applied sciences

MDPI

Article

Practical Medical Image Generation with Provable Privacy Protection Based on Denoising Diffusion Probabilistic Models for High-Resolution Volumetric Images

Hisaichi Shibata [1,*], Shouhei Hanaoka [1], Takahiro Nakao [2], Tomohiro Kikuchi [2,3], Yuta Nakamura [2], Yukihiro Nomura [2,4], Takeharu Yoshikawa [2] and Osamu Abe [1]

[1] Department of Radiology, The University of Tokyo Hospital, 7-3-1 Hongo, Bunkyo, Tokyo 113-8655, Japan
[2] Department of Computational Diagnostic Radiology and Preventive Medicine, The University of Tokyo Hospital, 7-3-1 Hongo, Bunkyo, Tokyo 113-8655, Japan
[3] Department of Radiology, School of Medicine, Jichi Medical University, 3311-1 Yakushiji, Shimotsuke, Tochigi 329-0498, Japan
[4] Center for Frontier Medical Engineering, Chiba University, 1-33 Yayoi-cho, Inage, Chiba 263-8522, Japan
* Correspondence: sh@g.ecc.u-tokyo.ac.jp

Abstract: Local differential privacy algorithms combined with deep generative models can enhance secure medical image sharing among researchers in the public domain without central administrators; however, these images were limited to the generation of low-resolution images, which are very insufficient for diagnosis by medical doctors. To enhance the performance of deep generative models so that they can generate high-resolution medical images, we propose a large-scale diffusion model that can, for the first time, unconditionally generate high-resolution ($256 \times 256 \times 256$) volumetric medical images (head magnetic resonance images). This diffusion model has 19 billion parameters, but to make it easy to train it, we temporally divided the model into 200 submodels, each of which has 95 million parameters. Moreover, on the basis of this new diffusion model, we propose another formulation of image anonymization with which the processed images can satisfy provable Gaussian local differential privacy and with which we can generate images semantically different from the original image but belonging to the same class. We believe that the formulation of this new diffusion model and the implementation of local differential privacy algorithms combined with the diffusion models can contribute to the secure sharing of practical images upstream of data processing.

Keywords: deep generative models; denoising; differential privacy; diffusion models; head magnetic resonance images

Citation: Shibata, H.; Hanaoka, S.; Nakao, T.; Kikuchi, T.; Nakamura, Y.; Nomura, Y.; Yoshikawa, T.; Abe, O. Practical Medical Image Generation with Provable Privacy Protection Based on Denoising Diffusion Probabilistic Models for High-Resolution Volumetric Images. *Appl. Sci.* **2024**, *14*, 3489. https://doi.org/10.3390/app14083489

Academic Editors: Zongwei Zhou, Hongming Shan and Ruibin Feng

Received: 19 March 2024
Revised: 19 April 2024
Accepted: 19 April 2024
Published: 20 April 2024

1. Introduction

Local differential privacy (LDP) [1,2] combined with the deep generative models [3,4] can simultaneously ensure the provable and controllable upper bound of information leakage and hold realistic utilities [5,6] upstream of data processing. Therefore, we consider that this combination can be a realistic tool to anonymize and then share private medical images among researchers in the public domain. On the other hand, the total dimension of data modeled by deep generative models for images is typically limited to, equivalent to, or below 2,097,152 (see the section on related works for details). However, to diagnose patients with volumetric medical images, medical doctors usually want those images to have a resolution of at least $256 \times 256 \times 256 = 16{,}777{,}216$ total dimensions. On the basis of these backgrounds, the purpose of the present study is to build a deep generative model that can learn volumetric images with state-of-the-art high-resolution ($256 \times 256 \times 256$) and sample from the probabilistic distribution of those images to apply LDP algorithms with it for high-resolution volumetric medical images. Specifically, we adopt diffusion models as the deep generative models because the diffusion models can divide the training problem

for the probabilistic distribution of images to virtually infinite subproblems to enhance the power of expression. With reference to this division, the stochastic differential equations for the diffusion models [7] are an interesting topic; however, we do not handle those equations in the present study. More strictly, we adopt pixel-space diffusion models [3] because they can straightforwardly handle real images without encoders, as in generative adversarial networks (GANs) [4] and latent-space diffusion models [8], which is an important and favorable feature in LDP processing. Furthermore, a pair of an encoder and a decoder for the latent-space diffusion models requires a significantly large GPU memory when we scale up the models to high-resolution volumetric images. Finally, we not only show unconditional image generation from a pseudo-random noise with the proposed diffusion model, but also apply the proposed model to generate high-resolution LDP volumetric medical images from a real image. Because differential privacy (DP) provides provable privacy protection, we decided to concentrate on clarifying the utility of those generated images, including LDP-processed images with visual evaluations by three medical doctors. We adopt head magnetic resonance (MR) images taken at the University of Tokyo Hospital throughout the present study. In summary, this study enhances diffusion models to facilitate modeling high-resolution volumetric images, contributing to the construction of practical medical systems. Additionally, we apply this improved diffusion model to propose and validate a method for sharing medical images with guaranteed privacy.

2. Related Works

2.1. Deep Generative Models for High-Dimensional Data

Shibata et al. [9] improved the training method of flow-based deep generative models [10–12] and successfully modeled three-dimensional chest computed tomography (CT) images with the models, but the image resolution was limited, that is, equivalent to or below $128 \times 128 \times 128$. Khader et al. [13] adopted denoising diffusion probabilistic models (DDPMs) to unconditionally generate three-dimensional medical images, but the total dimensions of image pixels were limited, that is, equivalent to or below 2,097,152. Bieder et al. [14] adopted DDPMs to segment three-dimensional high-resolution ($256 \times 256 \times 256$) medical images, but they did not report the results of their image generation. Dorjsembe et al. [15] adopted DDPMs to conditionally generate three-dimensional medical images, but their resolution was limited to $128 \times 128 \times 128$. Finally, Sun et al. [16] proposed a three-dimensional GAN, which can unconditionally generate high-resolution ($256 \times 256 \times 256$) medical images, but they did not report image generations conditioned with another image with a GAN, unlike in our present study.

2.2. Differential Privacy for Multidimensional Data

Fan [17] proposed an LDP algorithm for image anonymization. The algorithm directly adds a perturbation noise on images. In the present study, we adopt the algorithm, but we additionally postprocess the LDP-processed noisy images. Croft et al. [6], Li and Clifton [18], and Liu et al. [19] almost simultaneously proposed another LDP algorithm for the obfuscation of facial images, which adopts generative models to semantically change the identity in facial images while it can preserve the class (face) of those images. However, they did not adopt diffusion models as generative models. We stress that the above LDP algorithms are all different from DP-SGD [20], which adds perturbation noise against parameters of deep discriminative or generative models.

3. Methods

Figure 1 shows the flowchart of the proposed method.

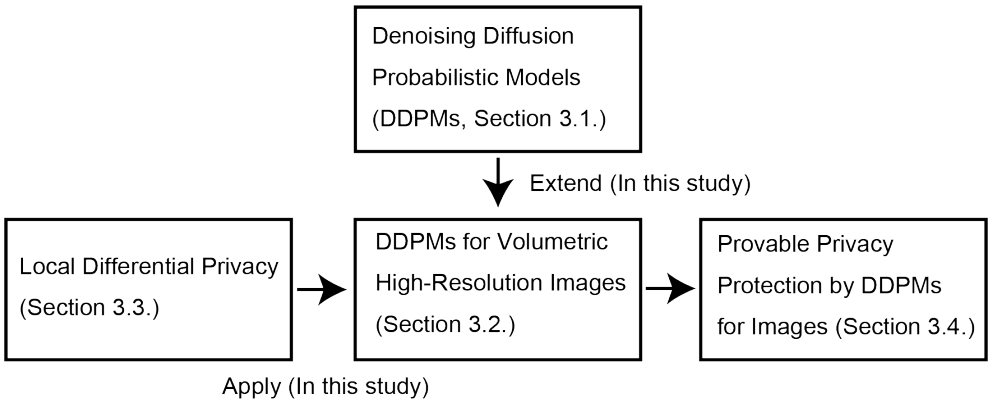

Figure 1. We extend the denoising diffusion probabilistic model (DDPM) to handle high-resolution volumetric images, and furthermore, propose and validate a novel method to remove noise induced by local differential privacy (LDP) using DDPM.

3.1. Score Matching and DDPMs

Given the domain-specific data (we represent the set of the data as $\mathcal{D}$ and one of the data as $\boldsymbol{x}$), the purpose of the generative models is to estimate the probabilistic distribution $\log p(\boldsymbol{x})$ of those data. However, the direct modeling of $\log p(\boldsymbol{x})$ is very difficult. Therefore, we focus on *score* $\nabla_{\boldsymbol{x}} \log p(\boldsymbol{x})$. However, the direct modeling of $\nabla_{\boldsymbol{x}} \log p(\boldsymbol{x})$ is still difficult. Therefore, we estimate the score with implicit score matching [21]. However, the scale up of implicit score matching to high-dimensional data is difficult and overfitting is significantly problematic. Therefore, we divide the problem. We add the perturbation noise of different signal-to-noise ratios (SNRs) to data and prepare datasets characterized with different SNRs. The implicit score matching now learns *denoising* from noisy data.

The denoising diffusion probabilistic models (DDPMs) [3] are another expression of score-based diffusion models and formulated to maximize the evidence lower bound (ELBO) of the Kullback–Leibler divergence. Maximizing the ELBO is equivalent to learning noise prediction from noisy data. Specifically, we train a model, which is represented as a vector function conditioned on the time step $\boldsymbol{f}_{t,\boldsymbol{\theta}}$, so that the model can predict the noise component at a time step t, which is linearly combined with the normalized image ($\boldsymbol{x}_t \in [-1, 1]$):

$$\boldsymbol{x}_t = \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sqrt{\bar{\beta}_t}\boldsymbol{\epsilon}, \tag{1}$$

and

$$\hat{\boldsymbol{x}}_{t-1} = \boldsymbol{f}_{t,\boldsymbol{\theta}}(\boldsymbol{x}_t), \tag{2}$$

where the hat represents the predicted quantity throughout this paper, and

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}), \tag{3}$$

$$\alpha_t := 1 - \beta_t, \tag{4}$$

$$\bar{\alpha}_t := \prod_{s=1}^{t} \alpha_s, \tag{5}$$

$$\bar{\beta}_t := 1 - \bar{\alpha}_t, \tag{6}$$

where $0 < \beta_1 < \beta_2 < \cdots < \beta_T < 1$ are the control parameters for the magnitude of deviations. In this study, we adopt the Sigmoid scheduling [22] for the betas.

3.2. Scale Up to High-Resolution Volumetric Images by Model Unrolling

Previous DDPMs recursively adopt the same neural network to execute denoising from images of different SNRs with the time step information prescribed inside the denoising network:

$$\hat{\boldsymbol{x}}_0 = \underbrace{\boldsymbol{f}_{t=1,\boldsymbol{\theta}} \circ \boldsymbol{f}_{t=2,\boldsymbol{\theta}} \circ \cdots \circ \boldsymbol{f}_{t=T,\boldsymbol{\theta}}}_{T\text{-steps}}(\boldsymbol{x}_T). \tag{7}$$

This DDPM network ($\boldsymbol{f}_{\boldsymbol{\theta}}$) requires a large power of expression, i.e., numerous parameters, to enable denoising from images of different SNRs, but it is very difficult to scale this up to high-resolution volumetric images owing to the GPU memory limitation. On the contrary, if we train a different DDPM network for each time step, the power of expression required for each model would be significantly relaxed. On the basis of this insight, we have

$$\hat{\boldsymbol{x}}_0 = \underbrace{\boldsymbol{f}_{\boldsymbol{\theta}_1} \circ \boldsymbol{f}_{\boldsymbol{\theta}_2} \circ \cdots \circ \boldsymbol{f}_{\boldsymbol{\theta}_T}}_{T\text{-steps}}(\boldsymbol{x}_T), \tag{8}$$

where $\boldsymbol{f}_{\boldsymbol{\theta}_i}$ learns denoising from $\boldsymbol{x}_t$ to estimate the noise component and it can generate the image in the previous time step $\boldsymbol{x}_{t-1}$. Specifically, for training, we prepare multiple instances of the network shown in Figure A1 for different time steps (t), and optimize the parameters included in them using the ADAM optimizer.

3.3. DP

We deal with the LDP apart from global differential privacy. The LDP can anonymize data itself upstream of data processing. The theoretical guarantee for privacy protection of LDP with the Gaussian mechanism [1,23] is given by

$$Pr(\tilde{\boldsymbol{x}}|\boldsymbol{x}) - e^{\epsilon} Pr(\tilde{\boldsymbol{x}}|\boldsymbol{x}') \leq \delta, \tag{9}$$

where $\tilde{\boldsymbol{x}}$ is an LDP-processed image, $\boldsymbol{x}$ and $\boldsymbol{x}'$ are different arbitrarily selected images in the probabilistic distribution of images, $\epsilon \geq 0$ and $0 \leq \delta < 1$ are the privacy budgets specified by administrators of the images, and $Pr(\tilde{\boldsymbol{x}}|\boldsymbol{x})$ [or $Pr(\tilde{\boldsymbol{x}}|\boldsymbol{x}')$] is the conditional probability that $\tilde{\boldsymbol{x}}$ is generated when $\boldsymbol{x}$ (or $\boldsymbol{x}'$) is given.

To generate $\tilde{\boldsymbol{x}}$ from $\boldsymbol{x}$ (or, $\boldsymbol{x}'$), which satisfies (9), we add a perturbation that obeys a normal (Gaussian) distribution on $\boldsymbol{x}$ as follows:

$$\tilde{x}_{ijk} = x_{ijk} + n, \tag{10}$$

$$n \sim \mathcal{N}(0, \sigma^2), \tag{11}$$

$$\sigma^2 = \frac{2 \ln 1.25/\delta_{ijk} \cdot (\Delta f_{ijk})^2}{\epsilon_{ijk}^2}, \tag{12}$$

where for a single pixel (x_{ijk}) of $\boldsymbol{x}$, ϵ_{ijk} and δ_{ijk} are the privacy budgets, and Δf_{ijk} is the sensitivity. After the addition of the perturbation, we clip the range of $\tilde{x}_{ijk}$ from -1 to 1.

3.4. Integration of DDPM and DP

First, we handle $(\epsilon_{ijk}, \delta_{ijk})$-Gaussian-LDP for a single fixed pixel (x_{ijk}) and extend this to handling all the pixels using the following composition theorem of DP:

$$\epsilon = \sum_{i,j,k} \epsilon_{ijk}, \tag{13}$$

$$\delta = \sum_{i,j,k} \delta_{ijk}, \tag{14}$$

where ϵ and δ are the total privacy budgets.

Because the deviations of perturbed images appear in DDPMs (1), we can write

$$\sigma_t^2 = \frac{\bar{\beta}_t}{\bar{\alpha}_t}. \tag{15}$$

We can now connect the above equation with (12) as

$$\sigma_t^2 = \sigma^2. \tag{16}$$

This indicates that we can compute the virtually infinite pairs of ϵ and δ if the noise scheduling (betas) of the DDPM and the sensitivity Δf_{ijk} are given.

Note that we can obtain LDP-processed images by just picking intermediate images $\boldsymbol{x}_t$ in the DDPM (the forward process), but those images do not preserve the class (head MR images in the present study) if the privacy budgets are small. To preserve the class even when the privacy budgets are small, we reversely apply the DDPM to the intermediate images as follows (the reverse process):

$$\hat{\boldsymbol{x}}_0 = \underbrace{\boldsymbol{f}_{\boldsymbol{\theta}_1} \circ \boldsymbol{f}_{\boldsymbol{\theta}_2} \circ \cdots \circ \boldsymbol{f}_{\boldsymbol{\theta}_t}}_{t\text{-steps}}(\boldsymbol{x}_t). \tag{17}$$

With this postprocessing, we can have images such that they (i) preserve the class, (ii) may be semantically different from original images, and (iii) have provable indistinguishability and therefore are practical for medical data sharing.

4. Numerical Experiments

4.1. Preparation of Head MR Images

The institutional review board of the University of Tokyo Hospital approved the use of head MR images (T1WI) taken in the hospital for the present retrospective study. From November 2006 to December 2017, high-resolution volumetric fast spoiled gradient-echo MR imaging was performed with a Signa EXCITE and a Discovery MR750 scanner (GE Healthcare Japan, Tokyo, Japan) (repetition time, 6.4 ms; echo time, 2.0 ms; inversion time, 450 ms; field-of-view, 25 cm; flip angle, 15 deg; acquisition matrix, 256 $\times$ 256; number of excitations, 0.5; and voxel dimensions 0.98 $\times$ 0.98 $\times$ 1.0 mm^3) using an 8-channel head coil. From January 2018 to April 2021, high-resolution volumetric magnetization-prepared rapid gradient-echo MR images were acquired with a Biograph mMR scanner (repetition time, 1660 ms; echo time, 2.4 ms; inversion time, 910 ms; field-of-view, 25 cm; flip angle, 8 deg; acquisition matrix, 256 $\times$ 256; number of excitations, 1; and voxel dimensions, 0.98 $\times$ 0.98 $\times$ 1.0 mm^3) using a 16-channel receiver coil (These protocols are the same as in [24].). We extracted only 1327 head MR images (volumes) from a large set (including both normal and abnormal cases) of images taken with the above protocols. The resolution of each two-dimensional MR image (sagittal slices) was 256 $\times$ 256. We stacked these two-dimensional images to create a three-dimensional image of 256 $\times$ 256 $\times$ 256 size with padding. These head MR images were processed without skull stripping. The images were divided into training (1224) and test (103) datasets. Each image was normalized so that the whole pixel range of all the images was included. We then mapped the range of these images onto $[-1, 1]$ for the training.

4.2. Training of DDPMs

We trained the models in an unsupervised manner with the DDPM network described in Appendix A without the time step information (t). We set the number of the total time steps ($T = 200$); hence, we trained 200 models ($\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_{200}$). Each model for the fixed time step contains 95 million trainable parameters and all the models together contain 19 billion trainable parameters. We efficiently trained the models in a serpentine folding manner. Specifically, we first trained the model $\boldsymbol{\theta}_1$ and when finished, we initialized the next model $\boldsymbol{\theta}_2$ with the parameters of $\boldsymbol{\theta}_1$ and then trained it. When all the models ($\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_{200}$) were updated once, we define this process as completinig one epoch. We

updated the model until four epochs. All the computations were executed on a single node of the supercomputer Wisteria/BDEC-01 (Fujitsu, Tokyo, Japan) (one node contains eight A100 GPUs with 40 GB of memory for each GPU), at the University of Tokyo.

5. Results

5.1. Unconditional Image Generation and Visual Evaluation

Using the trained model, we first unconditionally generated head MR images [i.e., $(\epsilon, \delta) = (0, 0)$]. Three medical doctors (radiologists) evaluated the quality of those images (six volumes) and real images (six volumes) to show the capability of the model to generate realistic but fictional high-resolution volumetric medical images. They evaluated the appearance of anatomical structures and the contrast of the cortical white matter in five stages on the basis of the criteria shown in Table 1. Tables 2 and 3 show the results of the evaluation of real and fake cases, respectively. Moreover, in Figure 2, we show three representative slices (i.e., axial, coronal, and sagittal slices) of a generated head MR image (we selected case 1 in Tables A1–A3.) from pseudo random noise, i.e., from $t = 200$ in (17). The medical images obtained in the present study were cropped and enlarged to enhance their visibility.

Table 1. Evaluation Criteria.

Stages	Criteria
1	The structure is invisible.
2	The structure is slightly identifiable.
3	The structure is visible but not sufficient.
4	The structure is visible as in real cases with the resolution of $256 \times 256 \times 256$.
5	The structure is well visible as in real cases with the resolution of $256 \times 256 \times 256$.

Table 2. Visual evaluation results. A, B, and C indicate averaged evaluation results for six real cases. Ave. indicates the averaged results (A, B, and C) from the three medical doctors.

Doctor	A	B	C	Ave. (256^3)
Brain regions				
Anterior commissure	5.0	5.0	5.0	5.0
Posterior commissure	5.0	5.0	5.0	5.0
Cerebral aqueduct	5.0	5.0	5.0	5.0
Tegmentum of midbrain	5.0	5.0	5.0	5.0
Cerebellar hemisphere sulcus	5.0	5.0	4.7	4.9
Cerebral peduncle	5.0	5.0	5.0	5.0
Corpus callosum	5.0	5.0	5.0	5.0
Third ventricle	5.0	5.0	5.0	5.0
Fourth ventricle	5.0	5.0	5.0	5.0
Lateral ventricle	5.0	5.0	5.0	5.0
Cortical white matter contrast				
Hippocampus	5.0	5.0	4.8	4.9
Frontal lobe	5.0	5.0	4.8	4.9
Occipital lobe	5.0	5.0	4.8	4.9
Temporal lobe	5.0	5.0	5.0	5.0
Parietal lobe	5.0	5.0	5.0	5.0
Basal ganglia	4.8	4.8	4.8	4.8
Other regions				
First cervical vertebra	5.0	5.0	5.0	5.0
Second cervical vertebra	5.0	5.0	5.0	5.0
Optic nerve	5.0	5.0	4.8	4.9
Extraocular muscles	5.0	5.0	5.0	5.0

Table 3. Visual evaluation results. A, B, and C indicate averaged evaluation results for six fake cases. Ave. indicates the averaged results (A, B, and C) from the three medical doctors.

Doctor	A	B	C	Ave. (256^3)
Brain regions				
Anterior commissure	3.5	2.0	3.2	2.9
Posterior commissure	1.8	1.5	2.3	1.9
Cerebral aqueduct	1.5	1.3	2.2	1.7
Tegmentum of midbrain	3.0	3.0	2.3	2.8
Cerebellar hemisphere sulcus	2.0	2.7	3.3	2.7
Cerebral peduncle	3.2	2.8	5.0	3.7
Corpus callosum	4.0	4.2	5.0	4.4
Third ventricle	3.2	4.2	4.7	4.0
Fourth ventricle	4.5	4.2	4.7	4.5
Lateral ventricle	5.0	4.2	3.7	4.3
Corticomedullary contrast				
Hippocampus	1.7	2.3	3.2	2.4
Frontal lobe	2.3	1.7	2.8	2.3
Occipital lobe	1.5	1.7	3.2	2.1
Temporal lobe	2.0	2.3	3.7	2.7
Parietal lobe	2.0	1.8	3.2	2.3
Basal ganglia	2.5	2.2	3.3	2.7
Other regions				
First cervical vertebra	1.7	1.7	4.2	2.5
Second cervical vertebra	2.5	2.3	3.2	2.7
Optic nerve	2.8	2.5	2.8	2.7
Extraocular muscles	4.0	2.7	3.2	3.3

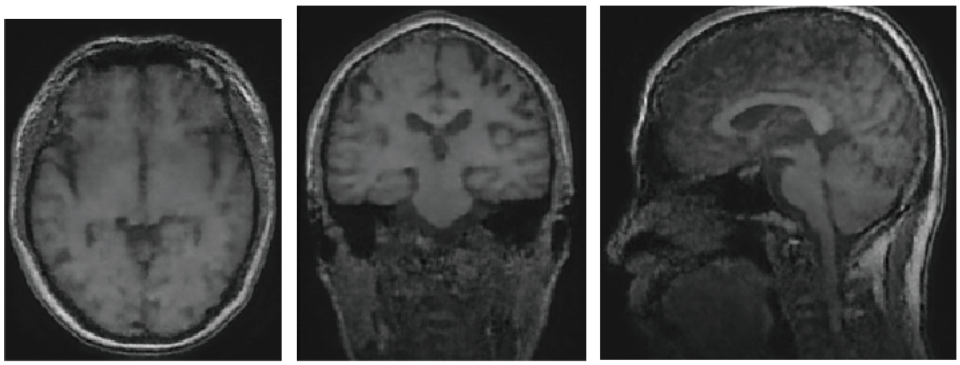

Figure 2. One of the unconditionally generated images (Case 1 in Table A1). (**a**) Axial slice. (**b**) Coronal slice. (**c**) Sagittal slice.

5.2. Conditional Image Generation and Equivalent Privacy Budget

Because we clarified that the model can generate realistic but fictional head MR images in the previous subsection, we now adopt the model to anonymize real head MR images. Specifically, we consider cases with nonzero privacy budgets, i.e., $(\epsilon, \delta) \neq (0, 0)$, for a real head MR image (this image corresponds to $\boldsymbol{x}$ in Section 3.3) and generate anonymized fictional head MR images (these images correspond to postprocessed $\tilde{\boldsymbol{x}}$ in Section 3.3).

Figure 3 shows three representative slices of the real images to be LDP-processed, three representative slices of the LDP-processed images, and three representative slices of the LDP-processed and postprocessed images when the total privacy budgets (In the case of $t = 50$, we have $\sigma^2 = 0.175$. Additionally, the l_2-sensitivity Δf_{ijk} is always 2 because we set $-1 \leq (x_0)_{ijk} \leq 1$. Therefore, for a given $\delta_{ijk} = 10^{-8}$, we have $\epsilon_{ijk} = 2.92 \times 10^1$ per pixel.

Therefore, the total privacy budgets in this case are $\epsilon = 4.90 \times 10^8$ and $\delta = 0.168$ for the image of $256 \times 256 \times 256$ size.) are $\epsilon = 4.90 \times 10^8$ and $\delta = 0.168$.

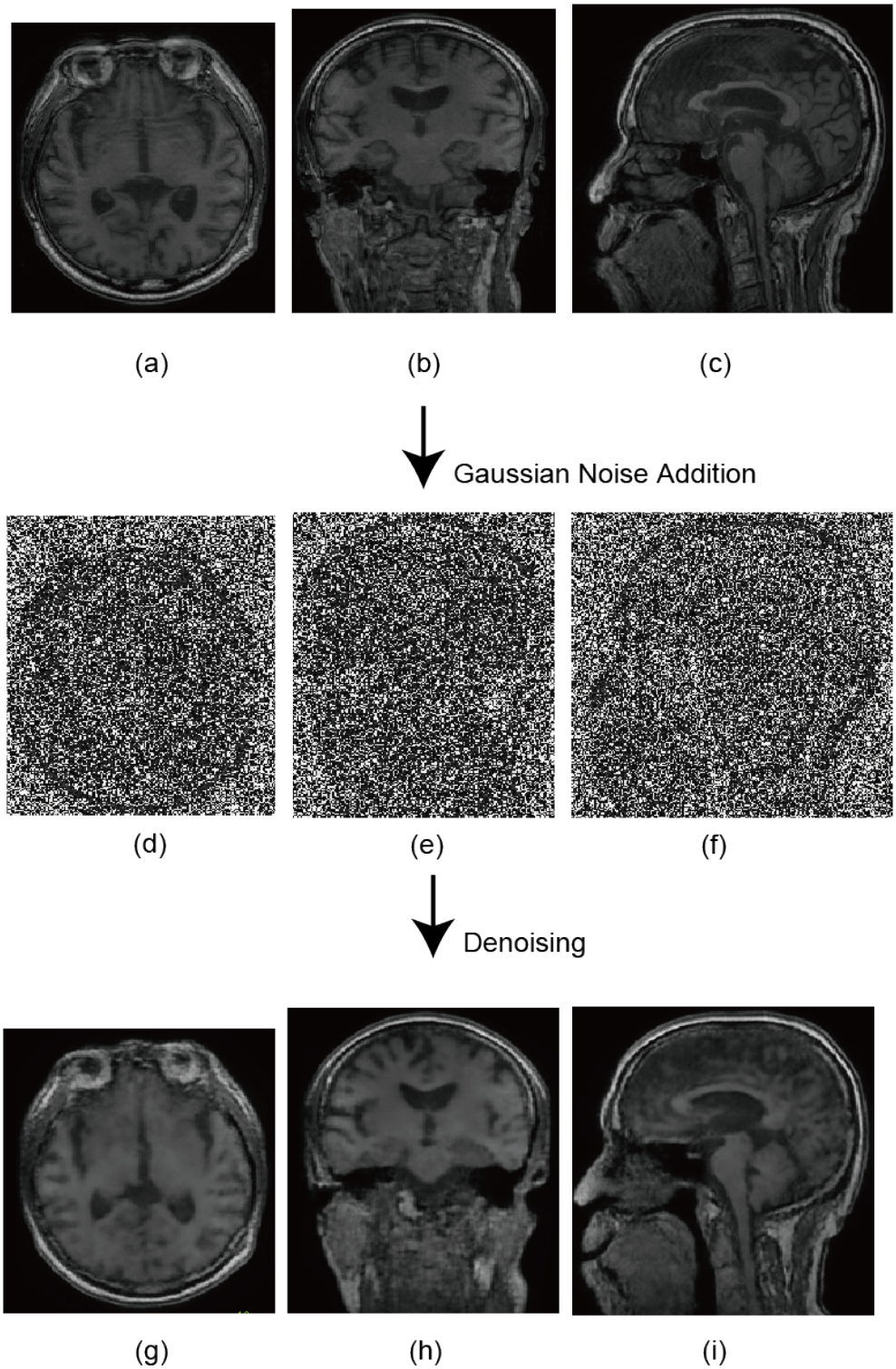

Figure 3. Results of applying Local Differential Privacy (LDP). (**a**) Axial slice of one of the real images. (**b**) Coronal slice of one of the real images. (**c**) Sagittal slice of one of the real image. (**d**) Axial slice of the LDP-processed real image. (**e**) Coronal slice of the LDP-processed real image. (**f**) Sagittal slice of the LDP-processed real image. (**g**) Axial slice of the LDP-processed and postprocessed real image. (**h**) Coronal slice of the LDP-processed and postprocessed real image. (**i**) Sagittal slice of the LDP-processed and postprocessed real image.

6. Discussion

6.1. Novelty

We realized state-of-the-art unconditional and conditional high-resolution volumetric image generation with this improved DDPM. The new architecture with low memory consumption for our proposed DDPM enabled this (see Section 3 and Appendix A for details). Moreover, the novel LDP and postprocessing algorithms that can generate semantically different images in the same class (head MR images) based on the DDPM were proposed and validated.

6.2. Quality Evaluation by Medical Doctors

Global and large structures, e.g., the lateral ventricle, are well unconditionally generated with the proposed DDPM. However, relatively small structures, e.g., the hippocampus and cortical white matter contrast, are not very well generated with the proposed DDPM, as commonly indicated by the three medical doctors in the evaluation results. Nevertheless, most of the anatomically important structures scored more than 2.0 on average. Since in this study a score of 2 means "slightly identifiable", we believe that our proposed method can reconstruct these important structures properly. In the clinical domain, the corticomedullary contrast and the volumes of the hippocampi are important in disease diagnosis and prognosis prediction (e.g., Alzheimer disease [25]). Therefore, we need to improve the depiction correctness of these structures. Note that such a clinically relative evaluation of AI-generated volumes was rarely performed in previous studies and is thus one of our contributions in this paper.

As the difficult structures (corticomedullary contrast and hippocampi) are slightly visualized in the results, we can also apply any additional postprocess to improve the depiction of these structures by, e.g., deep-learning-based filtering methods [26]. Generally, however, simple noise reduction postprocessing cannot recover the structures that are not represented in the original generated volume. In other words, for evaluating the quality of generated medical images, the ordinal contrast-noise ratio (CNR) or structural similarity index measure (SSIM) is not enough and the faithful recreation of relative anatomical structures must be confirmed by medical experts.

On the other hand, the amount of statistical variety of generalized images is another quality measure of generative models and it should be evaluated by further experiments. It is one of our future works. However, it would be difficult to fairly evaluate the amount of variety of medical images generated by a given model, because medical image generation should be not only diverse but also resemble the distribution of a real population of human beings. The ordinal inception score (IS) and Fréchet inception distance (FID) are based on and rely on another classifier model, which is rarely customized for medical images. Inventing a new methodology to evaluate the amount of variety of generated medical images will be another future work.

6.3. Limitation

First, without the model parallelization or high-capacity-memory GPUs, the straightforward scale up of the present model to the resolution of $512 \times 512 \times 512$ is not easy. Second, we showed LDP-processed images with the limited deviations σ_t^2, but we can set arbitrary deviations in theory. This requires the retraining of the diffusion model; otherwise, the model will generate a suboptimal image.

6.4. Future Works

If we increase the number of time steps (T) in the DDPM, it would significantly improve the quality of generated images including the contrast of the cortical white matter, and this is included in our future works. Moreover, we further plan to apply the proposed DDPM to other conditional image generation tasks, e.g., aging prediction, ultra-sparse view CTs [9].

7. Conclusions

We improved the diffusion models so that they can, for the first time, unconditionally generate high-resolution volumetric ($256 \times 256 \times 256$) medical images. Moreover, on the basis of this new diffusion model, we proposed another formulation of image anonymization with which the processed images can satisfy provable Gaussian local differential privacy and we can generate images semantically different from the original image but belonging to the same class. Furthermore, we validated the formulation with high-resolution volumetric medical image anonymization. This method assumes no specific class of images, making it potentially applicable to any type of natural image. Low-resolution medical images are not well-suited for practical medical systems, and traditional standard diffusion models have struggled to model high-resolution volumetric medical images. In this study, we overcame this challenge and paved the way for applying deep generative models to practical medical systems. We believe that this improvement of the DDPM and the formulation of LDP algorithms combined with the DDPM can contribute to the secure sharing of practical images upstream of data processing.

Author Contributions: Conceptualization, H.S.; methodology, H.S.; software, H.S.; validation, H.S.; formal analysis, H.S.; investigation, T.N., T.K. and Y.N. (Yuta Nakamura); resources, T.Y.; data curation, Y.N. (Yukihiro Nomura); writing—original draft preparation, H.S.; writing—review and editing, H.S., S.H., T.N., T.K., Y.N. (Yuta Nakamura), Y.N. (Yukihiro Nomura), T.Y. and O.A.; visualization, H.S.; supervision, T.Y. and O.A.; project administration, H.S.; funding acquisition, H.S. and S.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Japan Science and Technology Agency (JST), CREST Grant Number JPMJCR21M2, including the FY2023 AIP challenge program (Establishment of the diffusion models for high-resolution volumetric images; PI: H. Shibata), Japan.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of the University of Tokyo Hospital (protocol code: 1461-(9) and date of approval: 16 September 2020).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The code presented in this study is available on reasonable request from the corresponding author. The data are not publicly available due to privacy protection.

Acknowledgments: The Department of Computational Diagnostic Radiology and Preventive Medicine, the University of Tokyo Hospital, is sponsored by HIMEDIC Inc., and Siemens Healthcare K.K. This research was conducted using the FUJITSU Supercomputer PRIMEHPC FX1000 and FUJITSU Server PRIMERGY GX2570 (Wisteria/BDEC-01) at the Information Technology Center, the University of Tokyo.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

CT	Computed Tomography
DDPM	Denoising Diffusion Probabilistic Models
GAN	Generative Adversarial Networks
LDP	Local Differential Privacy
MR	Magnetic Resonance
SGD	Stochastic Gradient Descent

Appendix A. Network Architecture of the Proposed DDPM

Figure A1 illustrates the network architecture adopted in this study. We implemented a code on the basis of one of the PyTorch implementations [27] of the pixel space DDPMs for two-dimensional images [3] so that it can handle three-dimensional images and run on eight

NVIDIA A100 GPUs (NVIDIA Corporation, Santa Clara, CA, USA) with 40 GB of memory for each in a data-parallel manner. This code combines the attention mechanism [28], a linear attention mechanism, and U-Net [29] enabling local and global feature extraction and generation. All the `Conv2D` operations were converted into `Conv3D` operations. We adopted the SiLU activation function. The previous network of attention mechanisms was also converted into the equivalent three-dimensional network. We set the number of input image channels to 1. The depth of the three-dimensional U-Net in our code is 8. The Res-Net in different depths of the U-Net has the channel sizes of 4, 8, 16, 32, 64, 128, 256, and 512. For attention mechanisms, we set the dimension of each head to 8 and the number of heads to 4. For linear attention mechanisms, we set the dimension of each head to 8 and the number of heads to 2.

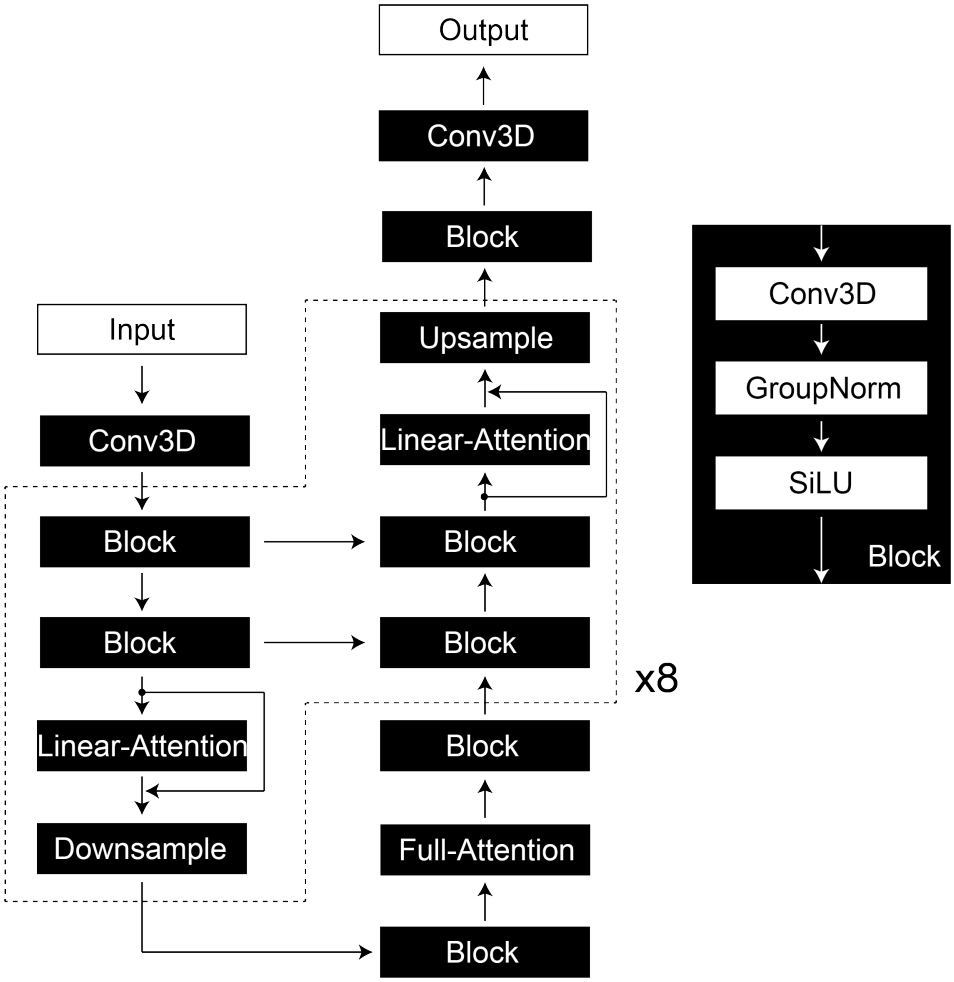

Figure A1. We adopted a UNet that includes attention to model the function f_θ (for details about function f_θ, refer to the main text).

Appendix B. All the Results of Evaluation by Three Medical Doctors

In Tables A1–A3, we show all the evaluation results by the three medical doctors for the six fake cases at resolution of $256 \times 256 \times 256$, respectively.

Table A1. Results of visual evaluation by medical doctor A.

Case (Fake)	1	2	3	4	5	6
Brain regions						
Anterior commissure	4	4	4	4	3	2
Posterior commissure	3	2	2	3	1	2

Table A1. *Cont.*

Case (Fake)	1	2	3	4	5	6
Cerebral aqueduct	2	1	2	2	1	1
Tegmentum of midbrain	3	3	3	3	3	3
Cerebellar hemisphere sulcus	1	2	2	2	2	3
Cerebral peduncle	4	4	4	4	3	3
Corpus callosum	4	4	4	4	4	4
Third ventricle	4	3	3	3	3	3
Fourth ventricle	4	5	4	5	4	5
Lateral ventricle	5	5	5	5	5	5
Cortical white matter contrast						
Hippocampus	1	2	3	1	1	2
Frontal lobe	3	2	3	1	2	2
Occipital lobe	2	2	2	1	1	1
Temporal lobe	2	2	3	1	2	2
Parietal lobe	2	2	2	1	3	2
Basal ganglia	2	3	3	2	3	2
Other regions						
First cervical vertebra (C1)	1	2	1	4	1	1
Second cervical vertebra (C2)	2	2	2	5	2	2
Optic nerve	1	5	3	4	2	2
Extraocular muscles	4	4	5	5	3	3

Table A2. Results of visual evaluation by medical doctor B.

Case (Fake)	1	2	3	4	5	6
Brain regions						
Anterior commissure	3	2	2	1	2	2
Posterior commissure	2	1	2	2	1	1
Cerebral aqueduct	1	1	2	2	1	1
Tegmentum of midbrain	3	3	3	3	3	3
Cerebellar hemisphere sulcus	2	3	2	3	3	3
Cerebral peduncle	3	3	3	3	2	3
Corpus callosum	4	4	4	5	4	4
Third ventricle	5	4	4	4	4	4
Fourth ventricle	5	4	4	4	4	4
Lateral ventricle	5	4	4	4	4	4
Cortical white matter contrast						
Hippocampus	2	2	3	2	3	2
Frontal lobe	3	1	3	1	1	1
Occipital lobe	3	2	2	1	1	1
Temporal lobe	3	2	2	2	3	2
Parietal lobe	3	2	2	1	2	1
Basal ganglia	2	2	3	2	2	2
Other regions						
First cervical vertebra (C1)	1	2	2	2	2	1
Second cervical vertebra (C2)	2	2	3	3	2	2
Optic nerve	1	4	2	4	2	2
Extraocular muscles	3	3	2	4	2	2

Table A3. Results of visual evaluation by medical doctor C.

Case (Fake)	1	2	3	4	5	6
Brain regions						
Anterior commissure	4	4	4	2	3	2
Posterior commissure	2	3	2	3	2	2
Cerebral aqueduct	2	2	2	2	2	3

Table A3. *Cont.*

Case (Fake)	1	2	3	4	5	6
Tegmentum of midbrain	2	2	2	2	3	3
Cerebellar hemisphere sulcus	3	4	4	3	3	3
Cerebral peduncle	5	5	5	5	5	5
Corpus callosum	5	5	5	5	5	5
Third ventricle	5	4	5	5	5	4
Fourth ventricle	5	3	3	3	4	4
Lateral ventricle	5	5	5	5	5	5
Cortical white matter contrast						
Hippocampus	3	3	3	4	3	3
Frontal lobe	2	3	3	3	3	3
Occipital lobe	3	3	3	3	4	3
Temporal lobe	3	4	4	4	4	3
Parietal lobe	3	3	3	3	4	3
Basal ganglia	2	3	3	4	4	4
Other regions						
First cervical vertebra (C1)	3	2	5	5	5	5
Second cervical vertebra (C2)	3	2	3	3	4	4
Optic nerve	2	5	2	4	2	2
Extraocular muscles	4	3	2	4	3	3

References

1. Dwork, C. Differential privacy. In Proceedings of the International Colloquium on Automata, Languages, and Programming ICALP 2006, Venice, Italy, 10–14 July 2006; pp. 1–12.
2. Dwork, C.; Kenthapadi, K.; McSherry, F.; Mironov, I.; Naor, M. Our data, ourselves: Privacy via distributed noise generation. In Proceedings of the Advances in Cryptology-EUROCRYPT 2006: 24th Annual International Conference on the Theory and Applications of Cryptographic Techniques, St. Petersburg, Russia, 28 May–1 June 2006; pp. 486–503.
3. Ho, J.; Jain, A.; Abbeel, P. Denoising diffusion probabilistic models. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 6840–6851.
4. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the Advances in Neural Information Processing Systems 2014, Montreal, QC, Canada, 8–13 December 2014; Volume 27.
5. Croft, W.L.; Sack, J.R.; Shi, W. Differentially private facial obfuscation via generative adversarial networks. *Future Gener. Comput. Syst.* **2022**, *129*, 358–379. [CrossRef]
6. Croft, L.; Sack, J.R.; Shi, W. Obfuscation of images via differential privacy: From facial images to general images. *Peer Netw. Appl.* **2021**, *14*, 1705–1733. [CrossRef]
7. Song, Y.; Sohl-Dickstein, J.; Kingma, D.P.; Kumar, A.; Ermon, S.; Poole, B. Score-Based Generative Modeling through Stochastic Differential Equations. In Proceedings of the International Conference on Learning Representations, Addis Ababa, Ethiopia, 26–30 April 2020.
8. Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; Ommer, B. High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 10684–10695.
9. Shibata, H.; Hanaoka, S.; Nomura, Y.; Nakao, T.; Takenaga, T.; Hayashi, N.; Abe, O. On the Simulation of Ultra-Sparse-View and Ultra-Low-Dose Computed Tomography with Maximum a Posteriori Reconstruction Using a Progressive Flow-Based Deep Generative Model. *Tomography* **2022**, *8*, 2129–2152. [CrossRef] [PubMed]
10. Dinh, L.; Krueger, D.; Bengio, Y. Nice: Non-linear independent components estimation. *arXiv* **2014**, arXiv:1410.8516.
11. Dinh, L.; Sohl-Dickstein, J.; Bengio, S. Density estimation using real nvp. *arXiv* **2016**, arXiv:1605.08803.
12. Kingma, D.P.; Dhariwal, P. Glow: Generative flow with invertible 1x1 convolutions. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 10236–10245.
13. Khader, F.; Müller-Franzes, G.; Tayebi Arasteh, S.; Han, T.; Haarburger, C.; Schulze-Hagen, M.; Schad, P.; Engelhardt, S.; Baeßler, B.; Foersch, S. Denoising diffusion probabilistic models for 3D medical image generation. *Sci. Rep.* **2023**, *13*, 7303. [CrossRef] [PubMed]
14. Bieder, F.; Wolleb, J.; Durrer, A.; Sandkuehler, R.; Cattin, P.C. Memory-Efficient 3D Denoising Diffusion Models for Medical Image Processing. In Proceedings of the Medical Imaging with Deep Learning, Nashville, TN, USA, 10–12 July 2023.
15. Dorjsembe, Z.; Odonchimed, S.; Xiao, F. Three-dimensional medical image synthesis with denoising diffusion probabilistic models. In Proceedings of the Medical Imaging with Deep Learning, Zurich, Switzerland, 6–8 July 2022.
16. Sun, L.; Chen, J.; Xu, Y.; Gong, M.; Yu, K.; Batmanghelich, K. Hierarchical amortized GAN for 3D high resolution medical image synthesis. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 3966–3975. [CrossRef] [PubMed]
17. Fan, L. Image pixelization with differential privacy. In Proceedings of the IFIP Annual Conference on Data and Applications Security and Privacy, Bergamo, Italy, 16–18 July 2018; pp. 148–162.

18. Li, T.; Clifton, C. Differentially private imaging via latent space manipulation. *arXiv* **2021**, arXiv:2103.05472.
19. Liu, B.; Ding, M.; Xue, H.; Zhu, T.; Ye, D.; Song, L.; Zhou, W. DP-Image: Differential Privacy for Image Data in Feature Space. *arXiv* **2021**, arXiv:2103.07073.
20. Abadi, M.; Chu, A.; Goodfellow, I.; McMahan, H.B.; Mironov, I.; Talwar, K.; Zhang, L. Deep learning with differential privacy. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, 24–28 October 2016; pp. 308–318.
21. Song, Y.; Ermon, S. Generative modeling by estimating gradients of the data distribution. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; Volume 32.
22. Jabri, A.; Fleet, D.J.; Chen, T. Scalable Adaptive Computation for Iterative Generation. In Proceedings of the International Conference on Machine Learning, PMLR, Honolulu, HI, USA, 23–29 July 2023; pp. 14569–14589.
23. Dwork, C. Differential privacy: A survey of results. In Proceedings of the International Conference on Theory and Applications of Models of Computation, Xi'an, China, 25–29 April 2008; pp. 1–19.
24. Fujita, S.; Mori, S.; Onda, K.; Hanaoka, S.; Nomura, Y.; Nakao, T.; Yoshikawa, T.; Takao, H.; Hayashi, N.; Abe, O. Characterization of Brain Volume Changes in Aging Individuals With Normal Cognition Using Serial Magnetic Resonance Imaging. *JAMA Netw. Open* **2023**, *6*, e2318153. [CrossRef] [PubMed]
25. Leandrou, S.; Petroudi, S.; Kyriacou, P.A.; Reyes-Aldasoro, C.C.; Pattichis, C.S. Quantitative MRI brain studies in mild cognitive impairment and Alzheimer's disease: A methodological review. *IEEE Rev. Biomed. Eng.* **2018**, *11*, 97–111. [CrossRef] [PubMed]
26. Chen, Z.; Pawar, K.; Ekanayake, M.; Pain, C.; Zhong, S.; Egan, G.F. Deep learning for image enhancement and correction in magnetic resonance imaging—State-of-the-art and challenges. *J. Digit. Imaging* **2023**, *36*, 204–230. [CrossRef] [PubMed]
27. Denoising Diffusion Probabilistic Model, in PyTorch. Available online: https://github.com/lucidrains/denoising-diffusion-pytorch/releases/tag/1.8.5 (accessed on 12 March 2024).
28. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems 2017, Long Beach, CA, USA, 4–9 December 2017; Volume 30.
29. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; pp. 234–241.