Improving 3D Imaging with Pre-Trained Perpendicular 2D Diffusion Models

Suhyeon Lee[1*], Hyungjin Chung[1*], Minyoung Park[2], Jonghyuk Park[2], Wi-Sun Ryu[2], Jong Chul Ye[1]
[1]Korea Advanced Institute of Science & Technology, [2]JLK Inc.,
{suhyeon.lee, hj.chung, jong.ye}@kaist.ac.kr, {mypark, jhpark01, wsryu}@jlkgroup.com

Abstract

Diffusion models have become a popular approach for image generation and reconstruction due to their numerous advantages. However, most diffusion-based inverse problem-solving methods only deal with 2D images, and even recently published 3D methods do not fully exploit the 3D distribution prior. To address this, we propose a novel approach using two perpendicular pre-trained 2D diffusion models to solve the 3D inverse problem. By modeling the 3D data distribution as a product of 2D distributions sliced in different directions, our method effectively addresses the curse of dimensionality. Our experimental results demonstrate that our method is highly effective for 3D medical image reconstruction tasks, including MRI Z-axis super-resolution, compressed sensing MRI, and sparse-view CT. Our method can generate high-quality voxel volumes suitable for medical applications. The code is available at https://github.com/hyn2028/tpdm

1. Introduction

The diffusion probabilistic model (DPM) uses neural networks to learn the gradient of the log probability distribution, $\nabla_{\boldsymbol{x}} \log p_{data}(\boldsymbol{x})$, also known as the score function. Sampling is done by either using Langevin dynamics [38] or solving the reverse stochastic differential equation (SDE) using the learned score function [40].

DPM has emerged as a leading generative model in the image field since its introduction [37, 15, 40], surpassing other models like GAN in achieving state-of-the-art performance [10, 30]. The confluence of the diffusion model with conditioning training is a noteworthy synergy, constituting a foundational framework within the domain of text-guided image generation [30, 32, 29]. Furthermore, its versatile applicability extends to novel realms like brain vision decoding [3, 41]. It is also being explored as a generative model in other various areas such as audio [28, 22, 18], video [2, 36, 26], radiance field [27, 35], and graph [43, 17].

Despite the slow sampling speed due to sequential sampling over multiple time steps, diffusion models offer significant advantages over other generative models, including sampling-time scalability. The pre-trained score function model can be used for conditional sampling without retraining, thanks to Bayes' theorem [10, 16]. This conditional sampling-based inverse problem-solving method can be interpreted as posterior sampling with diffusion generative priors. Thus, it effectively avoids bias and regression to the mean phenomena from the supervised likelihood optimization methods. In due course, the paradigm of diffusion-based inverse problem-solving methodology [40, 20, 7, 5, 4, 39, 6, 45] has risen to the forefront as a state-of-the-art technique within the realm of study.

Most contemporary diffusion-based inverse problem-solving methods are focused on 2D applications. However, a recent method called DiffusionMBIR [6] has been proposed to address 3D inverse problems in medical imaging. In DiffusionMBIR, the diffusion model trained on the primary XY-plane is used as the prior, and the generative prior is augmented with a model-based prior, namely total variation (TV), to enforce smoothness to the adjacent slices (Z-axis). While this approach has been effective for various tasks, it still has limitations because it does not fully learn the 3D prior distribution of the data. More specifically, the TV prior only imposes local dependencies that are derived from finite difference operators, whereas the true 3D prior should model global dependencies.

To overcome this limitation, we propose a new method called *Two Perpendicular 2D Diffusion Models (TPDM) for 3D generation*. TPDM fully leverages the 3D generative prior by modeling the 3D data distribution with a product distribution of 2D constituents, without relying on a model-based prior. This approach allows TPDM to effectively learn the 3D prior using only two 2D diffusion models: the primary model that operates on the XY-plane and an auxiliary model that learns the YZ-plane. Unlike the previous DiffusionMBIR approach, TPDM can model the global dependencies of the 3D structure, and it eliminates the need for sub-optimization schemes required to impose the TV constraint. It is worth mentioning that, unlike Diffu-

[*]These authors contributed equally to this work

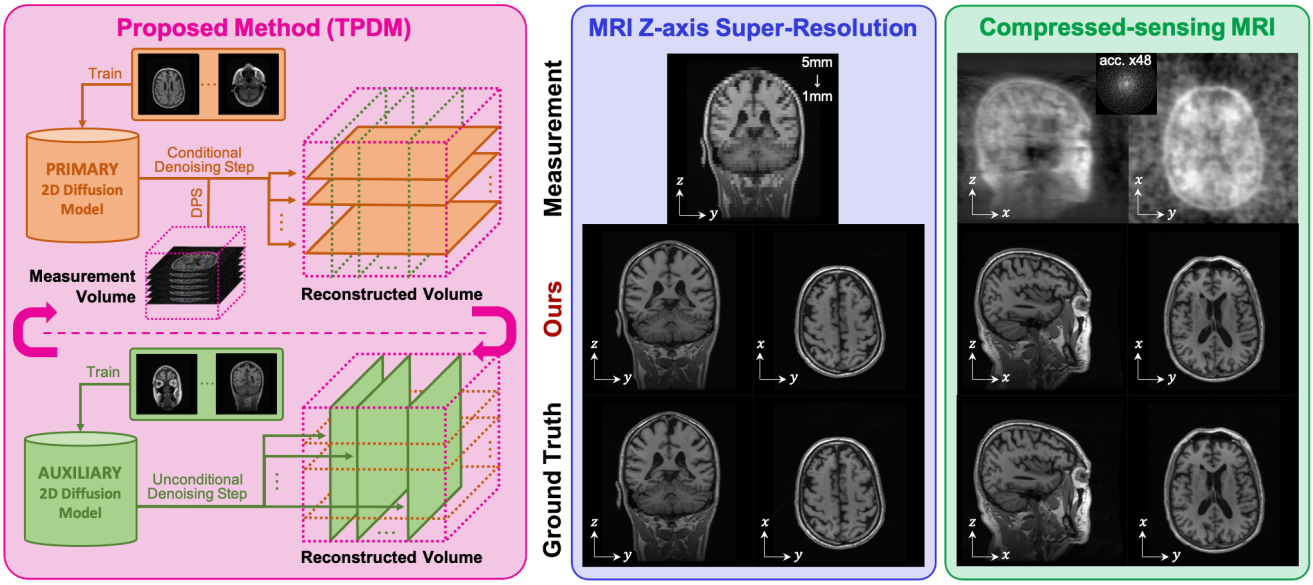

Figure 1. (Left) A visualization of our proposed method. (Right) We display the results of solving the 3D inverse problem using the proposed method, with MR-ZSR and CS-MRI techniques shown in the center and right panels, respectively. The first row shows the measurements, the second row displays the output from our proposed method, and the third row presents the ground truth. In the MR-ZSR approach, the slice thickness was improved from 5mm to 1mm using super-resolution techniques. In the CS-MRI approach, Poisson sub-sampling was used to accelerate the process by a factor of 48.

sionMBIR which is designed specifically for inverse problem solving, TDPM is a *fully general 3D generative model*, which can be used both for conditional and unconditional sampling.

In this paper, TPDM has been tested in various 3D medical imaging reconstruction problems such as MRI Z-axis (*i.e.* vertical axis) super-resolution (MR-ZSR), compressed sensing MRI (CS-MRI), and sparse view CT (SV-CT) and has produced the state-of-the-art results compared to existing methods. Especially, to the best of our knowledge, we have achieved the first successful attempt at a diffusion model-based MR-ZSR both technically and clinically (Fig. 2). We also demonstrated that TPDM can generate a very high-quality, complete 3D voxels volume as a pure generative model (Fig. 7). Our contributions can be summarized as follows.

1. We developed a novel, simple, yet effective method to solve the 3D volume inverse problem with two perpendicular 2D diffusion models as a 3D prior, in a fully unsupervised manner, without the need for re-training.
2. We applied it to various medical imaging reconstruction problems and achieved the best-known performance. In particular, TPDM succeeded in the first attempt at a diffusion model-based MR-ZSR.
3. Finally, we demonstrated that TPDM can also function as a 3D generative model, generating high-quality 3D voxel volumes.

2. Background

2.1. Score-based diffusion models

The diffusion model [37, 15, 40] is a model family that defines a process that noises the original data gradually, called a forward process, and expresses the generation process by performing the learned reverse process of this noising process. Among them, the score-based diffusion model introduced by Song *et al.* [40] defines the forward process through the following Itô stochastic differential equation (SDE). Throughout the diffusion process, the data $\boldsymbol{x}$ can be represented by $\boldsymbol{x}(t) = \boldsymbol{x}_t$, with continuous time index $t \in [0, 1]$. $\boldsymbol{x}_0 \sim p_{data}$ is the raw data distribution, and $\boldsymbol{x}_1 \sim p_0$ is the predefined prior distribution.

$$d\boldsymbol{x} = \boldsymbol{f}(\boldsymbol{x}, t)dt + g(t)d\boldsymbol{w}, \tag{1}$$

where the function $\boldsymbol{f} : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$ is the drift function, and the function $g : \mathbb{R} \to \mathbb{R}$ is the diffusion coefficient. $\boldsymbol{w}$ is the standard Wiener process, also called Brownian motion. The reverse-time SDE of Eq. (1) can be expressed as follows [1, 40]:

$$d\boldsymbol{x} = [\boldsymbol{f}(\boldsymbol{x}, t) - g(t)^2 \nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{x}_t)]dt + g(t)d\bar{\boldsymbol{w}} \tag{2}$$

where $\bar{\boldsymbol{w}}$ is also the standard Wiener process.

In order to solve the reverse-time SDE for the generation process, a time-dependent score function $\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{x}_t)$ is required, which can be obtained by training the neural

network-based score function estimator $\boldsymbol{s}_{\boldsymbol{\theta}}$ through the denoising score matching (DSM) objective [44, 40]

$$\min_{\boldsymbol{\theta}} \mathbb{E}_{\boldsymbol{x}_t|\boldsymbol{x}_0,\boldsymbol{x}_0}[\|\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}(t),t) - \nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{x}_t|\boldsymbol{x}_0)\|_2^2] \quad (3)$$

Setting $\boldsymbol{f}(\boldsymbol{x},t) = 0$ and $g(t) = \sqrt{\frac{d[\sigma^2(t)]}{dt}}$ with the positive time-dependent increasing noise scale function $\sigma(t)$, we achieve the so-called variance exploding SDE (VE-SDE). The sampling process of VE-SDE can be effectively solved by replacing the score function with the score network which is trained by the DSM objective.

2.2. Diffusion posterior sampling

Diffusion posterior sampling (DPS) is one of the state-of-the-art methods to solve the general noisy inverse problem introduced by Chung *et al*. [5] by using the diffusion model as a prior. Consider a general forward model of the inverse problem can be defined as:

$$\boldsymbol{y} = \boldsymbol{A}(\boldsymbol{x}_0) + \boldsymbol{n}, \qquad \boldsymbol{y}, \boldsymbol{n} \in \mathbb{R}^n, \boldsymbol{x} \in \mathbb{R}^d \quad (4)$$

where $\boldsymbol{A}$ is the forward measurement function and $\boldsymbol{n}$ is the measurement noise. To solve the inverse problem using the diffusion prior, we can use Bayes' rule to obtain

$$\begin{aligned}\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{x}_t|\boldsymbol{y}) &= \nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{x}_t) + \nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{y}|\boldsymbol{x}_t) \\ &\simeq \boldsymbol{s}_{\boldsymbol{\theta}^*}(\boldsymbol{x}_t, t) + \nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{y}|\boldsymbol{x}_t). \quad (5)\end{aligned}$$

Nonetheless, since there is no explicit relationship between $\boldsymbol{x}_t$ and $\boldsymbol{y}$, we cannot use (5) directly. To circumvent this problem, [5] proposes an approximation with a theoretically guaranteed upper bound on the approximation error

$$\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{y}|\boldsymbol{x}_t) \simeq \nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{y}|\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t)), \quad (6)$$

where

$$\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t) \coloneqq \mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t] = \boldsymbol{x}_t + \sigma^2(t)\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{x}_t) \quad (7)$$

is the Tweedie denoised estimate [11, 21]. Accordingly, when the measurement noise is Gaussian, one can use:

$$\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{x}_t|\boldsymbol{y}) \simeq \boldsymbol{s}_{\boldsymbol{\theta}^*}(\boldsymbol{x}_t, t) - \lambda\nabla_{\boldsymbol{x}_t}\|\boldsymbol{A}(\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t)) - \boldsymbol{y}\|_2^2. \quad (8)$$

3. Two Perpendicular 2D Diffusion Model

3.1. Modeling data distribution

To overcome the drawbacks of DiffusionMBIR, here we describe our method of applying priors that is closer to the actual 3D distribution than DiffusionMBIR. Our simple yet effective solution is, by modeling the 3D data distribution as the product distribution, to additionally use an auxiliary diffusion model trained on 2D slices in different directions of the volume, in addition to the primary 2D diffusion model to solve the inverse problem (Fig. 1). This allows us to effectively drive a diffusion model in high dimensional space, much like the utilization of factorization methods in diverse deep learning scenarios for efficiency [33, 13, 12].

Specifically, our proposal is to model the data distribution as the *product* distribution given by

$$p_{\boldsymbol{\theta},\boldsymbol{\phi}}(\boldsymbol{x}) = q_{\boldsymbol{\theta}}^{(p)}(\boldsymbol{x})^{\alpha} q_{\boldsymbol{\phi}}^{(a)}(\boldsymbol{x})^{\beta}/Z \quad (9)$$

$$\begin{aligned}&= [q_{\boldsymbol{\theta}}^{(p)}(\boldsymbol{x}_{[:,:,1]}) q_{\boldsymbol{\theta}}^{(p)}(\boldsymbol{x}_{[:,:,2]}) \cdots q_{\boldsymbol{\theta}}^{(p)}(\boldsymbol{x}_{[:,:,d_3]})]^{\alpha} \\ &\times [q_{\boldsymbol{\phi}}^{(a)}(\boldsymbol{x}_{[1,:,:]}) q_{\boldsymbol{\phi}}^{(a)}(\boldsymbol{x}_{[2,:,:]}) \cdots q_{\boldsymbol{\phi}}^{(a)}(\boldsymbol{x}_{[d_1,:,:]})]^{\beta}/Z, \quad (10)\end{aligned}$$

where Z is an appropriate normalizing partition function, $q_{\boldsymbol{\theta}}^{(p)}(\boldsymbol{x})$ is the distribution modeled by the primary model parameterized with $\boldsymbol{\theta}$, and $q_{\boldsymbol{\phi}}^{(a)}(\boldsymbol{x})$ is the distribution modeled by the auxiliary model parameterized with $\boldsymbol{\phi}$, for $\boldsymbol{x} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$. Moreover, α, β induces weighting between the two distributions according to the importance. We further assume that both $q_{\boldsymbol{\theta}}^{(p)}$ and $q_{\boldsymbol{\phi}}^{(a)}$ can be decomposed into independent 2D (slice) distributions.

Accordingly, when performing unconditional sampling from the prior distribution $p_{\boldsymbol{\theta},\boldsymbol{\phi}}(\boldsymbol{x})$, we can directly use

$$\begin{aligned}&\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{x}_t) = \alpha\nabla_{\boldsymbol{x}_t} \log q^{(p)}(\boldsymbol{x}_t) + \beta\nabla_{\boldsymbol{x}_t} \log q^{(a)}(\boldsymbol{x}_t) \\ &= \alpha\Sigma_{i=1}^{d_3}\nabla_{\boldsymbol{x}_t} \log q^{(p)}(\boldsymbol{x}_{t,[:,:,i]}) + \beta\Sigma_{i=1}^{d_1}\nabla_{\boldsymbol{x}_t} \log q^{(a)}(\boldsymbol{x}_{t,[i,:,:]}) \\ &\simeq \alpha\Sigma_{i=1}^{d_3}\boldsymbol{s}_{\boldsymbol{\theta}^*}^{3D}(\boldsymbol{x}_{t,[:,:,i]}) + \beta\Sigma_{i=1}^{d_1}\boldsymbol{s}_{\boldsymbol{\phi}^*}^{3D}(\boldsymbol{x}_{t,[i,:,:]}), \quad (11)\end{aligned}$$

where $\boldsymbol{x}_{t,[i,:,:]}$ and $\boldsymbol{x}_{t,[:,:,j]}$ denote the i and j-th x- and z-slice of $\boldsymbol{x}_t$, respectively, and

$$\begin{cases}\boldsymbol{s}^{3D}(\boldsymbol{x}_{t,[:,:,i]})_{[:,:,i]} = \boldsymbol{s}(\boldsymbol{x}_{t,[:,:,i]}) \\ \boldsymbol{s}^{3D}(\boldsymbol{x}_{t,[:,:,i]})_{[\text{otherwise}]} = 0\end{cases} \quad (12)$$

$$\begin{cases}\boldsymbol{s}^{3D}(\boldsymbol{x}_{t,[i,:,:]})_{[i,:,:]} = \boldsymbol{s}(\boldsymbol{x}_{t,[i,:,:]}) \\ \boldsymbol{s}^{3D}(\boldsymbol{x}_{t,[i,:,:]})_{[\text{otherwise}]} = 0\end{cases} \quad (13)$$

which used the trained 2D score estimator $\boldsymbol{s}(\cdot)$ due to our 2D slice independence assumption. However, care must be taken since simply using this approximation would be compute-heavy, as one would have to evaluate two forward passes *per* each iteration. In this regard, we propose a simple fix to this problem by using alternating updates

$$\begin{cases}\Sigma\boldsymbol{s}_{\boldsymbol{\theta}^*}^{3D}(\boldsymbol{x}_{t,[:,:,i]}), & \text{with } \mathbb{P} = \alpha/(\alpha+\beta) \\ \Sigma\boldsymbol{s}_{\boldsymbol{\phi}^*}^{3D}(\boldsymbol{x}_{t,[i,:,:]}), & \text{with } \mathbb{P} = \beta/(\alpha+\beta)\end{cases} \quad (14)$$

where $\mathbb{P}$ denotes the probability of each step to be performed. (14) can be implemented in regularly structured intervals or in a stochastic fashion, which we discuss in detail in Section 3.2.

Finally, in order to solve the inverse problem, we can leverage the following result:

$$\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{x}_t|\boldsymbol{y}) \simeq \alpha \nabla_{\boldsymbol{x}_t} \log q^{(p)}(\boldsymbol{x}_t) + \beta \nabla_{\boldsymbol{x}_t} \log q^{(a)}(\boldsymbol{x}_t) + \nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{y}|\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t)), \tag{15}$$

where is simplified similar to unconditional sampling in (14) as

$$\begin{cases} \Sigma \boldsymbol{s}^{3D}_{\boldsymbol{\theta}^*}(\boldsymbol{x}_{t,[:,:,i]}) + \gamma_t \nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{y}|\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t)), & \text{with } \mathbb{P} = \alpha/(\alpha+\beta) \\ \Sigma \boldsymbol{s}^{3D}_{\boldsymbol{\phi}^*}(\boldsymbol{x}_{t,[i,:,:]}), & \text{with } \mathbb{P} = \beta/(\alpha+\beta) \end{cases} \tag{16}$$

where γ_t is the step size that also absorbs the weighting factor induced by α and β.

3.2. Solving 3D reconstruction problem with TPDM

Training of TPDM is performed by training the primary and auxiliary 2D diffusion model (for the algorithm, see Appendix A.1). The primary 2D diffusion model $\boldsymbol{s}_{\boldsymbol{\theta}^*}$ selects an appropriate plane when solving the inverse problem and is trained with sliced images of 3D volumes into the corresponding plane. For example, in the case of CS-MRI and SV-CT, it is the axial plane, and in the case of MR-ZSR, it is the sagittal or coronal plane. An auxiliary 2D diffusion model $\boldsymbol{s}_{\boldsymbol{\phi}^*}$ is trained by selecting one of the two remaining planes of the volumes.

In order to solve the inverse problem, conditional sampling is performed alternately using the trained TPDM for each step of one time-step denoising (Algorithm 1). While the algorithms are presented individually for clarity, they can be batched for computational efficiency. In each denoising step, we use the primary diffusion model $\boldsymbol{s}_{\boldsymbol{\theta}^*}$ to constrain the consistency of measurements $\boldsymbol{y}$ and sample an image using the DPS [5]. The hyperparameter λ controls the strength of the measurement consistency. The auxiliary diffusion model $\boldsymbol{s}_{\boldsymbol{\phi}^*}$ is used to correct inconsistencies in the batch direction caused by the primary diffusion model. We adjust the contribution of the two models using the integer hyperparameter K (for non-integer values of K, see Appendix A.2). For example, if K=4, the primary model and the auxiliary model contribute to image generation at a ratio of 3:1, respectively.

Algorithm 1 Solving 3D Inverse Problem with TPDM

Require: $\boldsymbol{Y} \in \mathbb{N}^{d_1' \times d_2' \times d_3}$, $\boldsymbol{A}(\cdot) : \mathbb{N}^{d_1 \times d_2} \to \mathbb{N}^{d_1' \times d_2'}$, $\boldsymbol{s}_{\boldsymbol{\theta}^*}$, $\boldsymbol{s}_{\boldsymbol{\phi}^*}$, $\{\sigma_i\}_0^1$, N, K, λ

$\boldsymbol{X}_N \sim \mathcal{N}(\mathbf{0}, \sigma_1^2 \boldsymbol{I}) \in \mathbb{N}^{d_1 \times d_2 \times d_3}$
for i **in** $N-1:0$ **do**
  $t \leftarrow \frac{i}{N}$
  $\boldsymbol{X}_i \leftarrow$ torch.empty_like($\boldsymbol{X}_N$)
  if $\mod(i, K) \neq 0$ **then**
    for j **in** $1:d_3$ **do**
      $\boldsymbol{x} \leftarrow \boldsymbol{X}_{i+1}[:,:,j]$
      $\boldsymbol{y} \leftarrow \boldsymbol{Y}[:,:,j]$
      $\hat{\boldsymbol{x}}_0 \leftarrow \boldsymbol{x} + \sigma_t^2 \cdot \boldsymbol{s}_{\boldsymbol{\theta}^*}(\boldsymbol{x}, t)$
      $\boldsymbol{x}' \leftarrow$ step_2D_DPM($\boldsymbol{x}$, $\boldsymbol{s}_{\boldsymbol{\theta}^*}$, σ_t, t)
      $\boldsymbol{x}'' \leftarrow \boldsymbol{x}' - \lambda \nabla_{\boldsymbol{x}} \|\boldsymbol{A}(\hat{\boldsymbol{x}}_0) - \boldsymbol{y}\|_2^2$
      $\boldsymbol{X}_i[:,:,j] \leftarrow \boldsymbol{x}''$
    end for
  else
    for j **in** $1:d_1$ **do**
      $\boldsymbol{x} \leftarrow \boldsymbol{X}_{i+1}[j,:,:]$
      $\boldsymbol{x}' \leftarrow$ step_2D_DPM($\boldsymbol{x}$, $\boldsymbol{s}_{\boldsymbol{\phi}^*}$, σ_t, t)
      $\boldsymbol{X}_i[j,:,:] \leftarrow \boldsymbol{x}'$
    end for
  end if
end for
return $\boldsymbol{X}_0$

4. Methods

In this paper, we investigate various applications of TPDM, which include medical domain inverse problems such as 1) MRI Z-axis (vertical axis) super-resolution (MR-ZSR), 2) compressed sensing MRI (CS-MRI), and 3) sparse view CT (SV-CT). In addition to solving the 3D inverse problem that applies conditioned sampling of the diffusion model, 4) TPDM is also used to generate unconditioned high-fidelity 3D voxels volumes in Brain MRI.

4.1. Dataset

The MR-ZSR, CS-MRI, and the 3D volume voxel generation task used our IRB-approved in-house brain MRI image dataset (*i.e.* `BMR-ZSR-1mm` and `BMR-ZSR-5mm`). For detailed information, see Appendix B.1. All volumes are in the shape of a 256×256×256 cube and standard 3T T1-weighted images. `BMR-ZSR-1mm`, which is used for training and retrospective evaluation, has a 1mm slice thickness. 923 volumes (236,288 2D images) were used as a training dataset, and 1 volume was used as a test dataset with the retrospective slice thickness degradation or the CS-MRI subsampling simulation. `BMR-ZSR-5mm`, which is a prospective dataset acquired at a slice thickness of 5mm, was used for the prospective clinical evaluation of MR-ZSR.

SV-CT task used the public CT dataset provided in the AAPM 2016 CT low-dose grand challenge [25]. The dataset consists of a total of 10 volumes of contrast-enhanced abdominal CT. To make the volume a $256 \times 256 \times 256$ cube, we resized the XY-plane to 256×256 and cropped the common part in the Z-direction to make the length 256 (*i.e.* `LDCT-CUBE` Dataset). One of the 10 volumes was used as the test dataset with the retrospective measurement simulation, and the remaining 9 were used as the training dataset. The data we used for training was only 2304 2D im-

ages so that we can demonstrate reliable performance even when the training data was small.

4.2. Measurement model for inverse problems

MR-ZSR. The goal of this task is to perform super-resolution of a 5mm slice thickness MRI image to 1mm slice resolution for quantitative brain MR analysis such as cortical thickness measurement. Considering the slice selection process of MRI, the forward measurement kernel can be modeled by combining adjacent voxels in the Z-axis direction with averaging operation. For example, for 5mm to 1mm slice super-resolution, the forward model is an operation of degrading a 1mm slice image by grouping 5 adjacent XY-plane along the Z-axis direction and averaging each group to get a 5mm slice image. Here, we defined the number of pixels in the Z-axis direction of a group to be merged as *merge size* (M).

We used the forward kernel just presented when creating the retrospective degraded MRI dataset (1mm $\rightarrow$ 2, 5mm). We also used a slightly different forward measurement kernel used in the DPS step when solving the inverse problem MR-ZSR. The kernel is similar to the averaging process, but when averaging, divide by $\sqrt{M}$ instead of dividing by M, which is inspired by Song *et al*. [40] and Chung *et al*. [7]'s diffusion model-based image colorization method.

CS-MRI and SV-CT. The forward measurement kernel for compressive sensing MRI (CS-MRI) involves applying a 2D subsampling mask to each slice of the image after transforming it into a k-space using a 2D Fourier transform. The resulting measurement $\boldsymbol{y}$ is given in the k-space domain. In the case of sparse view CT (SV-CT), the forward measurement kernel is determined by the sparse view CT acquisition scenario, where angular projection views are subsampled at a sparse set of angles. The measurements are given in sinogram space.

4.3. DPM training and sampling

Both the MRI model and the CT model were trained and inferred under the common model setup and algorithms. The 2D image diffusion model constituting the TPDM used `ncsnpp` [40] using VE-SDE which is scheduled by a geometric sequence σ_0=0.01 to σ_1=378. All inputs were normalized between 0 to 1. In the MR-ZSR problem, the YZ-plane (coronal) was used for the primary model and the XY-plane (axial) was used for the auxiliary model. In all other problems, the XY-plane (axial) was used for the primary model and the YZ plane (coronal) for the auxiliary model. The training was conducted with batch size 8, and the MRI model and CT model performed 300K and 100K training iterations, respectively. For the sampling stage, N=2000 and predictor-corrector sampling [40] method were employed.

4.4. Comparison methods and evaluation

For the 3D medical inverse problem, our method was compared with DiffusionMBIR [6], DPS [5], MCG [7], score-MRI [8], score-CT [39], L1-Wavelet [24], FBPConvNet [19] and ADMM-TV. DiffusionMBIR is the state-of-the-art method to solve general 3D inverse problems which outperformed existing methods such as Score-MRI, DuDoRNet [47], U-Net [31] and Zero-filled in CS-MRI and outperformed previous methods such as MCG, Lahiri *et al*. [23], FBPConvNet, and ADMM-TV in SV-CT. As the MR-ZSR problem is a new endeavor, no diffusion-based method has been devised to address it specifically. Quantitative evaluation was performed using peak-signal-to-noise-ratio (PSNR) and structural similarity index measure (SSIM) [46] for the retrospective test dataset. PSNR was evaluated in a 3D volume, and SSIM measured the average value of results of 2D slices for each slice direction (axial, coronal, and sagittal).

For the evaluation of MR-ZSR's clinical implications, seven patients with ischemic stroke were included in the evaluation (`BMR-ZSR-5mm`). Visual assessments of cortical atrophy and white matter hyperintensity were conducted using the Global Cortical Atrophy scale [14] and Fazekas grade [34], respectively. Using the TDPM, the prospective standard T1-weighted images with a 5 mm thickness were reconstructed into 1mm images. Five out of seven patients had 3D volumetric 1 mm T1-weighted images acquired simultaneously with a 5 mm T1-weighted image. The mean cortical thickness obtained with an upscaled T1-weighted image was compared to the mean cortical thickness measured with a 1mm raw T1-weighted image as the ground truth. Using FreeSurfer [42] and ATROSCAN (JLK Inc., Seoul, Republic of Korea) based on Swin U-net [9], the cortical thickness was measured.

5. Experimental Results

5.1. MRI Z-axis super-resolution (MR-ZSR)

We first conducted MRI Z-axis $\times$5 super-resolution images of 5mm slices, which are mainly taken in clinical practice, to 1mm with the retrospective 5mm test dataset, and the results are in the Table 1 and Fig. 2. For another merge size, see Appendix C.1. MR-ZSR using TPDM showed quantitatively better results than any other diffusion-based 2D/3D inverse problem-solving methods [6, 5, 7], and no artifacts occurred in any slice direction of the volume. In addition, the use of the auxiliary model not only improves the quality of the slice in the auxiliary direction but also has the effect of improving the detail of the entire slice directions (see (c) and (d) of the Fig. 2).

Notably, DiffusionMBIR [6], which is known to be the highest-performing general linear 3D inverse problem solver, did not work at all for our custom-designed MR-ZSR

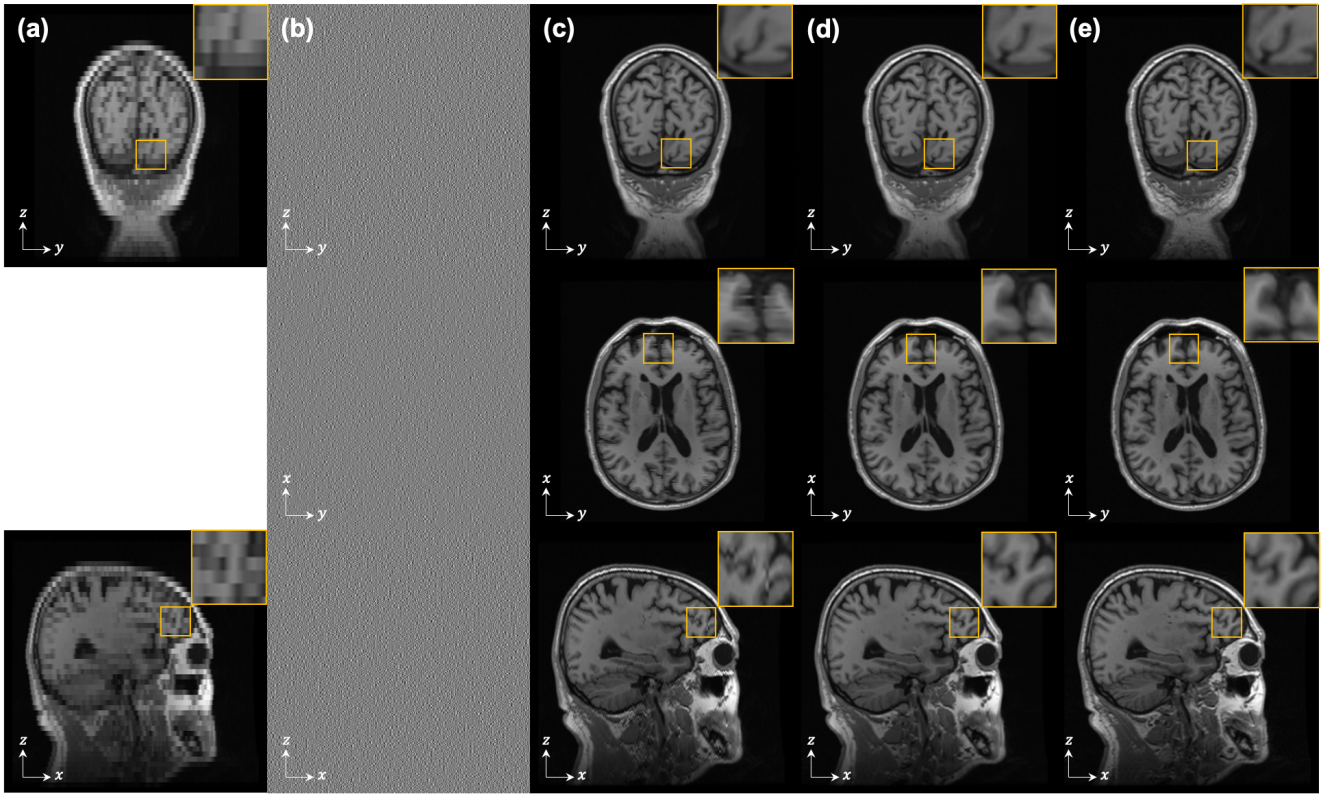

Figure 2. MR-ZSR results from 5mm→1mm (×5) of the retrospective test volume (first row: coronal slice, second row: axial slice, third row: sagittal slice). (a) measurement, (b) DiffusionMBIR [6], (c) DPS [5], (d) proposed method, (e) ground truth. For (d), first row: primary plane, second row: auxiliary plane.

Method	PSNR ↑	SSIM ↑		
		Axial^{+}	Coronal*	Sagittal
TPDM (ours)	**35.97**	**0.970**	**0.966**	**0.964**
TPDM-MEAN	32.84	0.963	0.957	0.955
TPDM-MCG	34.48	0.961	0.955	0.954
DiffusionMBIR [6]		N/W		
DPS [5]	34.77	0.965	0.963	0.960
MCG [7]	32.72	0.951	0.948	0.944

Table 1. Quantitative evaluation (PSNR, SSIM) of MR-ZSR (5mm→1mm; ×5) on the BMR-ZSR-1mm test set. TPDM-MCG: TPDM uses MCG instead of DPS, TPDM-MEAN: The forward model used to create the retrospective dataset is used. N/W: Not Working. *: primary plane, $^{+}$: auxiliary plane.

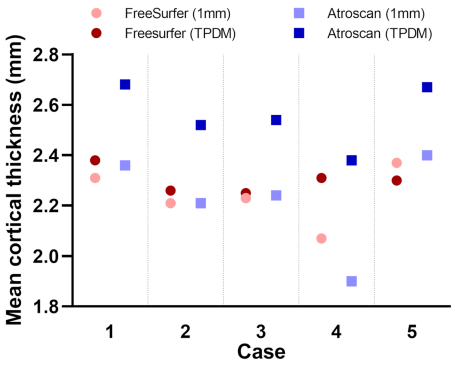

Figure 3. Comparison of the brain cortical thickness measurement result of paired ground truth 1mm volumes (1mm) and upscaled volumes from 5mm to 1mm with TPDM MR-ZSR.

forward measurement kernel. This problem is caused by the total variation loss term, which is a key point loss term that gives consistency in the batch direction in DiffusionMBIR.

As a forward model of TPDM, when merging slices, dividing the sum of slices by $\sqrt{M}$ (TPDM) instead of using N (TPDM-MEAN) yielded superior results. In addition, as a method for imposing measurement consistency constraints on the generation of the main model, TDPM with DPS (TPDM) exhibited superior outcomes than TDPM with MCG (TPDM-MCG), which is consistent with [5].

The cortical mask measured in the reconstructed 1mm image was comparable to the mask estimated in the raw 1mm image by FreeSurfer (Appendix. C.1), with a mean difference of 0.06±0.11 (paired t-test, p=0.28), indicating

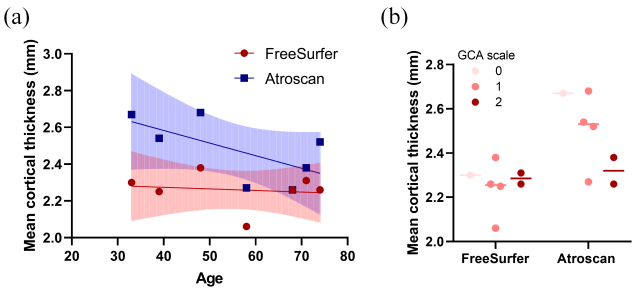

Figure 4. Relationship between mean cortical thickness, and age (a) and GCA scale (b) measured through brain MRI volume that upscaled from 5mm to 1mm with TPDM MR-ZSR.

that reconstructed T1 images by TDPM are reliably used for cortical thickness measurement that was not available in routine T1 image in clinical practice (Fig. 3). When the cortical thickness was measured by ATROSCAN, the cortical mask in the reconstructed 1mm image was larger than the cortical mask in the raw 1mm image; the mean difference was 0.34±0.08 (paired t-test; p<0.001). Nonetheless, the difference is in a quite reasonable range for clinical uses.

Although age-dependent cortical thickness decline was not clear in Freesurfer in the reconstructed T1 images, its general tendency was clearly observed in ATROSCAN as demonstrated by the dot plot (Fig. 4A). A similar trend was observed in the Global Cortical Atrophy scale (GCA), where cortical thickness by ATROSCAN was better correlated with than that by FreeSurfer (Fig. 4B), albeit the difference between scales was not significant due to the small sample (p=0.82 and 0.22, respectively).

Routine brain MR T1 images are typically acquired with 5mm thickness to save scan time. The findings from this study suggest that the TDPM model could significantly expand the pool of eligible images for volumetric measurement, which would facilitate cognitive decline research. This is particularly important given that current routine 5mm acquisition protocols are inadequate for such research. Further investigations with larger sample sizes and more diverse populations will be needed to fully demonstrate the clinical implications of image reconstruction with TDPM.

5.2. Compressed-sensing MRI (CS-MRI)

We also evaluated TPDM by performing reconstruction on retrospective ×48 acceleration Poisson sub-sampled CS-MRI volumes (Fig. 5, Table 2). For other acceleration factors, see Appendix C.2. Similarly to the outcomes by MR-ZSR, TPDM showed the best results compared to the prior art 2D/3D reverse problem-solving methods. Fig. 5 also shows that TPDM accurately reconstructed the details, surpassing all other methods.

Method	PSNR ↑	SSIM ↑		
		Axial*	Coronal+	Sagittal
TPDM (ours)	**37.17**	**0.966**	**0.967**	**0.965**
DiffusionMBIR [6]	34.83	0.907	0.909	0.906
ADMM-TV	27.01	0.812	0.802	0.812
DPS [5]	35.30	0.950	0.951	0.949
score-MRI [8]	32.75	0.849	0.853	0.855
L1-Wavelet [24]	23.15	0.557	0.530	0.535

Table 2. Quantitative evaluation (PSNR, SSIM) of CS-MRI (Poisson, ×48 acc.) on the BMR-ZSR-1mm test set. *: primary plane, +: auxiliary plane.

Method	PSNR ↑	SSIM ↑		
		Axial*	Coronal+	Sagittal
TPDM (ours)	**38.25**	**0.947**	**0.951**	**0.949**
DiffusionMBIR [6]	34.78	0.857	0.856	0.861
ADMM-TV	30.33	0.856	0.894	0.867
DPS [5]	38.20	0.942	0.943	0.941
score-CT [39]	37.56	0.922	0.922	0.924
FBPConvNet [19]	32.09	0.945	0.932	0.931

Table 3. Quantitative evaluation (PSNR, SSIM) of SV-CT (36-view) on the LDCT-CUBE test set. *: primary plane, +: auxiliary plane. Note that only 2304 2D images were used for training.

5.3. Sparse-view CT (SV-CT)

The CT problem was used for only 9 volumes (about 2000 2D images) as a train dataset data to test the performance of TPDM in extremely small data conditions. The experimental results for 36-view SV-CT are shown in Table 3, Fig. 6. Despite training with a highly limited dataset, the TPDM model performed well compared to the other models. Although the quantitative improvement over DPS [5] is not large, TPDM outperforms DPS significantly due to DPS being a 2D inverse problem solver which introduces artifacts in the batch direction when applied to 3D inverse problems. In the case of FBPConvNet [19], the SSIM in the axial direction has a small improvement, but since it is also a 2D model, it exhibits poor performance for other slice directions. Furthermore, the blurred outcomes commonly observed in convolutional networks trained using supervised techniques remain evident.

5.4. Unconditional 3D voxels volume generation

Using TPDM, we attempted to generate a full 3D voxel volume unconditionally (for the algorithm, see Appendix A.3). We trained the TPDM model using the `BMR-ZSR-1mm` dataset and used it to generate an MRI volume of the human head, the results of which are presented in Fig. 7. Notably, we were able to create a complete three-dimensional voxel volume with high resolution and quality, without relying on any measurement guidance. We believe

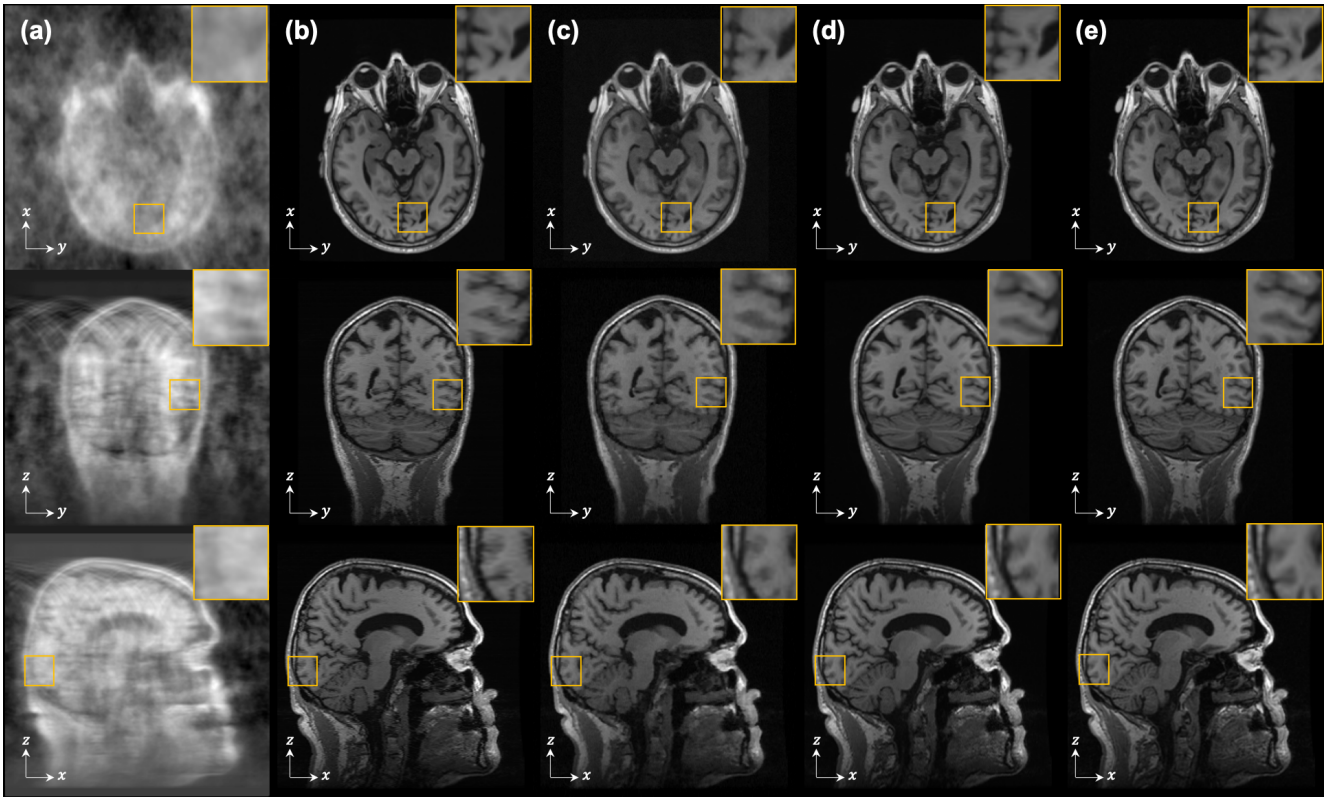

Figure 5. $\times 48$ acceleration Poisson sub-sampled CS-MRI reconstruction results of the retrospective test volume (first row: axial slice, second row: coronal slice, third row: sagittal slice). (a) Measurement, (b) DPS [5], (c) DiffusionMBIR [6], (d) proposed method, (e) ground truth. For (d), first row: primary plane, second row: auxiliary plane.

that TPDM's ability to generate 3D volumes is not solely due to the 2D image order guidance provided by the measurement in the DPS step, but also due to the alternative denoising algorithms of the two diffusion models. The empirical evidence presented here supports the reasonableness of our proposed data distribution assumptions.

6. Conclusion

In this study, we introduced TPDM, a method for solving the general 3D inverse problem and generating voxels volume with pre-trained two perpendicular 2D diffusion models. TPDM works in a completely unsupervised manner and does not require any fine-tuning for individual inverse problems. It handles 3D volume without using any 3D diffusion model by assuming a 3D distribution as the product distribution of 2D distributions, effectively avoiding the curse of dimensionality while still utilizing probability distributions of 3D volume. Our findings indicated that TPDM outperforms existing state-of-the-art 3D inverse problem-solving methods on several medical 3D reconstruction problems, even when trained with a significantly limited amount of data. Finally, using TPDM and a novel forward measurement model, we first-ever attempted diffusion-based Z-directional super-resolution of MRI images and demonstrated exceptional outcomes in both technical and clinical aspects.

Acknowledgement

This work was supported by the National Research Foundation of Korea under Grant NRF-2020R1A2B5B03001980, and by Field-oriented Technology Development Project for Customs Administration through National Research Foundation of Korea(NRF) funded by the Ministry of Science & ICT and Korea Customs Service(NRF-2021M3I1A1097938).

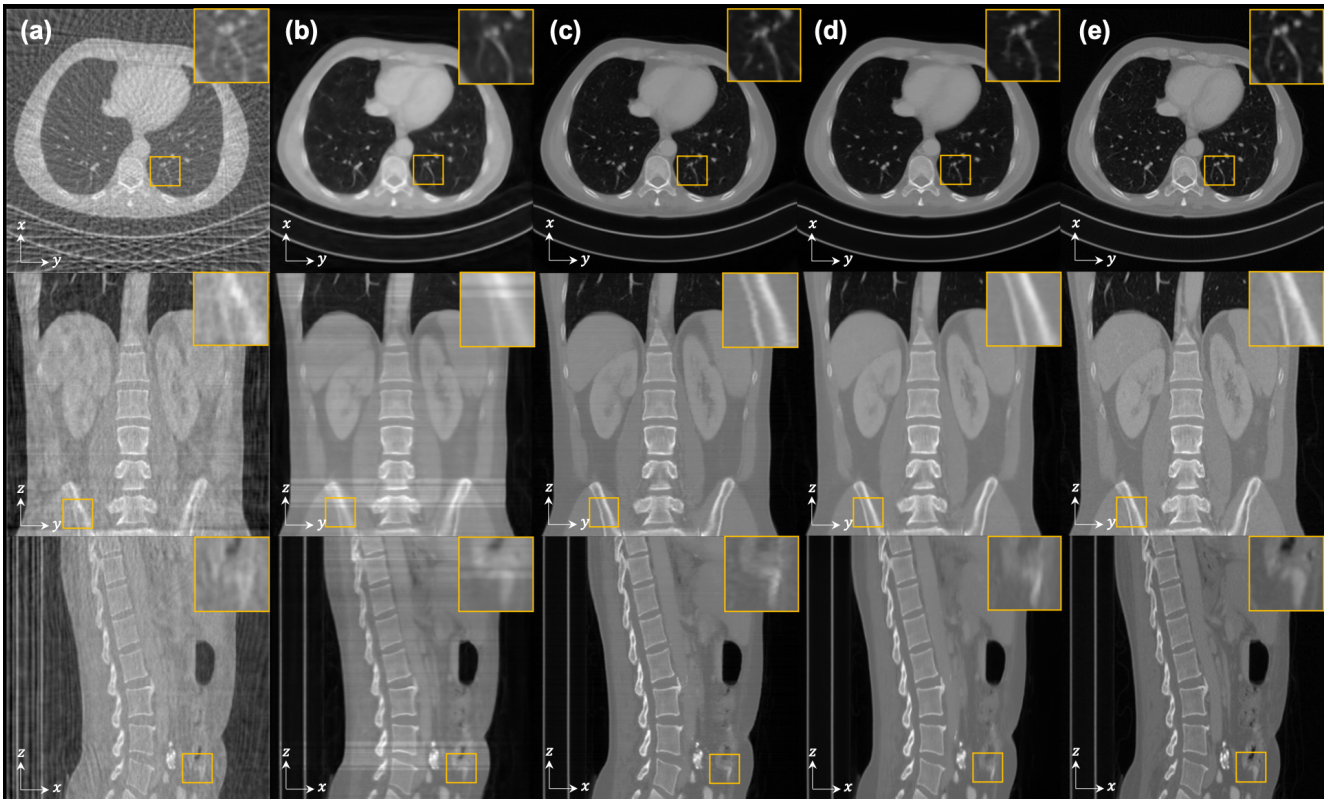

Figure 6. 36-view SV-CT reconstruction results of the retrospective test volume (first row: axial slice, second row: coronal slice, third row: sagittal slice). (a) Measurement, (b) FBPConvNet [19], (c) DPS [5], (d) proposed method, (e) ground truth. For (d), first row: primary plane, second row: auxiliary plane.

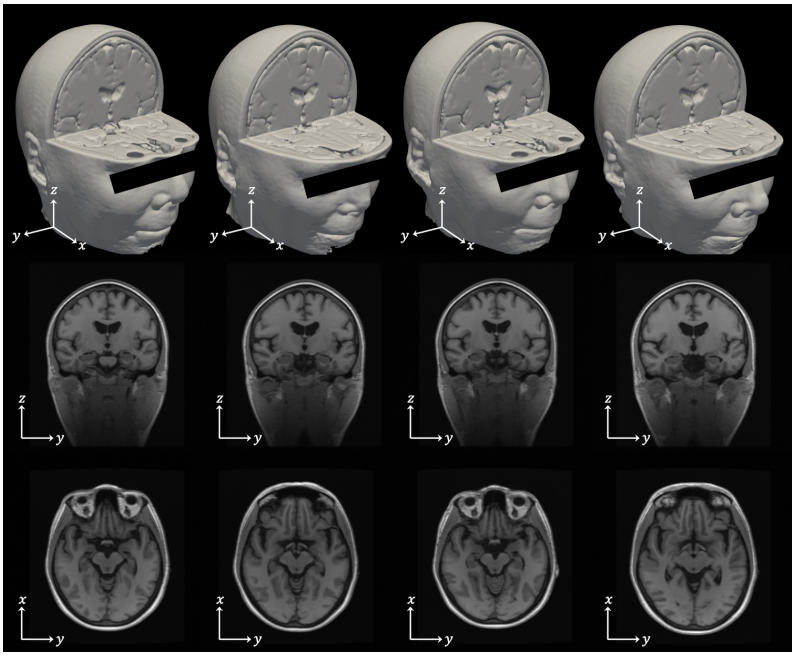

Figure 7. Results of the human head MRI volume generation using unconditioned TPDM. To visualize the volume, the iso-surface contour is expressed as a surface after removing a quarter of the volume.

References

[1] Brian DO Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982. 2

[2] Andreas Blattmann, Robin Rombach, Huan Ling, Tim Dockhorn, Seung Wook Kim, Sanja Fidler, and Karsten Kreis. Align your latents: High-resolution video synthesis with latent diffusion models. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 1

[3] Zijiao Chen, Jiaxin Qing, Tiange Xiang, Wan Lin Yue, and Juan Helen Zhou. Seeing beyond the brain: Conditional diffusion model with sparse masked modeling for vision decoding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22710–22720, 2023. 1

[4] Hyungjin Chung, Jeongsol Kim, Sehui Kim, and Jong Chul Ye. Parallel diffusion models of operator and image for blind inverse problems. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. 1

[5] Hyungjin Chung, Jeongsol Kim, Michael Thompson Mccann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *International Conference on Learning Representations*, 2023. 1, 3, 4, 5, 6, 7, 8, 9, 13, 14

[6] Hyungjin Chung, Dohoon Ryu, Michael T Mccann, Marc L Klasky, and Jong Chul Ye. Solving 3d inverse problems using pre-trained 2d diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. 1, 5, 6, 7, 8, 13, 14

[7] Hyungjin Chung, Byeongsu Sim, Dohoon Ryu, and Jong Chul Ye. Improving diffusion models for inverse problems using manifold constraints. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, editors, *Advances in Neural Information Processing Systems*, 2022. 1, 5, 6, 13

[8] Hyungjin Chung and Jong Chul Ye. Score-based diffusion models for accelerated mri. *Medical Image Analysis*, 80:102479, 2022. 5, 7, 13, 14

[9] Simon Dahan, Abdulah Fawaz, Logan ZJ Williams, Chunhui Yang, Timothy S Coalson, Matthew F Glasser, A David Edwards, Daniel Rueckert, and Emma C Robinson. Surface vision transformers: Attention-based modelling applied to cortical analysis. In *International Conference on Medical Imaging with Deep Learning*, pages 282–303. PMLR, 2022. 5

[10] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*, 34:8780–8794, 2021. 1

[11] Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011. 3

[12] Stephan J. Garbin, Marek Kowalski, Matthew Johnson, Jamie Shotton, and Julien Valentin. Fastnerf: High-fidelity neural rendering at 200fps. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 14346–14355, October 2021. 3

[13] Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He. Deepfm: A factorization-machine based neural network for ctr prediction. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, IJCAI'17, page 1725–1731. AAAI Press, 2017. 3

[14] Lorna Harper, Frederik Barkhof, Nick C Fox, and Jonathan M Schott. Using visual rating to diagnose dementia: a critical evaluation of mri atrophy scales. *Journal of Neurology, Neurosurgery & Psychiatry*, 86(11):1225–1233, 2015. 5

[15] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020. 1, 2

[16] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. In *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*, 2021. 1

[17] Han Huang, Leilei Sun, Bowen Du, Yanjie Fu, and Weifeng Lv. Graphgdp: Generative diffusion processes for permutation invariant graph generation. In *2022 IEEE International Conference on Data Mining (ICDM)*, pages 201–210. IEEE, 2022. 1

[18] Rongjie Huang, Max WY Lam, Jun Wang, Dan Su, Dong Yu, Yi Ren, and Zhou Zhao. Fastdiff: A fast conditional diffusion model for high-quality speech synthesis. *arXiv preprint arXiv:2204.09934*, 2022. 1

[19] Kyong Hwan Jin, Michael T McCann, Emmanuel Froustey, and Michael Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, 26(9):4509–4522, 2017. 5, 7, 9

[20] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, editors, *Advances in Neural Information Processing Systems*, 2022. 1

[21] Kwanyoung Kim and Jong Chul Ye. Noise2score: tweedie's approach to self-supervised image denoising without clean images. *Advances in Neural Information Processing Systems*, 34:864–874, 2021. 3

[22] Zhifeng Kong, Wei Ping, Jiaji Huang, Kexin Zhao, and Bryan Catanzaro. Diffwave: A versatile diffusion model for audio synthesis. *arXiv preprint arXiv:2009.09761*, 2020. 1

[23] Anish Lahiri, Gabriel Maliakal, Marc L Klasky, Jeffrey A Fessler, and Saiprasad Ravishankar. Sparse-view cone beam ct reconstruction using data-consistent supervised and adversarial learning from scarce training data. *IEEE Transactions on Computational Imaging*, 9:13–28, 2023. 5

[24] Michael Lustig, David Donoho, and John M Pauly. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007. 5, 7

[25] Cynthia H McCollough, Adam C Bartley, Rickey E Carter, Baiyu Chen, Tammy A Drees, Phillip Edwards, David R Holmes III, Alice E Huang, Farhana Khan, Shuai Leng, et al. Low-dose ct for the detection and classification of metastatic liver lesions: results of the 2016 low dose ct grand challenge. *Medical physics*, 44(10):e339–e352, 2017. 4, 13

[26] Eyal Molad, Eliahu Horwitz, Dani Valevski, Alex Rav Acha, Yossi Matias, Yael Pritch, Yaniv Leviathan, and Yedid Hoshen. Dreamix: Video diffusion models are general video editors. *arXiv preprint arXiv:2302.01329*, 2023. 1

[27] Norman Müller, Yawar Siddiqui, Lorenzo Porzi, Samuel Rota Bulo, Peter Kontschieder, and Matthias Nießner. Diffrf: Rendering-guided 3d radiance field diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4328–4338, 2023. 1

[28] Vadim Popov, Ivan Vovk, Vladimir Gogoryan, Tasnima Sadekova, and Mikhail Kudinov. Grad-tts: A diffusion probabilistic model for text-to-speech. In *International Conference on Machine Learning*, pages 8599–8608. PMLR, 2021. 1

[29] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 2022. 1

[30] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10684–10695, 2022. 1

[31] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 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*, pages 234–241. Springer, 2015. 5

[32] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L Denton, Kamyar Ghasemipour, Raphael Gontijo Lopes, Burcu Karagol Ayan, Tim Salimans, et al. Photorealistic text-to-image diffusion models with deep language understanding. *Advances in Neural Information Processing Systems*, 35:36479–36494, 2022. 1

[33] Tara N Sainath, Brian Kingsbury, Vikas Sindhwani, Ebru Arisoy, and Bhuvana Ramabhadran. Low-rank matrix factorization for deep neural network training with high-dimensional output targets. In *2013 IEEE international conference on acoustics, speech and signal processing*, pages 6655–6659. IEEE, 2013. 3

[34] Philip Scheltens, Timo Erkinjunti, Didier Leys, Lars-Olaf Wahlund, Domenico Inzitari, Theodoro del Ser, Florence Pasquier, Frederik Barkhof, Riita Mäntylä, John Bowler, et al. White matter changes on ct and mri: an overview of visual rating scales. *European neurology*, 39(2):80–89, 1998. 5

[35] J Ryan Shue, Eric Ryan Chan, Ryan Po, Zachary Ankner, Jiajun Wu, and Gordon Wetzstein. 3d neural field generation using triplane diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20875–20886, 2023. 1

[36] Uriel Singer, Adam Polyak, Thomas Hayes, Xi Yin, Jie An, Songyang Zhang, Qiyuan Hu, Harry Yang, Oron Ashual, Oran Gafni, et al. Make-a-video: Text-to-video generation without text-video data. *arXiv preprint arXiv:2209.14792*, 2022. 1

[37] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pages 2256–2265. PMLR, 2015. 1, 2

[38] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32, 2019. 1

[39] Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models. In *International Conference on Learning Representations*, 2022. 1, 5, 7, 13

[40] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021. 1, 2, 3, 5

[41] Yu Takagi and Shinji Nishimoto. High-resolution image reconstruction with latent diffusion models from human brain activity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14453–14463, 2023. 1

[42] Nicholas J Tustison, Philip A Cook, Arno Klein, Gang Song, Sandhitsu R Das, Jeffrey T Duda, Benjamin M Kandel, Niels van Strien, James R Stone, James C Gee, et al. Large-scale evaluation of ants and freesurfer cortical thickness measurements. *Neuroimage*, 99:166–179, 2014. 5

[43] Clement Vignac, Igor Krawczuk, Antoine Siraudin, Bohan Wang, Volkan Cevher, and Pascal Frossard. Digress: Discrete denoising diffusion for graph generation. *arXiv preprint arXiv:2209.14734*, 2022. 1

[44] Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural computation*, 23(7):1661–1674, 2011. 3

[45] Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. *The Eleventh International Conference on Learning Representations*, 2023. 1

[46] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 5

[47] Bo Zhou and S Kevin Zhou. Dudornet: learning a dual-domain recurrent network for fast mri reconstruction with deep t1 prior. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4273–4282, 2020. 5