusnezov, D., 2019. The need for uncertainty quantification in machine-assisted medical decision making. Nature Machine Intelligence 1, 20–23.

Beluch, W.H., Genewein, T., Nürnberger, A., Köhler, J.M., 2018. The power of ensembles for active learning in image classification, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 9368–9377.

Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M., 2023. Swin-unet: Unet-like pure transformer for medical image segmentation, in: Computer Vision–ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part III, Springer. pp. 205–218.

Cao, Y., Liu, S., Peng, Y., Li, J., 2020. Denseunet: densely connected unet for electron microscopy image segmentation. IET Image Processing 14, 2682–2689.

Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y., 2021. Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 .

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence 40, 834–848.

Dalmaz, O., Yurt, M., Çukur, T., 2022. Resvit: residual vision transformers for multimodal medical image synthesis. IEEE Transactions on Medical Imaging 41, 2598–2614.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al., . An image is worth 16x16 words: Transformers for image recognition at scale, in: International Conference on Learning Representations.

Esser, P., Rombach, R., Blattmann, A., Ommer, B., 2021. Imagebart: Bidirectional context with multinomial diffusion for autoregressive image synthesis. Advances in Neural Information Processing Systems 34, 3518–3532.

Gal, Y., Ghahramani, Z., 2016. Dropout as a bayesian approximation: Representing model uncertainty in deep learning, in: international conference on machine learning, PMLR. pp. 1050–1059.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2020. Generative adversarial networks. Communications of the ACM 63, 139–144.

Guo, X., Lu, S., Yang, Y., Shi, P., Ye, C., Xiang, Y., Ma, T., 2022a. Modeling annotator variation and annotator preference for multiple annotations medical image segmentation, in: 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE. pp. 977–984.

Guo, X., Yang, Y., Ye, C., Lu, S., Xiang, Y., Ma, T., 2022b. Accelerating diffusion models via pre-segmentation diffusion sampling for medical image segmentation. arXiv preprint arXiv:2210.17408 .

Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H.R., Xu, D., 2022. Unetr: Transformers for 3d medical image segmentation, in: Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp. 574–584.

He, S., Grant, P.E., Ou, Y., 2021. Global-local transformer for brain age estimation. IEEE transactions on medical imaging 41, 213–224.

Ho, J., Jain, A., Abbeel, P., 2020. Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems 33, 6840–6851.

Hu, D., Tao, Y.K., Oguz, I., 2022. Unsupervised denoising of retinal oct with diffusion probabilistic model, in: Medical Imaging 2022: Image Processing, SPIE. pp. 25–34.

Kingma, D., Salimans, T., Poole, B., Ho, J., 2021. Variational diffusion models. Advances in neural information processing systems 34, 21696–21707.

Kingma, D.P., Welling, M., 2014. Auto-encoding variational bayes. stat 1050, 1.

Kohl, S., Romera-Paredes, B., Meyer, C., De Fauw, J., Ledsam, J.R., Maier-Hein, K., Eslami, S., Jimenez Rezende, D., Ronneberger, O., 2018. A probabilistic u-net for segmentation of ambiguous images. Advances in neural information processing systems 31.

Kuijf, H.J., Biesbroek, J.M., De Bresser, J., Heinen, R., Andermatt, S., Bento, M., Berseth, M., Belyaev, M., Cardoso, M.J., Casamitjana, A., et al., 2019. Standardized assessment of automatic segmentation of white matter hyperintensities and results of the wmh segmentation challenge. IEEE transactions on medical imaging 38, 2556–2568.

Li, H., Yang, Y., Chang, M., Chen, S., Feng, H., Xu, Z., Li, Q., Chen, Y., 2022. Srdiff: Single image super-resolution with diffusion probabilistic models. Neurocomputing 479, 47–59.

Liao, Z., Xie, Y., Hu, S., Xia, Y., 2022. Learning from ambiguous labels for lung nodule malignancy prediction. IEEE Transactions on Medical Imaging 41, 1874–1884.

Liu, H., Wang, Y., Wang, M., Rui, Y., 2022. Delving globally into texture and structure for image inpainting, in: Proceedings of the 30th ACM International Conference on Multimedia, pp. 1270–1278.

Lyu, Z., Xu, X., Yang, C., Lin, D., Dai, B., 2022. Accelerating diffusion models via early stop of the diffusion process. arXiv preprint arXiv:2205.12524 .

Malekzadeh, S., 2019. Mri hippocampus segmentation. Kaggle .

Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al., 2014. The multimodal brain tumor image segmentation benchmark (brats). IEEE transactions on medical imaging 34, 1993–2024.

Nichol, A.Q., Dhariwal, P., 2021. Improved denoising diffusion probabilistic models, in: International Conference on Machine Learning, PMLR. pp. 8162–8171.

Oktay, O., Schlemper, J., Le Folgoc, L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al., . Attention u-net: Learning where to look for the pancreas, in: Medical Imaging with Deep Learning.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B., 2022. High-resolution image synthesis with latent diffusion models, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10684–10695.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part

III 18, Springer. pp. 234–241.

Salimans, T., Ho, J., . Progressive distillation for fast sampling of diffusion models, in: International Conference on Learning Representations.

San-Roman, R., Nachmani, E., Wolf, L., 2021. Noise estimation for generative diffusion models. arXiv preprint arXiv:2104.02600 .

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S., 2015. Deep unsupervised learning using nonequilibrium thermodynamics, in: International Conference on Machine Learning, PMLR. pp. 2256–2265.

Song, J., Chen, X., Zhu, Q., Shi, F., Xiang, D., Chen, Z., Fan, Y., Pan, L., Zhu, W., 2022. Global and local feature reconstruction for medical image segmentation. IEEE Transactions on Medical Imaging 41, 2273–2284.

Song, J., Meng, C., Ermon, S., . Denoising diffusion implicit models, in: International Conference on Learning Representations.

Vahdat, A., Kreis, K., Kautz, J., 2021. Score-based generative modeling in latent space. Advances in Neural Information Processing Systems 34, 11287–11302.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Advances in neural information processing systems 30.

Watson, D., Chan, W., Ho, J., Norouzi, M., 2022. Learning fast samplers for diffusion models by differentiating through sample quality, in: International Conference on Learning Representations.

Wolleb, J., Sandkühler, R., Bieder, F., Valmaggia, P., Cattin, P.C., 2022. Diffusion models for implicit image segmentation ensembles, in: International Conference on Medical Imaging with Deep Learning, PMLR. pp. 1336–1348.

Xiao, X., Lian, S., Luo, Z., Li, S., 2018. Weighted res-unet for high-quality retina vessel segmentation, in: 2018 9th international conference on information technology in medicine and education (ITME), IEEE. pp. 327–331.

Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2881–2890.

Zheng, H., He, P., Chen, W., Zhou, M., 2022. Truncated diffusion probabilistic models and diffusion-based adversarial auto-encoders. arXiv preprint arXiv:2202.09671 .

Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J., 2019. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE transactions on medical imaging 39, 1856–1867.