Annealed Score-Based Diffusion Model for MR Motion Artifact Reduction

Gyutaek Oh, Jeong Eun Lee, and Jong Chul Ye, *Fellow, IEEE*

Abstract—Motion artifact reduction is one of the important research topics in MR imaging, as the motion artifact degrades image quality and makes diagnosis difficult. Recently, many deep learning approaches have been studied for motion artifact reduction. Unfortunately, most existing models are trained in a supervised manner, requiring paired motion-corrupted and motion-free images, or are based on a strict motion-corruption model, which limits their use for real-world situations. To address this issue, here we present an annealed score-based diffusion model for MRI motion artifact reduction. Specifically, we train a score-based model using only motion-free images, and then motion artifacts are removed by applying forward and reverse diffusion processes repeatedly to gradually impose a low-frequency data consistency. Experimental results verify that the proposed method successfully reduces both simulated and in vivo motion artifacts, outperforming the state-of-the-art deep learning methods.

Index Terms—MRI, motion artifact, score-based models, diffusion models

I. INTRODUCTION

MAGNETIC resonance imaging (MRI) is an imaging technique that provides various types of contrast images without radiation exposure or invasive procedure. Despite many advantages, MRI requires a long acquisition time due to its imaging physics. Furthermore, the long acquisition time leads to motion artifacts due to the movement of the patient, so the motion artifact is considered one of the main problems of MRI.

In addition, the contrast agent injection may cause motion artifacts in MRI. For example, gadoxetic acid (Gd-EOB-DTPA) is one of the liver-specific MRI contrast agents that can help the diagnosis of diseases such as hepatocellular carcinoma, liver metastases [1], [2] by providing hepatobiliary phase (HBP) imaging [3]. However, the administration of Gd-EOB-DTPA can occur acute transient dyspnea, resulting in transient severe motion (TSM) [4]. If TSM occurs, the image quality of the arterial phase is degraded, and the accuracy of diagnosis can be affected. So, an algorithm to correct the motion artifact due to TSM is required.

There have been several attempts to reduce the motion artifact of MRI by tracking the motion [5], [6], or changing

G. Oh is with the Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea (e-mail: okt0711@kaist.ac.kr). J. C. Ye is with the Kim Jaechul Graduate School of AI, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea (e-mail: jong.ye@kaist.ac.kr). J.E. Lee is with the Department of Radiology, Chungnam National University Hospital, Chungnam National University College of Medicine, 282 Munhwa-ro, Jung-gu, Daejeon 35015, Republic of Korea (e-mail: leeje290@gmail.com).

sampling trajectory or imaging sequence [7]–[9]. However, they require additional devices or scan time, and the types of motion artifacts corrected by these methods are limited.

Motion artifact correction algorithms based on compressed sensing (CS) [10]–[12] also have been investigated. CS-based algorithms have shown high-quality results, but they have limitations such as the difficulty of hyperparameter tuning and high computational complexity. Furthermore, many CS algorithms require raw k-space data, which are rarely obtained in clinical environments due to storage limitations.

Recent studies for MRI motion artifact reduction are based on deep learning methods [13]–[19]. Deep learning methods have shown improved performance and reduced run time compared to previous methods. However, most of the deep learning methods are based on supervised learning approaches. Since paired motion-free and corrupted data are difficult to obtain, these methods usually utilize simulated motion artifact images to train the networks. Therefore, it is difficult to apply them to real motion-corrupted data.

To overcome the limitation of simulation-based deep learning methods, deep learning methods using unpaired data have also been explored. Some methods interpret the motion artifact reduction problem as image-to-image translation [20], [21], and address them based on CycleGAN architecture [22]. Although they utilize real motion artifact data, the performance of these algorithms is often limited because there is no explicit motion artifact rejection mechanism.

Recently, we proposed an algorithm for MR motion artifact reduction using bootstrap subsampling and aggregation [23]. Under the assumption that the motion artifact appears as k-space outliers, the method removes the motion artifact by rejecting k-space outliers in a probabilistic manner. Although our prior method outperforms other simulation-based or unpaired deep learning methods, there exist limitations if the motion artifact does not appear as sparse outliers in k-space.

Recently, score-based diffusion models [24]–[26] have shown remarkable performance in the field of image generation. In score-based diffusion models, a network that estimates the score, the gradient of the log probability density function, is first trained, and then images can be generated by solving the reverse-time stochastic differential equation (SDE). Furthermore, it has been verified that unconditionally trained diffusion models can be applied to solve various inverse problems by adjusting sampling procedure using constraints [27]–[31]. Importantly, the unconditionally trained diffusion models do not require paired data, so it is possible to solve inverse problems in an unsupervised manner.

Inspired by this, here we propose a novel MRI motion

Fig. 1: The overall procedure of our method. $\boldsymbol{x}_{N'}$ is generated from the motion-corrupted image $\boldsymbol{x}_0$ by the forward diffusion. Then the MR image with reduced motion artifact $\boldsymbol{x}_0$ is sampled by solving the reverse-time SDE. $\boldsymbol{y}$ denotes the measurement (k-space of motion-corrupted image), and it is used in the data consistency step to prevent the severe deformation of the output image. The output goes through forward and reverse diffusion iteratively to obtain the final reconstructed image.

artifact reduction method using score-based diffusion models. Fig. 1 shows the overall procedure of the proposed method. During reverse diffusion, the low-frequency data consistency is gradually imposed in an iterative manner so that the overall structure of the original image is maintained and helps to remove only motion artifact components.

In particular, our constraint is designed based on the observation that the motion artifacts in MRI usually occurred in the high-frequency region of k-space. This is because k-space acquisition is usually performed first in the center region and motion occurs after a certain period after the start of acquisition, so k-space samples that include motion artifacts generally appear in high-frequency regions. Therefore, during the reverse diffusion, the low-frequency region needs to be maintained and only the high-frequency region should be corrected by diffusion sampling. However, because the high-frequency region of k-space also contains the information of details, the detailed structures of images can be altered or vanished if the data consistency step in Eq. (9) is applied directly. To address this issue, we propose an annealed reverse sampling procedure where the data consistency step is gradually applied in a repeated manner to maintain high-frequency details of measurements.

The remaining parts of the paper are constructed as follows. Section II introduces backgrounds of score-based diffusion models. Section III contains the key idea of the proposed method. The experimental setting is explained in Section IV, and qualitative and quantitative results are shown in Section V. Section VI and VII contains the discussion and conclusion of our paper.

II. BACKGROUNDS

A. Score-Based Diffusion Models

A continuous diffusion process can be represented as $\{\boldsymbol{x}(t)\}_{t=0}^1$, where $t \in [0, 1]$ denotes the time variable. Here, $\boldsymbol{x}(0) \sim p_{data}$ where p_{data} is the data distribution, and

$\boldsymbol{x}(1) \sim p_1$ where p_1 refers to the noise distribution which is commonly set to Gaussian distribution. Then, the diffusion process can be modeled by the solution of the following stochastic differential equation:

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

where $\boldsymbol{f} : \mathbb{R}^d \to \mathbb{R}^d$ is a drift coefficient, $g : \mathbb{R} \to \mathbb{R}$ is a diffusion coefficient, and $\boldsymbol{w}$ denotes a standard Wiener process. By solving Eq. (1), it is possible to transmit a sample from the data distribution to that of the noise distribution through the forward diffusion process.

If it is possible to reverse the diffusion process in Eq. (1), then we can obtain samples of the data distribution from samples of the noise distribution. In [32], it was shown that the reverse process is also a diffusion process that can be modeled by following reverse SDE:

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

where $\bar{\boldsymbol{w}}$ is also a standard Wiener process from time 1 to 0, and $\nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x})$ denotes the score function. Therefore, if the score function can be estimated, it is possible to derive the reverse diffusion process and generate samples of data distribution from random Gaussian noise.

Among the many possible choices of $\boldsymbol{f}$ and g, we choose variance exploding SDE (VE-SDE) [26], where $\boldsymbol{f}$ and g are defined by

$$\boldsymbol{f} = 0, \quad g = \sqrt{\frac{d[\sigma^2(t)]}{dt}}, \qquad (3)$$

and

$$\sigma(t) = \sigma_{\min} \left(\frac{\sigma_{\max}}{\sigma_{\min}} \right)^t. \qquad (4)$$

Then, the reverse SDE in Eq. (2) can be rewritten as:

$$d\boldsymbol{x} = -\frac{d[\sigma^2(t)]}{dt} \nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x})dt + \sqrt{\frac{d[\sigma^2(t)]}{dt}}d\bar{\boldsymbol{w}}. \qquad (5)$$

Algorithm 1 CCDF with PC sampler

Require: $\boldsymbol{x}_0$, $\boldsymbol{y}$, N', $\{\sigma_i\}_{i=1}^{N'}$, $\{\epsilon_i\}_{i=1}^{N'}$, $\boldsymbol{s}_{\boldsymbol{\theta}}$
1: $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$
2: $\boldsymbol{x}_{N'} \leftarrow \boldsymbol{x}_0 + \sigma_{N'}\boldsymbol{z}$ ▷ Forward diffusion
3: **for** $i = N'$ to 1 **do**
4:     $\boldsymbol{x}'_{i-1} \leftarrow \boldsymbol{x}_i + (\sigma_i^2 - \sigma_{i-1}^2)\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}_i, \sigma_i)$
5:     $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$
6:     $\boldsymbol{x}_{i-1} \leftarrow \boldsymbol{x}'_{i-1} + \sqrt{\sigma_i^2 - \sigma_{i-1}^2}\boldsymbol{z}$ ▷ Predictor
7:     $\boldsymbol{x}_{i-1} \leftarrow$ Data consistency$(\boldsymbol{x}_{i-1}, \boldsymbol{y})$
8:                                   ▷ Data consistency
9:     $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$
10:    $\boldsymbol{x}_{i-1} \leftarrow \boldsymbol{x}_{i-1} + \epsilon_i \boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}_i, \sigma_i) + \sqrt{2\epsilon_i}\boldsymbol{z}$ ▷ Corrector
11:    $\boldsymbol{x}_{i-1} \leftarrow$ Data consistency$(\boldsymbol{x}_{i-1}, \boldsymbol{y})$
12:                                   ▷ Data consistency
13: **end for**

Here, the time index t is usually discretized uniformly into N intervals, and $\boldsymbol{x}_i$ and σ_i can be defined as

$$\boldsymbol{x}_i := \boldsymbol{x}(t)|_{t=\frac{i-1}{N-1}}, \quad \sigma_i := \sigma_{\min}\left(\frac{\sigma_{\max}}{\sigma_{\min}}\right)^{\frac{i-1}{N-1}}. \quad (6)$$

The score function $\nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x})$ is generally estimated by training a neural network $\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}(t), t)$ with denoising score matching [33]. The training of the score-based model with denoising score matching can be done by minimizing the following objective function:

$$\min_{\boldsymbol{\theta}} \mathbb{E}_t\left\{\lambda(t)\mathbb{E}_{\boldsymbol{x}(0)}\mathbb{E}_{\boldsymbol{x}(t)|\boldsymbol{x}(0)}\Big[\right.$$
$$\left.\left\|\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}(t), t) - \nabla_{\boldsymbol{x}(t)} \log p_{0t}(\boldsymbol{x}(t)|\boldsymbol{x}(0))\right\|_2^2\Big]\right\}. \quad (7)$$

After training the neural network and plugging it into Eq. (5), the reverse SDE can be solved by numerical SDE solvers or predictor-corrector (PC) samplers [26].

B. Come-Closer-Diffuse-Faster (CCDF)

The main drawback of score-based diffusion models is their slow sampling time. Because the sampling starts from the random Gaussian noise and usually requires thousands of steps, the sampling time of score-based diffusion models is too long. In the prior work [28], the authors proposed a method called Come-Closer-Diffuse-Faster (CCDF) to reduce the sampling time of diffusion models when solving inverse problems. Specifically, instead of starting sampling from random Gaussian noise, the forward diffusion is first applied from the initial reconstruction, leading to only few steps of reverse diffusion to get the final reconstruction.

More specifically, Algorithm 1 shows the CCDF sampling procedure using the PC sampler, where $\boldsymbol{y}$ denotes the initial measurement, and $N' = Nt'$ is the number of reverse diffusion steps where $t' \in [0, 1]$. Here, the data consistency step should be non-expansive to maintain the stochastic contraction mapping nature of reverse diffusion sampling [28]. With a better initialization followed by one-step forward diffusion, CCDF largely reduces the reverse sampling time for solving inverse problems [28].

Fig. 2: The data consistency step of the proposed method.

III. MAIN CONTRIBUTION

A. Motivation

In our prior work [23], we solved the motion artifact reduction problem by regarding motion artifacts as sparse outliers in k-space. Specifically, if the motion is occurred by translation or rotation, it is assumed to result in k-space phase shift or rotation at the specific phase encoding lines:

$$\hat{\boldsymbol{y}}(k_x, k_y) = \begin{cases} \boldsymbol{F}\{\boldsymbol{R}_\alpha \boldsymbol{x}\}e^{-j\Phi}, & k_y \in \mathbb{K} \\ \boldsymbol{F}\{\boldsymbol{x}\}, & \text{otherwise,} \end{cases} \quad (8)$$

where $\hat{\boldsymbol{y}}$ denotes the motion-corrupted k-space with the indices along the frequency encoding direction k_x and phase encoding direction k_y, and $\boldsymbol{x}$ is the motion-free image, $\boldsymbol{F}$ denotes the Fourier transform, $\boldsymbol{R}_\alpha$ denotes the rotation operation with the angle α, Φ is the displacement in radian, and $\mathbb{K}$ is the phase encoding indices where the rotation or translation occurred.

Based on this assumption, the network is trained to reconstruct fully sampled motion-free images from randomly subsampled images along the phase encoding direction in which the corrupted k-space data can be removed in a probabilistic manner. Although this method does not require simulated motion artifact images and shows improved performance, it has a limitation in that it is difficult to apply when the motion artifacts cannot be considered as sparse outliers in k-space. Furthermore, because the index of outliers is not known, some outliers that are not removed by subsampling can remain in the reconstructed image.

B. Proposed Method

Rather than using the sparse outlier assumption in Eq. (8), our method is based on a more relaxed assumption that the motion artifacts in MRI mainly occur in the high-frequency region of k-space. This is because k-space acquisition is usually performed first in the center region and motion occurs after a certain period after the start of acquisition so that k-space samples with motion artifacts generally appear in high-frequency regions. Therefore, the high-frequency region of k-space should be corrected to remove the motion artifact.

The application of CCDF in Algorithm 1 starts from the one-step forward diffusion from the initialization. Then, a naïve way of using data consistency for reverse diffusion would be to impose the low-frequency region consistency:

$$\boldsymbol{x}_{i-1} = (\boldsymbol{I} - \boldsymbol{F}^{-1}\boldsymbol{P}_\Omega \boldsymbol{F})\boldsymbol{x}'_{i-1} + \boldsymbol{F}^{-1}\boldsymbol{P}_\Omega \boldsymbol{y}, \quad (9)$$

Algorithm 2 MR Motion Artifact Reduction

Require: $\boldsymbol{x}_0$, $\boldsymbol{y}$, N', $\{\sigma_i\}_{i=1}^{N'}$, $\{\epsilon_i\}_{i=1}^{N'}$, $\{\lambda_i\}_{i=1}^{N'}$, $\boldsymbol{s_\theta}$
1: **for** $j = 1$ to M **do**
2:     $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$
3:     $\boldsymbol{x}_{N'} \leftarrow \boldsymbol{x}_0 + \sigma_{N'}\boldsymbol{z}$ ▷ Forward diffusion
4:     **for** $i = N'$ to 1 **do**
5:         $\boldsymbol{x}'_{i-1} \leftarrow \boldsymbol{x}_i + (\sigma_i^2 - \sigma_{i-1}^2)\boldsymbol{s_\theta}(\boldsymbol{x}_i, \sigma_i)$
6:         $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$
7:         $\boldsymbol{x}_{i-1} \leftarrow \boldsymbol{x}'_{i-1} + \sqrt{\sigma_i^2 - \sigma_{i-1}^2}\boldsymbol{z}$ ▷ Predictor
8:         $\boldsymbol{x}_{i-1} \leftarrow (1 - \lambda_i)(\boldsymbol{I} - \boldsymbol{F}^{-1}\boldsymbol{P}_\Omega\boldsymbol{F})\boldsymbol{x}_{i-1}$
9:             $+ \lambda_i\boldsymbol{F}^{-1}(\boldsymbol{I} - \boldsymbol{P}_\Omega)\boldsymbol{y} + \boldsymbol{F}^{-1}\boldsymbol{P}_\Omega\boldsymbol{y}$
10:                         ▷ Data consistency
11:         $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$
12:         $\boldsymbol{x}_{i-1} \leftarrow \boldsymbol{x}_{i-1} + \epsilon_i\boldsymbol{s_\theta}(\boldsymbol{x}_i, \sigma_i) + \sqrt{2\epsilon_i}\boldsymbol{z}$ ▷ Corrector
13:         $\boldsymbol{x}_{i-1} \leftarrow (1 - \lambda_i)(\boldsymbol{I} - \boldsymbol{F}^{-1}\boldsymbol{P}_\Omega\boldsymbol{F})\boldsymbol{x}_{i-1}$
14:             $+ \lambda_i\boldsymbol{F}^{-1}(\boldsymbol{I} - \boldsymbol{P}_\Omega)\boldsymbol{y} + \boldsymbol{F}^{-1}\boldsymbol{P}_\Omega\boldsymbol{y}$
15:                         ▷ Data consistency
16:     **end for**
17: **end for**

---

where $\boldsymbol{P}_\Omega$ is the operator that samples only the low-frequency region of k-space. In other words, during reverse diffusion, the low-frequency region is maintained so that only the high-frequency region is corrected by the diffusion model.

However, because the high-frequency region of k-space also contains the information of details, the detailed structures of images can be altered or vanished if the data consistency step in Eq. (9) is applied directly. To address this issue, we propose an annealed data consistency step to maintain high-frequency details of measurements as illustrated in Fig. 2:

$$
\begin{aligned}
\boldsymbol{x}_{i-1} = (1 - \lambda_i)(\boldsymbol{I} - \boldsymbol{F}^{-1}\boldsymbol{P}_\Omega\boldsymbol{F})\boldsymbol{x}'_{i-1} \\
+ \lambda_i\boldsymbol{F}^{-1}(\boldsymbol{I} - \boldsymbol{P}_\Omega)\boldsymbol{y} + \boldsymbol{F}^{-1}\boldsymbol{P}_\Omega\boldsymbol{y},
\end{aligned}
\tag{10}
$$

where $\lambda_i \in [0, 1]$ is the annealing hyperparameter to control the weight of high-frequency components of the measurement. Furthermore, as shown in Algorithm 2, we choose relatively small N', and repeat forward and reverse processes M times so that the high-frequency components of the measurement are gradually added at each data consistency step.

Here, Eq. (10) can be written as

$$
\boldsymbol{x}_{i-1} = \boldsymbol{T}(\boldsymbol{x}_{i-1}) := \boldsymbol{A}\boldsymbol{x}'_{i-1} + \boldsymbol{b},
$$

where

$$
\begin{aligned}
\boldsymbol{A} &= (1 - \lambda_i)(\boldsymbol{I} - \boldsymbol{F}^{-1}\boldsymbol{P}_\Omega\boldsymbol{F}), \\
\boldsymbol{b} &= \lambda_i\boldsymbol{F}^{-1}(\boldsymbol{I} - \boldsymbol{P}_\Omega) + \boldsymbol{F}^{-1}\boldsymbol{P}_\Omega.
\end{aligned}
$$

Since $\left\|\boldsymbol{I} - \boldsymbol{F}^{-1}\boldsymbol{P}_\Omega\boldsymbol{F}\right\| \leq 1$ [28], it is also true that $\|\boldsymbol{A}\| = \left\|(1 - \lambda_i)(\boldsymbol{I} - \boldsymbol{F}^{-1}\boldsymbol{P}_\Omega\boldsymbol{F})\right\| \leq 1$. Therefore, $\boldsymbol{T}$ is a non-expansive mapping, so it can accelerate the reverse diffusion process through the CCDF principle [28].

C. Implementation Details

In our implementation, we choose VE-SDE, which results in the following one-step forward sampling:

$$
\boldsymbol{x}(t) = \boldsymbol{x}(0) + \sigma(t)\boldsymbol{z}
\tag{11}
$$

where $\boldsymbol{z} \sim \mathcal{N}(0, \boldsymbol{I})$ and $\boldsymbol{x}(0)$ is the clean training data. By plugging this in Eq. (7), we have the following cost function [26]:

$$
\min_\theta \mathbb{E}_t\mathbb{E}_{\boldsymbol{x}(0)}\mathbb{E}_{\boldsymbol{x}(t)|\boldsymbol{x}(0)}\left[\left\|\sigma(t)\boldsymbol{s_\theta}(\boldsymbol{x}(t), t) - \frac{\boldsymbol{x}(t) - \boldsymbol{x}(0)}{\sigma(t)}\right\|_2^2\right].
\tag{12}
$$

Here, we choose the number of discretized steps $N = 1000$, and $\sigma_{\min}$ and $\sigma_{\max}$ in Eq. (6) are set to $\sigma_{\min} = 0.01$ and $\sigma_{\max} = 50$, respectively. We train the score model for 1.3M iterations and follow [30] for the setting of other hyperparameters such as optimization, batch size, learning rate, gradient clipping, or exponential moving average.

In addition, for N', M and λ_i in Algorithm 2, we choose $N' = 10$, $M = 3$, and

$$
\lambda_i = \frac{\lambda_{N'}}{N' - 1}(i - 1),
\tag{13}
$$

where $\lambda_{N'} = 0.01$. In other words, λ_i linearly decreases to 0 as i goes to 1, so the weight of high-frequency components of the measurement decreases as reverse diffusion proceeds.

In CCDF [28], it was shown that a better initialization provides faster reverse sampling. Accordingly, the neural network (NN) initialization could be utilized if available as it is better than the original artifact-corrupted images. Accordingly, we also employed NN initialization with [20] for the brain dataset, and [23] for the liver dataset.

IV. METHODS

A. Experimental Data

In our experiments, we use two MR datasets. The first dataset is the human connectome project (HCP) dataset which is the public dataset that contains human brain MR images. This dataset is acquired by Siemens 3T system with 3D spin echo sequence, and the scan parameters are as follows: TR = 3200 ms, TE = 565 ms, echo train duration = 1105, matrix size = 320×320, voxel size = 0.7×0.7×0.7 mm³, and phase encoding direction = anterior-posterior. Because the HCP dataset does not contain motion-corrupted images, it is used for quantitative evaluation with motion artifact simulation. The score model is trained with 3000 motion-free MR images from 150 subjects, and other 800 images from 40 subjects are used for testing.

The second dataset is collected from Chungnam National University Hospital (CNUH), and it includes Gd-EOB-DTPA-enhanced liver MR images. It is obtained by a 3T Philips Achieva MR system with the following scan parameters: TR = 3.1 ms, TE = 1.5 ms, flip angle = 10°, field of view = 256×256 mm², slice thickness/intersection gap = 2/0 mm, acquisition matrix = 320×192, and phase encoding direction = anterior-posterior. Also, dynamic imaging including various phases was obtained, but only arterial phase images are used for experiments because TSM usually occurs during the arterial phase. The liver dataset consists of two groups, motion-free images, and motion-corrupted images. For the training of the score model, 3097 motion-free images from 18 subjects are

used. After training, 444 simulated motion-corrupted images from 5 subjects are selected for the quantitative evaluation, and 38 MR volumes with in vivo motion-corrupted images are used for qualitative and radiologist evaluations.

B. Artifact Simulation

For the quantitative evaluation, we used simulated motion artifacts. The simulation was performed similarly to prior works [16], [23]. The first type of motion artifact that we simulate is random translation and rotation. We simulate the first type of motion artifact with the HCP brain dataset. The motion artifact with random translation and rotation can be simulated by Eq. (8) with

$$(\alpha, \Phi) = \begin{cases} (\alpha_{k_y}, k_y\Delta_{k_y} + k_x\Delta_{k_x}), & |k_y| > k_0 \\ (0,0), & \text{otherwise,} \end{cases} \quad (14)$$

where α_{k_y} denotes the rotation angle, Δ_{k_y} and Δ_{k_x} denote the degree of motion along x and y direction, respectively, and k_0 is the delay time of the phase error due to the centric k-space filling. In our simulation, k_0 is fixed to $\pi/10$, α_{k_y} is randomly sampled from $[-2°, 2°]$, Δ_{k_y} and Δ_{k_x} are sampled from $[-1\text{cm}, 1\text{cm}]$ and $[-0.5\text{cm}, 0.5\text{cm}]$, respectively, at each k-space line.

The second type of simulated motion is respiratory motion, which appears as a sinusoidal function in k-space [16], [23]:

$$(\alpha, \Phi) = \begin{cases} (0, k_y\Delta_{k_y}\sin(mk_y + n)), & |k_y| > k_0 \\ (0,0), & \text{otherwise,} \end{cases} \quad (15)$$

where Δ_{k_y}, m, and n denote the amplitude, period, and phase shift of the sinusoidal function, respectively. Because the respiratory motion appears in abdominal MR images, we simulate it with the liver MR image dataset. Parameters for the simulation are sampled as follows: $k_0 \sim U[\pi/10, \pi/5]$, $\Delta_{k_y} \sim U[1\text{cm}, 1.5\text{cm}]$, $m \sim U[0.1, 5.0]$, and $n \sim U[0, \pi/4]$, where $U[a, b]$ denotes the uniform distribution with the interval $[a, b]$

C. Comparison Methods

We compared our method with three state-of-the-art methods to verify the performance of the method. The first comparison method is MARC [16], a method for reducing liver MRI motion artifacts. Because it is a supervised method, we train MARC models using simulated motion-corrupted images with Eqs. (14) and (15).

The second comparison method is Cycle-MedGAN V2.0 [20], an unpaired deep learning method based on CycleGAN [22]. Cycle-MedGAN V2.0 can be trained with both simulated or in vivo motion-corrupted data, but we train it with only simulated motion-corrupted data because the training of Cycle-MedGAN V2.0 was unstable when using in vivo data.

We also employed the bootstrap subsampling and aggregation method in [23] as a comparison method. Because this method requires only motion-free images during training, simulated or in vivo motion-corrupted images were not used during training.

D. Evaluation Methods

For the quantitative evaluation, we used the peak signal-to-noise ratio (PSNR) and the structural similarity index metric (SSIM). Because there is no ground truth matched with in vivo motion-corrupted images, the quantitative evaluation was performed with simulated motion-corrupted images.

In addition, we also conducted a clinical evaluation with the results using in vivo motion-corrupted data. Specifically, a radiologist with 13 years of experience in abdominal MR imaging performed an analysis of the results of various methods. The image analysis was conducted from various perspectives. First, the performance in reducing motion artifacts is rated using a 5-point scoring system: 1 = non-diagnostic (severe artifacts causing impaired diagnostic capability of the readers); 2 = substantial artifacts with image quality decrease, but diagnostic performance impairment; 3 = mild artifacts, no significant (only mild) image quality disturbance; 4 = minimal artifacts, sharp image; 5 = no artifacts. The image noise level is also evaluated with the following scoring system: 1 = non-diagnostic (severe noise causing impaired diagnostic capability of the readers); 2 = substantial noise with image quality decrease, but diagnostic performance impairment; 3 = mild noise, no significant (only mild) image quality disturbance; 4 = minimal noise; 5 = no noise. Next, the blurring can be induced when reducing the motion artifact, so the rating of image blurring level is performed: 1 = non-diagnostic (severely pixelated texture causing impaired diagnostic capability of the readers); 2 = substantially pixelated, artificial sensation with concerns about the loss of normal texture, without diagnostic performance impairment; 3 = mildly pixelated, artificial sensation, without image quality decrease; 4 = minimal alteration of image texture; 5 = no alteration of image texture. Furthermore, because the hepatic artery (HA) on the arterial phase should be visualized clearly, the vessel clarity is evaluated with a scoring system: 1 = not delineated due to motion or low signal-to-noise ratio (SNR); 2 = blur or decreased SNR; 3 = clear common hepatic artery (CHA) and proper hepatic artery (PHA), but blurred HA and gastroduodenal artery (GDA); 4 = entire HA is clearly visible, clear CHA, GDA, bilateral HA; 5 = strong contrast-to-noise ratio with score 4. Last, the overall image quality is assessed by following scoring system: 1 = non-diagnostic; 2 = not satisfactory image quality, but re-examination is not needed; 3 = acceptable image quality (image quality may not be very good, but clinically acceptable); 4 = good image quality without significant artifact; 5 = excellent image quality without artifact and good spatial resolution. The score is rated for each volume in all assessments. Also, the results were presented to the radiologist in a random order without any labeling for a fair comparison.

V. RESULTS

A. Results with Simulated Data

Fig. 3 shows the motion artifact reduction results of various methods with random simulated motion-corrupted data. As shown in Fig. 3(a), it is hard to recognize detailed structures of brains due to motion artifacts. MARC [16] reduces the motion artifact but the output images of MARC are too blurry or

Fig. 3: The simulated random motion artifact reduction results with the HCP brain dataset: (a) motion-corrupted input image, (b) MARC [16], (c) Cycle-MedGAN V2.0 [20], (d) bootstrap subsampling and aggregation [23], (e) the proposed method, and (f) motion-free label image. The difference maps show the difference between each image and the label image. PSNR and SSIM values of each image are shown in the corner of the images.

Fig. 4: The simulated respiratory motion artifact reduction results with the CNUH liver dataset: (a) motion-corrupted input image, (b) MARC [16], (c) Cycle-MedGAN V2.0 [20], (d) bootstrap subsampling and aggregation [23], (e) the proposed method, and (f) motion-free label image. The difference maps show the difference between each image and the label image. PSNR and SSIM values of each image are shown in the corner of the images.

smoothed (Fig. 3(b)). In the results of MARC, the boundary between gray matter and white matter is not clear (the first row in Fig. 3(b)), and the structure of the choroid plexus is not properly restored (the second row in Fig. 3(b)). Next, in Fig. 3(c) and (d), Cycle-MedGAN V2.0 [20] and bootstrap sub-

sampling and aggregation [23] remove random motion artifacts significantly and show increased quantitative results compared to input images. However, there are some differences between label images and outputs of Cycle-MedGAN V2.0 as shown in difference maps, and bootstrap subsampling and aggregation

[23] shows blurrier edge details compared to label images. On the other hand, as shown in Fig. 3(e), the proposed method shows the best qualitative and quantitative results among all methods. Especially, the proposed method shows the sharpest boundary between gray and white matters among methods as shown in the first row of Fig. 3.

Next, we compare motion artifact reduction methods using simulated respiratory motion-corruption data. In Fig. 4(a), the vasculature of the liver is damaged or blurred due to motion artifacts. Especially, artifacts appear most severe around blood vessels. MARC removes motion artifacts and achieves high quantitative metric values, but the blood vessels still look blurry as shown in Fig. 4(b). On the other hand, Cycle-MedGAN V2.0 [20] sharp reconstructed results but the PSNR of results of Cycle-MedGAN V2.0 is lower than that of input images (4(c)). It is maybe because Cycle-MedGAN V2.0 changes image intensity or details. Results of bootstrap subsampling and aggregation [23] are shown in Fig. 4(d), resulting in images with reduced motion artifacts and improved quantitative metrics compared to input images. However, some motion artifacts near the blood vessels remain (the first row of Fig. 4(d)), and it is hard to recognize the vessel due to blurring and remaining artifacts (the second row of Fig. 4(d)). Meanwhile, the proposed method shows the most similar restoration results to the label images as shown in Fig. 4(e) and (f). Specifically, the vascular structure is most clearly and accurately restored by the proposed method. Furthermore, our method significantly reduces motion artifacts around the blood vessels compared to other methods.

TABLE I shows the quantitative metric values of motion artifact reduction methods. In experiments using simulated random motion-corrupted data, the proposed method achieves the highest PSNR and SSIM, and it is consistent with the qualitative results in Figs. 3 and 4. On the other hand, MARC shows the highest quantitative results when using simulated respiratory motion-corrupted data. However, as confirmed in Figs. 3 and 4, reconstructed images by MARC are extremely blurred, so the detailed structures are indistinguishable. Compared to MARC, the proposed method removes the motion artifacts without losing information on image details. Furthermore, the quantitative metric value of our method is the highest among that of unpaired/unsupervised methods.

TABLE I: Quantitative results of various methods with simulated motion-corrupted data (Cycle: Cycle-MedGAN V2.0, BSA: Bootstrap Subsampling and Aggregation).

	Method	PSNR (dB)	SSIM
Brain Random motion	Input	27.83	0.751
	MARC [16]	29.29	0.891
	Cycle [20]	28.79	0.894
	BSA [23]	30.18	0.839
	Proposed	**31.40**	**0.916**
Liver Respiratory motion	Input	36.15	0.912
	MARC [16]	**37.87**	**0.947**
	Cycle [20]	35.54	0.926
	BSA [23]	36.45	0.932
	Proposed	37.01	0.940

Fig. 5: The in vivo motion artifact reduction results with the CNUH liver dataset: (a) motion-corrupted input image, (b) MARC [16], (c) Cycle-MedGAN V2.0 [20], (d) bootstrap subsampling and aggregation [23], and (e) the proposed method.

B. Results with In Vivo Data

Because the simulated motion artifacts only consider rigid motion artifacts, it should be verified that the method can also be applied to non-rigid in vivo motion artifact removal. In Fig. 5(a), motion artifacts due to transient dyspnea degrade the quality of liver MR image. We attempt to remove motion artifacts in Fig. 5(a), and results are shown in Fig. 5(b) to (e). Again, MARC removes not only motion artifacts but also detailed structures of blood vessels, so the reconstructed image is extremely blurry (Fig. 5(b)). Conversely, in Fig. 5(c), Cycle-MedGAN V2.0 makes the image sharper, but it also amplifies motion artifacts or noise in the input image. Next, the bootstrap subsampling and aggregation method also fails to remove the motion artifacts. Specifically, as shown in the yellow and green boxes of Fig. 5(d), motion artifacts around the blood vessels remain in the output image. Unlike comparison methods, the proposed method successfully removes the motion artifacts and reduces the noise level of the input image. Furthermore, our method reconstructs detailed structures. For example, in the yellow box of Fig. 5(e), the sharpness of the lesion increased as the motion artifact disappeared. Also, the vascular structure is recovered due to the reduction of motion artifacts as shown in the green box of Fig. 5(e). Through the experiment using in vivo motion-corrupted data, we confirmed that the proposed method also removes in vivo motion artifacts that contain the non-rigid motion of patients.

TABLE II: Clinical evaluation results of various methods with in vivo motion-corrupted data (average ± standard deviation) (Cycle: Cycle-MedGAN V2.0, BSA: Bootstrap Subsampling and Aggregation). Higher scores indicate higher performance.

Method	Motion artifact	Noise	Blurring	Vessel clarity	Overall quality
Input	3.03 ± 0.91	3.03 ± 0.68	3.92 ± 0.71	3.45 ± 1.20	3.00 ± 1.09
MARC [16]	3.37 ± 0.79	3.34 ± 0.81	2.29 ± 0.87	2.97 ± 1.15	2.50 ± 1.03
Cycle [20]	3.42 ± 0.92	3.13 ± 0.81	3.97 ± 0.94	3.47 ± 1.18	3.21 ± 1.09
BSA [23]	3.45 ± 1.22	3.39 ± 0.75	3.89 ± 1.06	3.45 ± 1.29	3.29 ± 1.18
Proposed	3.63 ± 1.10	3.58 ± 0.76	3.97 ± 0.91	3.71 ± 1.31	3.45 ± 1.25

C. Clinical Evaluation

Because it is impossible to quantitatively evaluate results using in vivo motion-corrupted datasets due to the lack of paired motion-free data, we evaluate motion artifact reduction results by clinical evaluation.

TABLE II shows the scores by evaluating each method on various criteria. MARC achieved scores of 3.37 and 3.34 in terms of motion artifact and noise evaluation, respectively, while input images score 3.03 in both evaluations. These results indicate that MARC was good in motion artifact improvement or noise reduction. However, MARC scored 2.29 in the blurring evaluation, which is lower than the score of input images (score: 3.92). The blurring effect of MARC also can be confirmed in Fig. 5(b). Therefore, the overall quality score of MARC (score: 2.50) is lower than that of input images (score: 3.00). On the other hand, Cycle-MedGAN V2.0 got the highest score in the blurring evaluation (score: 3.97). However, Cycle-MedGAN V2.0 scored 3.13 in noise evaluation, which is lower than the scores of other methods. This high level of noise affects the image quality drop of

Cycle-MedGAN V2.0, so Cycle-MedGAN V2.0 gets only 3.29 points in terms of the overall image quality. As shown in Fig. 5 and TABLE II, the bootstrap subsampling and aggregation method shows higher scores than the other existing methods in most assessments. However, the outputs of the bootstrap subsampling and aggregation method were slightly blurred, so its score was lower than the input images in the blurring evaluation.

While the other methods each showed drawbacks, the proposed method achieved the highest performance in all evaluations. First, in terms of motion artifact removal, the proposed method achieves the highest score (score: 3.63) while other methods get similar lower scores (score: 3.37-3.45). Next, our method scored 3.58 and 3.97 in the noise and blurring evaluations, respectively. From these results, we confirm that our method does not amplify image noise level or blur output images through the clinical evaluation. Moreover, the motion-corrupted input images scored 3.45 in terms of vessel clarity. The proposed method shows a significant improvement in vessel clarity score (score: 3.71) while the vessel clarity of the other three methods is similar to or lower than that of motion-corrupted input images (score: 2.97-3.47). Finally, our method gets the best score (score: 3.45) for overall image quality. To sum up, the proposed method achieves the highest score in all clinical evaluations, and this result indicates that our method is useful in clinical practice.

VI. DISCUSSION

A. Comparison with Other Methods

In Section V, it was verified that MARC [16] generates blurry outputs in both simulation and in vivo study. The blurring results may be a limitation of methods based on supervised learning. Because the supervised learning minimizes the loss (e.g. L1, mean squared error (MSE)) between output and label, it achieves high quantitative results as shown in TABLE I. However, it can also lead to the loss of information on image details because L1 or MSE losses do not assure the perceptual quality of output images.

Unlike MARC, Cycle-MedGAN V2.0 [20] is an unpaired method that does not require paired input and label images. Instead of using losses between input and label, it translates an image from one domain to another domain by utilizing cycle consistency loss and adversarial loss. Because the discriminators of Cycle-MedGAN V2.0 distinguish real and fake generated images, the generators of Cycle-MedGAN V2.0 provide realistic images with sharp details. However, we have confirmed that Cycle-MedGAN V2.0 also magnifies the artifacts or noise of images. We conjecture that it is because the networks of Cycle-MedGAN V2.0 consider resolution degradation due to the motion artifacts to be the main difference between the two image domains. Therefore, the networks of Cycle-MedGAN V2.0 try to improve resolution rather than eliminate motion artifacts.

Compared to the previous two methods, bootstrap subsampling and aggregation [23] showed stable qualitative and quantitative results. Nevertheless, because [23] works under the assumption that the motion artifact appears as sparse

TABLE III: Ablation studies on hyperparameters with simulated respiratory motion-corrupted data. The gray rows indicate the hyperparameters that are selected in our experiments.

Hyperparameters		PSNR (dB)	SSIM	Time/image (sec)
$\lambda_{N'}$	0	36.36	0.935	19.30
	0.01	**37.01**	**0.940**	19.30
	0.1	36.58	0.927	19.30
N'	1	36.45	0.935	1.834
	10	**37.01**	**0.940**	19.30
	100	36.88	0.938	195.6
M	1	36.43	0.934	6.358
	3	37.01	0.940	19.30
	5	**37.28**	**0.942**	32.48
$N' \times M$	10×3	**37.01**	**0.940**	19.30
	30×1	36.90	0.938	19.30

outliers in k-space, the performance of this method is degraded if the assumption is not satisfied. For example, we simulated the respiratory motion with Eq. (15), so the respiratory motion appears as a continuous sinusoidal form in k-space. Because the motion did not appear as sparse outliers, the performance of [23] was dropped compared to when it works with simulated random motion-corrupted data.

On the other hand, our proposed method presented outstanding results compared to other comparison methods. The proposed method successfully removes motion artifacts and retrieves high-frequency image details in both simulation and in vivo studies.

Nevertheless, our method is not free of limitations. Because the score-based diffusion models require several steps of reverse diffusion, it takes a long time to generate outputs. Although we utilized the CCDF algorithm to reduce the inference time, our method also requires several seconds as shown in TABLE III. Therefore, the acceleration of the proposed method should be done for clinical use.

B. Effects of Annealing Hyperparameters

In our method, we injected high-frequency components of measurements (k-space of motion-corrupted images) with the hyperparameter $\lambda_{N'}$ to preserve detailed structures of MR images. To confirm the effect of high-frequency component injection, we conduct our method for simulated liver motion-corrupted images with various $\lambda_{N'}$. As shown in TABLE III, the proposed method with $\lambda_{N'} = 0$ shows lower quantitative results than the proposed method with $\lambda_{N'} = 0.01$. It is because detailed structures such as vessels cannot be reconstructed perfectly without high-frequency component injection. When $\lambda_{N'} = 0.1$, the quantitative results drop again compared to results with $\lambda_{N'} = 0.01$. We conjecture that it is because the high-frequency component of measurements also contains motion artifacts, and the remaining artifacts degrade the quality of reconstructed images. Therefore, we choose to inject high-frequency components with $\lambda_{N'} = 0.01$ in our experiments.

Next, we also confirm the effect of the selection of N'. When $N' = 1$, the motion artifacts remain in output images, so the quantitative results deteriorate. On the other hand, our method also shows the degraded performance when $N' = 100$. It may be because the structures that cannot be seen in the

input image were generated during the iterations of the reverse diffusion process. Moreover, the required inference time of the proposed method with $N' = 100$ is quite long as shown in TABLE III, so we choose $N' = 10$ that shows the best qualitative and quantitative performance.

Finally, the number of iterations of the reverse diffusion process M is also one of the important hyperparameters of our method. Through the experiments on M, we find that the proposed method cannot completely remove motion artifacts when $M = 1$. On the other hand, when $M = 5$, the required inference time for one image is too long while the performance gain is negligible compared to when $M = 3$. Therefore, $M = 3$ is selected in our experiments.

In addition, we also verify the effect of the combination of N' and M. The proposed method shows different results depending on the combination of N' and M as shown in TABLE III, even if it takes the same inference time. The proposed method with $N' = 30$, $M' = 1$ shows lower quantitative performance compared to the method with $N' = 10$, $M' = 3$. It is because the motion artifacts cannot be removed perfectly with only one iteration of the diffusion process even though N' is large. Through the experiment, we verify that the combination of $N' = 10$, $M' = 3$ is better than $N' = 30$, $M' = 1$ for the performance of our proposed method.

VII. CONCLUSION

In this paper, we proposed a novel MRI motion artifact reduction method using the annealed score-based diffusion model. By applying the diffusion process iteratively and gradually imposing data consistency with high-frequency injection, the proposed method successfully reduced simulated and in vivo motion artifacts in MR images. Furthermore, we verified that our method provides higher-quality images and more clinical meaning compared to other state-of-the-art deep learning methods. We believe that our algorithm can be a useful framework for MRI motion artifact reduction.

VIII. ACKNOWLEDGEMENT

REFERENCES

[1] A. Nishie, D. Kakihara, Y. Asayama, Y. Ushijima, Y. Takayama, N. Fujita, D. Shimamoto, K. Shirabe, T. Hida, and H. Honda, "Detectability of hepatocellular carcinoma on gadoxetic acid-enhanced MRI at 3T in patients with severe liver dysfunction: clinical impact of dual-source parallel radiofrequency excitation," *Clinical radiology*, vol. 70, no. 3, pp. 254–261, 2015.

[2] N. Verloh, K. Utpatel, M. Haimerl, F. Zeman, C. Fellner, S. Fichtner-Feigl, A. Teufel, C. Stroszczynski, M. Evert, and P. Wiggermann, "Liver fibrosis and Gd-EOB-DTPA-enhanced MRI: A histopathologic correlation," *Scientific reports*, vol. 5, no. 1, pp. 1–10, 2015.

[3] K. Kubota, T. Tamura, N. Aoyama, M. Nogami, N. Hamada, A. Nishioka, and Y. Ogawa, "Correlation of liver parenchymal gadolinium-ethoxybenzyl diethylenetriaminepentaacetic acid enhancement and liver function in humans with hepatocellular carcinoma," *Oncology letters*, vol. 3, no. 5, pp. 990–994, 2012.

[4] M. S. Davenport, B. L. Viglianti, M. M. Al-Hawary, E. M. Caoili, R. K. Kaza, P. S. Liu, K. E. Maturen, T. L. Chenevert, and H. K. Hussain, "Comparison of acute transient dyspnea after intravenous administration of gadoxetate disodium and gadobenate dimeglumine: effect on arterial phase image quality," *Radiology*, vol. 266, no. 2, pp. 452–461, 2013.

[5] N. White, C. Roddey, A. Shankaranarayanan, E. Han, D. Rettmann, J. Santos, J. Kuperman, and A. Dale, "Promo: real-time prospective motion correction in MRI using image-based tracking," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 63, no. 1, pp. 91–105, 2010.

[6] N. Todd, O. Josephs, M. F. Callaghan, A. Lutti, and N. Weiskopf, "Prospective motion correction of 3d echo-planar imaging data for functional mri using optical tracking," *Neuroimage*, vol. 113, pp. 1–12, 2015.

[7] J.-R. Liao, J. M. Pauly, T. J. Brosnan, and N. J. Pelc, "Reduction of motion artifacts in cine MRI using variable-density spiral trajectories," *Magnetic resonance in medicine*, vol. 37, no. 4, pp. 569–575, 1997.

[8] J. G. Pipe, "Motion correction with PROPELLER MRI: application to head motion and free-breathing cardiac imaging," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 42, no. 5, pp. 963–969, 1999.

[9] S. S. Vasanawala, Y. Iwadate, D. G. Church, R. J. Herfkens, and A. C. Brau, "Navigated abdominal T1-W MRI permits free-breathing image acquisition with less motion artifact," *Pediatric radiology*, vol. 40, no. 3, pp. 340–344, 2010.

[10] M. Usman, D. Atkinson, F. Odille, C. Kolbitsch, G. Vaillant, T. Schaeffter, P. G. Batchelor, and C. Prieto, "Motion corrected compressed sensing for free-breathing dynamic cardiac MRI," *Magnetic resonance in medicine*, vol. 70, no. 2, pp. 504–516, 2013.

[11] Z. Yang, C. Zhang, and L. Xie, "Sparse MRI for motion correction," in *2013 IEEE 10th International Symposium on Biomedical Imaging*. IEEE, 2013, pp. 962–965.

[12] K. H. Jin, J.-Y. Um, D. Lee, J. Lee, S.-H. Park, and J. C. Ye, "MRI artifact correction using sparse+ low-rank decomposition of annihilating filter-based hankel matrix," *Magnetic resonance in medicine*, vol. 78, no. 1, pp. 327–340, 2017.

[13] B. A. Duffy, W. Zhang, H. Tang, L. Zhao, M. Law, A. W. Toga, and H. Kim, "Retrospective correction of motion artifact affected structural MRI images using deep learning of simulated motion," 2018.

[14] I. Oksuz, J. Clough, B. Ruijsink, E. Puyol-Antón, A. Bustin, G. Cruz, C. Prieto, D. Rueckert, A. P. King, and J. A. Schnabel, "Detection and correction of cardiac mri motion artefacts during reconstruction from k-space," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2019, pp. 695–703.

[15] J. Liu, M. Kocak, M. Supanich, and J. Deng, "Motion artifacts reduction in brain MRI by means of a deep residual network with densely connected multi-resolution blocks (DRN-DCMB)," *Magnetic resonance imaging*, vol. 71, pp. 69–79, 2020.

[16] D. Tamada, M.-L. Kromrey, S. Ichikawa, H. Onishi, and U. Motosugi, "Motion artifact reduction using a convolutional neural network for dynamic contrast enhanced MR imaging of the liver," *Magnetic resonance in medical sciences*, vol. 19, no. 1, p. 64, 2020.

[17] Q. Lyu, H. Shan, Y. Xie, A. C. Kwan, Y. Otaki, K. Kuronuma, D. Li, and G. Wang, "Cine cardiac MRI motion artifact reduction using a recurrent neural network," *IEEE Transactions on Medical Imaging*, vol. 40, no. 8, pp. 2170–2181, 2021.

[18] M. A. Al-Masni, S. Lee, J. Yi, S. Kim, S.-M. Gho, Y. H. Choi, and D.-H. Kim, "Stacked U-Nets with self-assisted priors towards robust correction of rigid motion artifact in brain MRI," *NeuroImage*, vol. 259, p. 119411, 2022.

[19] E. Kuzmina, A. Razumov, O. Y. Rogov, E. Adalsteinsson, J. White, and D. V. Dylov, "Autofocusing+: Noise-resilient motion correction in magnetic resonance imaging," *arXiv preprint arXiv:2203.05569*, 2022.

[20] K. Armanious, A. Tanwar, S. Abdulatif, T. Küstner, S. Gatidis, and B. Yang, "Unsupervised adversarial correction of rigid MR motion artifacts," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2020, pp. 1494–1498.

[21] S. Liu, K.-H. Thung, L. Qu, W. Lin, D. Shen, and P.-T. Yap, "Learning MRI artefact removal with unpaired data," *Nature Machine Intelligence*, vol. 3, no. 1, pp. 60–67, 2021.

[22] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[23] G. Oh, J. E. Lee, and J. C. Ye, "Unpaired MR motion artifact deep learning using outlier-rejecting bootstrap aggregation," *IEEE Transactions on Medical Imaging*, vol. 40, no. 11, pp. 3125–3139, 2021.

[24] Y. Song and S. Ermon, "Generative modeling by estimating gradients of the data distribution," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[25] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in Neural Information Processing Systems*, vol. 33, pp. 6840–6851, 2020.

[26] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," in *International Conference on Learning Representations*, 2020.

[27] Y. Song, L. Shen, L. Xing, and S. Ermon, "Solving inverse problems in medical imaging with score-based generative models," *arXiv preprint arXiv:2111.08005*, 2021.

[28] H. Chung, B. Sim, and J. C. Ye, "Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 12 413–12 422.

[29] H. Chung, B. Sim, D. Ryu, and J. C. Ye, "Improving diffusion models for inverse problems using manifold constraints," in *Advances in Neural Information Processing Systems*, A. H. Oh, A. Agarwal, D. Belgrave, and K. Cho, Eds., 2022. [Online]. Available: https://openreview.net/forum?id=nJJjv0JDJju

[30] H. Chung and J. C. Ye, "Score-based diffusion models for accelerated MRI," *Medical Image Analysis*, p. 102479, 2022.

[31] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, "Image super-resolution via iterative refinement," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[32] B. D. Anderson, "Reverse-time diffusion equation models," *Stochastic Processes and their Applications*, vol. 12, no. 3, pp. 313–326, 1982.

[33] P. Vincent, "A connection between score matching and denoising autoencoders," *Neural computation*, vol. 23, no. 7, pp. 1661–1674, 2011.