Active CT Reconstruction with a Learned Sampling Policy

Ce Wang†
Institute of Computing Technology, CAS
Beijing, China
Suzhou Institute of Intelligent Computing Technology, CAS
Suzhou, Jiangsu, China
foguever@icloud.com

Kun Shang†
Shenzhen Institute of Advanced Technology, CAS
Shenzhen, Guangdong, China
kun.shang@gmail.com

Haimiao Zhang
Beijing Information Science and Technology University
Beijing, China
hmzhang@bistu.edu.cn

Shang Zhao
School of Biomedical Engineering, Division of Life Sciences, USTC
Suzhou Institute for Advanced Technology, USTC
Suzhou, Jiangsu, China
shangzhao@ustc.edu.cn

Dong Liang
Shenzhen Institute of Advanced Technology, CAS
Shenzhen, Guangdong, China
dong.liang@siat.ac.cn

S. Kevin Zhou†
School of Biomedical Engineering, Division of Life Sciences, USTC
Beijing, China
Suzhou Institute for Advanced Technology, USTC
Suzhou, Jiangsu, China
skevinzhou@ustc.edu.cn

ABSTRACT
Computed tomography (CT) is a widely-used imaging technology that assists clinical decision-making with high-quality human body representations. To reduce the radiation dose posed by CT, sparse-view (SV) CT is developed with preserved image quality. However, these methods are still stuck with a fixed uniform SV (USV) sampling strategy, which inhibits the possibility of acquiring a better image with an even reduced dose. In this paper, we explore this possibility via learning an active SV (ASV) sampling policy that optimizes the sampling positions for regions of interest (RoI)-specific, high-quality reconstruction. To this end, we design an sampling agent for the recommendation of ASV sampling positions based on on-the-fly reconstruction with obtained sinograms in a progressive fashion. With such a design, we achieve better performances on the NIH-AAPM dataset over popular USV sampling, especially when the number of views is small. Finally, such a design enables the RoI-aware reconstruction with improved local quality within the RoI that are clinically important. Experiments on the VerSe dataset demonstrate the ability of the proposed sampling policy, which is difficult to achieve with USV sampling.

CCS CONCEPTS
• Computing methodologies → Reconstruction.

†Both authors contributed equally to this research.
†The corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Proceedings of the 31st ACM International Conference on Multimedia (MM ’23), October 29–November 3, 2023, Ottawa, ON, Canada.
© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-8-4007-0108-5/23/10... $15.00
https://doi.org/10.1145/3581783.3611746

1 INTRODUCTION
Computed tomography (CT) reconstructs highly-detailed, cross-sectional maps of an object from X-ray sinograms, benefiting effective clinical diagnosis. Unfortunately, the carcinogenic nature of ionizing radiation means one needs to reduce the radiation dose that the patient is exposed to in a CT scan, which raises safety concerns for patients. In contrast, the lower dose means noisier data, and degrades the imaging quality. This increases the difficulty of diagnosis, even misdiagnosis, causing a more challenging imaging problem. To balance the demand of two aspects, sparse-view (SV) CT is developed [28, 42], which reconstructs images with a few uniform sparse-view (USV) sampling sinograms.

Traditional knowledge-based methods iteratively reconstruct an image from USV acquisitions with representative priors, such as Total Variation based methods [18, 24, 34], Non-Local based methods [33, 38], Wavelet-based methods [8, 9, 20], and Low-Rank-based methods [10, 15, 22, 40]. These approaches typically introduce an iterative reconstruction process that is computationally expensive. Then, the Deep Learning (DL) based algorithms [4, 5, 13, 14, 32, 36, 41, 43] have been proposed and achieved satisfactory performances with less time consumption, which directly learn a mapping function between image pairs. While, the procedure introduces a risk of falling into the local minima of training data, thereby being unable to generalize to unseen signals. To improve the generalizability, a series of deep-unfolding frameworks are proposed [1, 4, 6, 32, 37, 39] by further combining the iterative...
algorithms with DL, which inherits both the generalizability of knowledge-based methods and the high-performances of DL-based methods and hence boosts the reconstruction performances by a large margin.

Despite the success achieved by the above USV-based imaging methods, they ignore the differences across patients with different anatomy representations, making the commonly-used USV sampling non-adaptive [23]. Targeting an adaptive sampling policy, PLANet[35] proposes to pre-learn the sampling policy through the optimization of the parameterized Gaussian Mixture distribution on the training data. Although the method provides a better sampling policy on the whole data rather than USV, the offline process can hardly handle the individual factors across patients. Then UF-AEC [23] marks the first attempt that simultaneously learns the policy via reinforcement learning (RL) and the reconstruction with PD-Net [1] in an on-the-fly SVCT imaging progress. Such a design learns the policy and tunes the dose distributed along with each X-Ray signal, rendering better reconstruction performances.

In order to promote the contribution of CT imaging to intelligent treatment, further taking into account region of interest (RoI)-specific diagnosis requirements in the reconstruction would provide better visual quality in clinically concerned regions [17, 31]. Specifically, patients with spinal diseases need better image quality in the spine region for diagnosis; whereas COVID-19 examination necessitates more details within lung reconstruction. Although the above-mentioned UF-AEC provides overall better reconstructions than USV, they still fail to present the required image quality granularity in RoI across patients from specific diagnosis requirements, making the method clinically ineffective for personalized clinical treatment. In addition, a physically developed method [7, 12] for CT imaging uses a controllable beam filter to impose a high radiation dose only in the RoI, resulting in a better RoI reconstruction. However, the dose distributed inside and outside RoI needs to be carefully balanced for different demands. Till now, to the best of our knowledge, there are few works considering this problem from an algorithmic perspective.

In this paper, we incorporate the active SV (ASV) sampling into our reconstruction framework, which learns to jointly acquire sinograms and reconstruct the image, and therefore, adapts to the clinical RoIs of individual patients. Specifically, we separately learn the two subtasks with two corresponding modules, one for learning the ASV sampling policy, named Sampling Agent (SA), and the other for the reconstruction, named Reconstructor (R). Facing no ground truth for SA, we pretrain SA to be capable of ranking sinograms in terms of the contribution to CT reconstructions via self-supervised learning and a proposed metric learning-based loss function. Then, the pretrained SA is incorporated to recommend the most reconstruction-benefiting sinogram sampling position. Next, we design an iterative reconstruction process to gradually (i) rank all projected sinograms of the current reconstruction with SA and select the candidates with higher scores; and (ii) reconstruct with sinograms composed of the previous stages and the selected candidates in (i). To compensate for the optimization shift between learning better ranking with SA and better reconstruction with R, we employ an alternating optimization design. Via sufficient training, our method learns to actively obtain sinograms benefiting high-quality reconstructions and is able to go beyond what can be done by doing either separately.

In sum, our contributions are as follows:

- We propose a method to jointly learn ASV sampling policy and reconstruction. Targeting this, we convert the continuous optimization of sampling policy learning into a discrete sampling position selection problem. The learned sampling policy adaptively select sinograms which contributes more to the final reconstruction, and reduces radiation dose when achieving comparable reconstruction quality with USV.
- We propose an RoI-aware reconstruction framework by additionally bridging the interaction between RoI-specific information and our sampling policy learning process. Such RoI-aware imaging makes it convenient to adapt to various clinical downstream tasks.
- Empirical experiments on NIH-AAPM [19] and VerSe [21] benchmark datasets demonstrate that our learning-based policy achieves better reconstruction on both the whole image and the predefined RoI. The visualization of the learned sampling mask verifies the adaptiveness and its effectiveness in the interaction with the RoI information.

## 2 BACKGROUND AND MOTIVATION

CT reconstruction aims to reconstruct structural representations of external/internal tissues of the human body \( u \in \mathbb{R}^{H \times W} \) (H and W are the image height and width) from the corresponding sinogram \( y_f \in \mathbb{R}^{TD} \) (\( T \) is the sampling times, and \( D \) is the number of detector photons), which is formulated as

\[
y_f = A_f u + n,
\]

where the forward projection matrix \( A_f \in \mathbb{R}^{TD \times HW} \) represents the full-sampling Radon transform and \( n \) denotes the imaging noise. With USV sampling, the sampling times \( M \) is much less than \( T \), such that we define an additional sub-sampling matrix \( P \in \mathbb{R}^{MD \times TD} \).
composed of \( \{0, 1\} \) elements to choose suitable subvectors from \( y_f \). Thus we have the down-sampled observation \( y = Py_f \).

In iterative methods, DL methods, or deep-unrolling methods, the sampling matrix \( P \) is pre-defined, and the included trajectory is uniformly distributed around the patient (usually limited in 180º since the penetrability of X-Ray leads to repetitive information). This evenly-acquired sinogram collects global body representation and brings overall satisfactory reconstruction. However, such a predefined sampling policy does not consider task-specific characteristics. Taking spine checking as a motivating example in Fig. 1, we here only consider three partial sinograms within a fixed angle range (labeled with curves) to simplify the problem. With the USV trajectory (three sinograms distributed equally within the range) as in Fig. 1 (a), the Region of Cross (RoC) formed by acquired sinograms does not bear a significant overlap with the RoI. Indeed, this is not expected for the shown CT image with the RoI (spine parts) deviating a lot from the center, resulting in the reconstruction that seems overall clear but with degraded quality in the RoI. The phenomena make the deployed USV sampling less effective when clinicians require to examine the patient’s spinal region.

To avoid such undesirable properties in USV, additional optimization of the sampling matrix \( P \) is necessary, which provides a potential to achieve a personalized and clinically effective imaging process. Recall the above example, with a rearrangement of the positions as in Fig. 1 (b), the resulted RoC highly overlaps with the RoI. Motivated by the insignificant overlap between RoC and RoI in USV and the RoC changes with rearrangement, we explore to design an ASV sampling policy as in Fig. 1 (d), which gradually learns the sampling matrix \( P \) and reconstruction simultaneously, to achieve a suitable arrangement of the sampling policy for corresponding clinical requirements.

With technology advancement, dynamically moving the radioactive source is realizable. Thus, further optimizing \( P \) for a better reconstruction has recently become attractive yet challenging. Targeting this issue, we present a controllable policy, and learn the optimal \( P \) simultaneously with reconstruction to impose the positive interaction, which better aligns with clinical scenarios.

3 METHOD

To this end, we reformulate the reconstruction problem as follows:

\[
\min_{u, P} \frac{1}{2} \| P A_f u - y \|^2 + R_1(u) + R_2(P),
\]

where \( R_1(u) \) and \( R_2(P) \) are the regularization terms used to impose the prior information. Traditionally, the minimization problem (1) can be efficiently solved via the splitting method. The variables \( u \) and \( P \) in the above are split into two blocks:

\[
\begin{align*}
    u^{k+1} &= \arg\min_u \frac{1}{2} \| PA_f u^{k} - y \|^2 + R_1(u), \\
    p^{k+1} &= \arg\min_P \frac{1}{2} \| PA_f u^{k+1} - y \|^2 + R_2(P).
\end{align*}
\]

Rethinking the above-introduced \( P \) as in Fig. 2, \( P \) is a very high-dimensional and sparse matrix. More specifically, it’s composed of \( M \times T \) submatrices, which are either an identity or all-zero matrix. Due to its special form, direct optimization of \( P \) in (2) is difficult and computationally expensive. To speed up the iterative reconstruction, recall the physical significance of \( P \). For the fixed fully sampled sinogram \( y_f \in \mathbb{R}^{TD} \), \( P \) adopts \( M \)-times (with \( M \ll T \)) sampling on the temporal dimension, which is represented by the identity matrix in Fig. 2, and each time a projection \( y_{f,k} \in \mathbb{R}^{D}(1 \leq k \leq T) \) is returned. In practice, \( y_{f,k} \) represents one imaging line on the detector (i.e., the D-element subvector in Fig. 2). With such analysis, we simplify the optimization of \( P \) to the iterative selection of \( M \)-times choice of identity matrix in our framework, that is, adding one more imaging line per iteration. In this way, it becomes a discrete selection rather than a continuous optimization problem.

For this purpose, we use the neural networks to learn the solutions of Eq. (2) in each iteration, which gives

\[
\begin{align*}
    \text{Reconstruction step} : \quad & u^{k+1} = u^k + R_\theta(u^k, p^k), \\
    \text{Acquisition step} : \quad & p^{k+1} = \mathbf{SA}_\phi(u^{k+1}, p^k),
\end{align*}
\]

where \( R_\theta \) and \( \mathbf{SA}_\phi \) are parameterized sub-networks to iteratively compute the intermediate estimations. In this way, we have decoupled the optimization problem (1) into the acquisition step and reconstruction step, where the alternating direction method is employed to solve the multi-variable optimization problem. The overall active reconstruction ↔ acquisition process is depicted in Fig. 3. The complete process is alternatively between the optimization of Reconstructor (\( \mathbf{R} \)) and Sampling Agent (\( \mathbf{SA} \)). Especially, we first pre-define \( k_0 \) initial sinogram-sampling positions (we empirically set initial positions uniformly in experiments) to give a suitable initialization of \( u^0 \) and \( p^0 \). According to the current image quality, \( \mathbf{SA} \) is then utilized to recommend the next \( k \) acquisition positions within a certain angle range in the acquisition step. The proposed iterative process outputs a series of iterates:

\((u^1, p^1), (u^2, p^2), (u^3, p^3), \ldots\)
and continues until the total number of acquired sinograms equals $k_{\text{max}}$, which is the pre-defined in experiments. Next, we introduce the modules $R$, $SA$, and the training paradigm in detail.

### 3.1 Reconstructor

From the perspective of active learning, the obtained $k_0 + nk$ sinograms after $n$-time iterations provide partial and low-quality representation, which can be utilized to suggest the next sinogram acquisition positions. To ensure the power of the latter sampling policy design, we realize $R$ with the commonly-used backbone U-Net, which can also be replaced with other structures. In the iterative sampling and reconstruction process, $R$ is used to transit estimations from different acquisition scenarios, and the following $SA$ would evaluate the correlation between each candidate and the final reconstruction performance. Specifically, in the $n$-th iteration, the reconstructor $R$ receives $k_0 + nk$ acquired sinograms $y_n$, and outputs the reconstruction $\hat{u}_n$. Towards an efficient computation, we employ the same parameters among each iteration, which also renders better robustness of our reconstructor. The final loss function for $R$ is defined as follows:

$$L_R = ||(1 + M) \odot (\hat{u} - u_{gt})||_2^2,$$

where $I$ is the identity matrix, $M$ is a pre-defined 0-1 mask representing the coarse RoI, $\odot$ is the Hadamard product, and $\hat{u}$ and $u_{gt}$ are the final estimation and the corresponding ground-truth CT image, respectively. With the additional guidance of $M$, the interaction between $R$ and $SA$ results in images that pay more attention to RoI reconstruction quality with the help of rearrangement of sinogram positions. This is especially for a constrained radiation budget: the Region of Cross (RoC) is more likely to match with the RoI.

### 3.2 Sampling Agent

Given an estimated reconstruction, an experienced clinician is capable of deciding which part of the reconstruction is sufficient for diagnosis, and therefore suggests a coarse sampling range to enhance the reconstruction. Nevertheless, such an expert-dominated sampling suggestion is expensive because of the scarcity of clinicians. A desirable substitution is a Sampling Agent (SA) able to seek out the most reconstruction-benefited sampling positions with the current estimated $\hat{u}_n$. But this is an extremely difficult problem since the image domain information cannot directly guide the sinogram position searching. To solve the problem, we propose a two-stage projection-image domain correlation sampling policy, denoted by $SA$, as in Fig. 3. Specifically, with the current estimation, we first transform $\hat{u}_n$ into the Radon domain with $A_f$. Obtaining the current full-projections $A_f(\hat{u}_n)$, we next use a fully-connected network to output a confidence score for each sinogram, which tells the system how each obtained projection, i.e., $\{A_f(\hat{u}_n)\}_{i=1}^{k_{\text{max}}}$ of the current reconstruction, correlates with the final reconstruction. Targeting such a key point, we have designed a self-supervised strategy for the position sampling process since there are no ground-truth sampling positions for each CT image in practical reconstruction. Firstly, we compute the reliability of the current projection position with the following metric:

$$A_{r_d,i} = \exp(-||A_f(\hat{u}_n)_i - A_f(u_{gt})||^2).$$

where $A_f$ is the fully-sampled transform matrix, $\hat{u}_n$ is the current reconstruction with currently-sampled sinograms, and $u_{gt}$ is the corresponding ground-truth CT image. The metric indeed softly evaluates how the current sinograms match with ground-truth sinograms, and would assign $\rightarrow 1$ to the most approximated one and $\rightarrow 0$ to the fakest one. With the computed $A_{r_d,i}$ as the supervision of SA, our SA learns to select sinograms closest to ground-truth ones, which benefits the reconstruction mostly. The final loss function for the evaluator is as follows:

$$L_{SA} = \| \sum_i SA(\hat{u}_n)_i - A_{r_d,i} \|^2.$$
Table 1: Quantitative comparisons of our SAS and GDS policies with RS and US. Reconstructor performance with our policies achieve significant improvement with $p < 0.001$ in terms of PSNR when $15 < k_{max} < 60$. For $k_{max} = 90$, where sampling freedom is very limited, our policies achieve significant improvement with $p < 0.05$ in terms of PSNR. The best performance in each column highlighted in bold and the second best is underlined.

| NIH-AAPM | $k_{max} = 15$ | $k_{max} = 30$ | $k_{max} = 60$ | $k_{max} = 90$ |
|----------|----------------|----------------|----------------|----------------|
| PSNR | SSIM | RMSE | PSNR | SSIM | RMSE | PSNR | SSIM | RMSE | PSNR | SSIM | RMSE |
| RS | 23.17 ± 0.81 | 0.793 ± 0.029 | 0.070 ± 0.004 | 25.85 ± 0.106 | 0.836 ± 0.012 | 0.053 ± 0.008 | 28.44 ± 0.309 | 0.858 ± 0.015 | 0.053 ± 0.008 | 29.40 ± 0.319 | 0.873 ± 0.014 | 0.034 ± 0.005 |
| US | 24.98 ± 0.59 | 0.823 ± 0.017 | 0.057 ± 0.004 | 27.89 ± 0.59 | 0.845 ± 0.017 | 0.041 ± 0.003 | 30.79 ± 0.78 | 0.874 ± 0.016 | 0.032 ± 0.003 | 31.22 ± 0.57 | 0.885 ± 0.016 | 0.028 ± 0.002 |
| SAS (ours) | 26.16 ± 0.38 | 0.829 ± 0.018 | 0.047 ± 0.003 | 28.69 ± 0.57 | 0.855 ± 0.019 | 0.034 ± 0.003 | 30.02 ± 0.78 | 0.875 ± 0.016 | 0.032 ± 0.003 | 30.58 ± 0.59 | 0.882 ± 0.018 | 0.028 ± 0.002 |
| GDS (ours) | 25.69 ± 0.71 | 0.829 ± 0.018 | 0.047 ± 0.003 | 28.47 ± 0.66 | 0.851 ± 0.019 | 0.034 ± 0.003 | 29.84 ± 0.53 | 0.871 ± 0.017 | 0.032 ± 0.002 | 31.00 ± 0.68 | 0.884 ± 0.018 | 0.028 ± 0.002 |

3.3 Training strategy
As described above, the included R and SA target different functions: one for reconstruction and the other for selecting reliable sinograms. Simultaneously optimizing them together makes it confusing for the whole system. Motivated by the optimization procedure in GANs [3, 11, 26], we propose to optimize them in an alternative fashion (the specific algorithm is depicted in Appendix A) to compensate for the optimization direction shift. Concretely, we first train R for 2 epochs with fixed SA since the training of such reconstruction is much easier than SA. Then, with fixed R, we optimize SA to search for a better sampling trajectory that is most suitable for the current R. With sufficient training, the two modules would cooperate, targeting a better reconstruction.

3.4 Active sampling and reconstruction
With the well-trained R and SA, the testing phase utilizes them to simultaneously sample a suitable trajectory and output a better reconstruction. Here the hyperparameters in testing are denoted with a superscript to distinguish them from training ones. Specifically, with pre-defined $\alpha_{max}$ and $k_{max}$ according to the physical setting, we need to additionally tune hyperparameters $k_g$, $k'$, $\alpha_p$ and $\alpha_q'$ to strive for the best one. Note that we would make these hyperparameters different in training and testing since (i) we release more searching freedom to the algorithm when training to help the model face more cases, and (ii) we tighten the searching space in testing to eliminate poor entities with empirical experience.

4 EXPERIMENTAL RESULTS

Dataset. We use the “2016 NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge” (AAPM) [19] dataset and VerSe [21] dataset in the experiments. The former is used to verify the active reconstruction framework and ablation experiments, and the latter is used for Roll-aware reconstruction. For the AAPP dataset, we follow the original data partition with three anatomies, including the chest, abdomen, and brain. We first conduct ablation experiments on chest CT scans to confirm the advantage over USV sampling policy, followed by ablation experiments about the hyperparameters choices, and the robustness on the other two anatomies. To further demonstrate the deployment flexibility of the framework with known clinical tasks, we experiment with VerSe volumes, which cover the whole spine of patients, to further reconstruct and optimize sampling policy for better recovery within the ROI.

Implementation details. Models are implemented using the PyTorch framework. We use the Adam optimizer [16] with ($\beta_1$, $\beta_2$) = (0.9, 0.999) to train these models. The learning rate starts from 0.0001 for the R and 0.0002 for the SA. Models are all trained on an NVIDIA 3090 GPU card for 50 epochs with a batch size of 1. All the sinograms are generated in a fan-beam CT geometry using ODL [2], and the number of full-view is set to 360. The CT image resolution is 512 × 512.

Evaluation metrics. Reconstructed CT images are quantitatively measured by the SSIM [29, 30], PSNR, and Root MSE(RMSE), and we compute the RMSE on normalized images.

Comparison methods. To verify the effectiveness of our learning-based ASV sampling policy, we compare it with Uniform Sampling (US) and Random Sampling (RS). For RS, we infer with the trained model, followed by hyperparameters chosen. Then, we conduct experiments on additional two anatomies to confirm the robustness.

4.1 Ablation Studies and Analysis
To verify the effectiveness of the proposed ASV sampling policies, SAS and GDS, we compare them with US (which is indeed a special case in ASV) and RS on chest CT scans of the NIH-AAPM dataset, followed by hyperparameters chosen. Then, we conduct experiments on additional two anatomies to confirm the robustness.
4.1.1 Effectiveness of ASV sampling policies. Firstly, we quantitatively evaluate the effectiveness of our proposed SAS and GDS policies, where the introduced $M$ in Eq. 4 is set to zero matrix since we don’t utilize RoI-specific information. The results are reported in Table 1. Reconstructions with SAS and GDS achieve significant improvement across different settings except when $k_{\text{max}}=90$, where GDS policy is a little worse or comparable with the US since the restricted searching space limits its various hyperparameter selection (i.e., the performances). Especially, the improvements over the US become enlarged when decreasing sinograms, which is caused by the increasing "sinogram choosing freedom". In other words, when decreasing $k_{\text{max}}$ from 90 to 15, the whole selection range (0, $e_{\text{max}}$) is fixed and the selection rate is reduced to 1/6. Therefore, we have more candidate angle positions in each selection. This highlights the advantage of the trajectory optimization of SAS and GDS in such extremely sparse scenarios. As the clinical setup is different across hospitals, the model robustness to the involved photo noises is important, and we test these policies with noise levels $L_1$ (Poisson noise level $5e^5$) and $L_2$ (Poisson noise level $1e^6$). Across these noise levels, our SAS and GDS achieve consistently better results, which confirms its feasibility in practice.

To further qualitatively characterize the interaction between the learning-based policies and the final reconstruction, we visualize the reconstructions of RS, US, and SAS in Fig. 4. The first three rows correspond to the difference images, reconstructed images, and zoom-in images when $k_{\text{max}} = 30$, respectively; the last three rows are when $k_{\text{max}} = 15$. Following the above quantitative results, the reconstructed images with SAS show better performances in both intrathoracic tissues and bones. With such extremely sparse views, projection information is limited, and the details in intrathoracic parts are difficult to be recovered with both RS and US as shown in the bottom right part of Fig. 4. While simultaneously optimizing sampling trajectories provides a potential to focus on the projections that are the most important for the final reconstructions. Besides, comparing the four images representing very different tissue distributions, the dynamically chosen sinograms are indeed individually adaptive, caused by the different interactions between the reconstruction and each sampling sinogram. As the

**Table 2: Testing results on the other anatomies, i.e., brain, and abdomen, when $k_{\text{max}} = 30$. Our SAS policy achieves consistently better performances on all the anatomies, implying the robustness of the proposed policy.**

| NIH-AAPM | Brain            | Abdomen       |
|----------|------------------|---------------|
|          | PSNR(db) | SSIM     | RMSE     | PSNR(db) | SSIM     | RMSE     |
| RS       | 29.03±1.19 | .908±.032 | .036±.009 | 29.67±2.29 | .947±.016 | .034±.011 |
| US       | 33.33±2.89 | .942±.025 | .023±.008 | 31.69±1.67 | .952±.011 | .027±.006 |
| SAS(ours)| 33.74±1.89 | .954±.010 | .021±.005 | 33.12±1.70 | .955±.012 | .022±.005 |

Figure 4: We visualize the reconstructions based on RS, US, and learning-based SAS in the 2nd and 5th rows, where $k_{\text{max}} = 30$ and 15 respectively. Accordingly, we also show the corresponding difference images and zoom-in images. With the same radiation dose, obviously, the reconstructions with sinograms recommended by SAS achieve better results, where the reconstruction details are clearer than the others. Besides, we visualize the learned sampling trajectories of SAS across different cases, which indeed contributes to the final reconstruction. The display window is set to [-1000, 1000] in all cases.
quantitative and qualitative performances are competitive, we only experiment with the SAS policy later for simplicity.

4.1.2 Hyperparameters. We then discuss the choice of contained hyperparameters in terms of chest CT image reconstruction when $k_{\text{max}} = 30$. The conducted ablations of four hyperparameters $\{k_0, k', \alpha'_p, \alpha'_q\}$ used in testing is exhibited in Fig. 6. In subfigure (a), when we fix the other hyperparameters and gradually increase $k_0$ from 1 to 10, the reconstruction performance decreases step by step. Note that when increasing $k_0$, the searching freedom is gradually tightened which limits the final reconstructions. With a similar procedure, we have chosen the best $k', \alpha'_p$ and $\alpha'_q$ for reconstruction.

4.1.3 Robustness on anatomical structures. Since CT-based clinical decision-making concerns multiple anatomical structures, we then conduct experiments with brain and abdomen CT images to verify the robustness. In experiments, we select 10 patients with abdomen CT images and 12 patients with brain CT images for training and use additional 4 patients with abdomen CT images and 3 patients with brain CT images for testing. The $k_{\text{max}}$ is set to 30 and the results are shown in Table 2. We observe a consistent improvement of our SAS policy over RS and US. Especially, for the abdomen images, our policy obtains the 1.43 dB improvement over the US. Accordingly, we visualize the images in Fig. 5, where SAS reconstructions provide better clinical RoI recovery in zoom-in images. The accompanying learned trajectories are shown in the last column.

4.2 Radiation dose reduction analysis.

Since the radiation dose raises safety concerns for patients, reducing it while keeping the reconstructed image quality is indispensable in CT imaging. To exhibit how our reconstruction framework achieves such demand, we visualize the sampling and reconstruction process when $k_{\text{max}} = 15, 30$ in Fig. 7. As shown with the red and blue lines, the performance is improved faster in the beginning, while suffering instability across samples. Then, with sufficient sampled views, the improvement slows down when sampling the last 13.3% views, while model deviation on samples is smaller. Moreover, we mark the performance of US as the baseline with green dotted lines. With such comparison, we observe that our SAS achieves comparable performances with US when $n = 13, 26$ in the two scenarios, respectively. This means that the sampling trajectory optimization of SAS is capable to reduce about 13.3% views, i.e., **13.3% radiation dose**. When sampling these views with additional dose, we can further improve the reconstruction performances. The characteristic is especially important for protecting patient safety, and ensures that dynamically changing sampling trajectory is clinically necessary.

Accordingly, we show three intermediate images when $n = 10, 20, 26$, and the corresponding learned trajectory. The dynamic process shows that the sampling sparsity is still important as in the US, which helps cover body information as much as possible. Then,"
As discussed in Sec 3.1, the reconstruction framework with ASV sampling renders a potential to combine with downstream tasks to introduce the prior, i.e., clinically-concerned RoI. Targeting the exploration of the interaction between the loss function introduced in Eq. (4) and our SAS policy, we investigate RoI-aware reconstruction and conduct experiments on VerSe [21] covering the spine. Specifically, we employ partial slices of seven patients covering spines, of which five are for training and the others for testing. To verify the influence, we equip the weighted loss for both US and SAS policies, named US_RoI and SAS_RoI. The related mask $M$ is very coarse, instead of the well-defined spine mask since such information is expensive to obtain in practice, and we report the quantitative results in Table 3. Comparing US and SAS, our policy achieves better reconstruction of both the whole image and RoI. Further, when additionally employing Eq. (4), SAS_RoI increases about 1dB on the region containing spines, even if the SAS has achieved 27.83 dB. The improvements on the RoI ensure that the active sampling process is guided by the weighted reconstruction loss. Besides, we visualize the RoI image in Fig. 8, including three cases with different spine parts. Coinciding with the quantitative results, the US is almost not affected by the weighted RoI loss. In contrast, SAS_RoI provides clearer recovery on the clinical concerned region, which is valuable for practical diagnosis.

5 CONCLUSIONS

In this work, we break through the limitation of USV sampling in CT by proposing a novel framework of ASV sampling that is adaptive for CT reconstruction with a reduced radiation dose. It is done by the specifically-designed SA module for selecting the sinograms. In this way, the whole sampling trajectory takes into account the patient differences, and the later sampled sinogram position is determined by the pre-sampled ones. With such a design, we propose two ASV sampling policies, called SAS and GDS, both achieving better reconstruction performances with the same reconstructor. Furthermore, when fusing the RoI information of clinically-concerned anatomy, the ASV sampling policy achieves even better reconstruction, especially in the RoI.

Table 3: Testing performance on the VerSe datasets ($k_{\text{max}} = 30$). The additional weighted loss indeed improves SAS policy performance with a larger margin over the US policy, verifying the interaction between the loss and view selection.

| VerSe-Spine | Whole Image | Region of Interest (RoI) |
|-------------|-------------|--------------------------|
| | PSNR (dB) | SSIM | RMSE | PSNR (dB) | SSIM | RMSE |
| US | 28.75 ± 1.15 | .911 ± .009 | .038 ± .010 | 26.13 ± 1.96 | .893 ± .019 | .051 ± .015 |
| US_RoI | 28.82 ± 1.29 | .904 ± .015 | .037 ± .006 | 26.47 ± 1.74 | .894 ± .022 | .047 ± .010 |
| SAS (ours) | 29.24 ± 1.01 | .919 ± .009 | .035 ± .004 | 27.83 ± 1.93 | .917 ± .017 | .042 ± .009 |
| SAS_RoI (ours) | 29.91 ± 1.00 | .913 ± .009 | .032 ± .004 | 28.86 ± 1.41 | .916 ± .016 | .037 ± .006 |

ACKNOWLEDGMENTS

This work was supported by the National Key Research and Development Program of China under Grant No. 2020AAA0105802, 2020AAA0105800, in part by the National Natural Science Foundation of China (NSFC) under Grant No. 62271465, 12001180, 12101061, in part by the Natural Science Foundation of Jiangsu Province under Grant SBK2023044571, in part by the Open Fund Project of Guangdong Academy of Medical Sciences, China (No. YKY-KF202206).
active CT reconstruction with a learned sampling policy

MM ’23, October 29–November 3, 2023, Ottawa, ON, Canada.

REFERENCES

[1] Jonas Adler and Ozan Öktem. 2018. Learned primal-dual reconstruction. IEEE transactions on medical imaging 37, 6 (2018), 1332–1332.

[2] Adler, Jonas and Khojol Holger and Öktem, Ozan. 2017. Operator discretization library (ODL). Software available from https://github.com/odlgroup/odl.

[3] Martin Arjovsky, Soumith Chintala, and Leon Bottou. 2017. Wasserstein generative adversarial networks. In International conference on machine learning. PMLR, 214–223.

[4] Hu Chen, Yi Zhang, Yunjin Chen, Junfeng Zhang, Weihuang Zhang, Huaiqiang Sun, Yang Lv, Peixi Liao, Jiliu Zhou, and Ge Wang. 2018. LEARN: Learned experts’ assessment-based reconstruction network for sparse-data CT. IEEE transactions on medical imaging 37, 6 (2018), 1333–1347.

[5] Chen, Hu, Zhang, Yang, Lv, Liao, Peixi, Zhou, Jiliu, and Wang, Ge. 2017. Low-dose CT with a residual encoder-decoder convolutional neural network. IEEE transactions on medical imaging 36, 12 (2017), 2524–2535.

[6] Weilin Cheng, Wu, Yang, Hongwei, Li, and Yuping, Duan. 2020. Learned Full-Sampling Reconstruction From Incomplete Data. IEEE transactions on Computational Imaging 6 (2020), 945–957.

[7] Ravishankar N. Chityala, Kenneth R. Hoffmann, Daniel R. Bednarek, and Stephen Rudin. 2004. Region of interest (ROI) computed tomography. In Medical Imaging. 5366. SPIE. 534–541.

[8] Jae Kyu Choi, Bin Dong, and Xiaoqin Zhang. 2016. Limited tomography reconstruction via tight frame and simultaneous sinogram extrapolation. Journal of Computational Mathematics (2016), 575–589.

[9] Bin Dong, Jia Li, and Zuowei Shen. 2013. X-ray CT image reconstruction via wavelet frame based regularization and Radon domain inpainting. Journal of Scientific Computing 54 (2013), 333–349.

[10] Hao Gao, Jiao Feng, Cai, Zuowei, Shen, and Hongkai. Zhao. 2011. Robust principal component analysis-based four-dimensional computed tomography. Physics in Medicine & Biology 56, 11 (2011), 3181.

[11] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Hao Gao, Jian-Feng Cai, Zuowei Shen, and Hongkai Zhao. 2011. Robust principal component analysis-based four-dimensional computed tomography. Physics in Medicine & Biology 56, 11 (2011), 3181.

[12] Dominic J Heuscher and Frederic Noo. 2011. CT dose reduction using dynamic collimation. In Proceedings of the 2011 IEEE Nuclear Science Symposium Conference Record. IEEE, 3470–3473.

[13] Jonas Adler and Ozan Öktem. 2018. Learned primal-dual reconstruction. IEEE transactions on medical imaging 37, 6 Part35 (2016), 3759–3760.

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

[15] Emrehan Kang, Junhong Min, and Jung Chul Ye. 2017. A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction. Medical physics 44, 10 (2017), e960–e975.

[16] Kyyungsoo Kim, Jung Chul Ye, William Worstell, Jinsong Ouyang, Yotkin Rakvongthai, Georges El Fakhri, and Quanzheng Li. 2014. Sparse-view spectral CT reconstruction using spectral patch-based low-rank penalty. IEEE transactions on medical imaging 34, 3 (2014), 748–760.

[17] Diederick P. Kuimga and Jimmy Ra. 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014).

[18] Juihong Liu, Angelica I. Aviles-Rivero, Hui Ji, and Carola-Bibiane Schönlieb. 2019. Bilevel medical image reconstruction via shape prior, going deeper and faster: Deep joint indirect registration and reconstruction. Medical image Analysis 68 (2021), 101938.

[19] Faisal Mahmood, Nauman Shahid, Ulfi Skoglund, and Pierre Vanderbeyght. 2018. Adaptive graph-based total variation for tomographic reconstructions. IEEE Signal Processing Letters 25, 5 (2018), 700–704.

[20] C. McCollough. 2016. TU-FG-20A-04: Overview of the Low Dose CT Grand Challenge. Medical physics 43, 6Part5 (2016), 3739–3760.

[21] Massimiliano Miracoli, Simopekka Vanska, Seppo Jarvenpaa, Martti Kalke, Matti Lassac, Jan Moberg, and Samuli Siltanen. 2006. Wavelet-based reconstruction for limited-angle X-ray tomography. IEEE transactions on medical imaging 25, 2 (2006), 210–217.

[22] Arjuny Sekhooyina, Malek E. Hussein, Amirhossein Bayat, Maximilian Löfler, Hans Liebl, Hongwei Li, Giles Tetteh, Jan Kukačka, Christian Payer, Darko Stern, et al. 2021. VerSe: A Vertebrae labelling and segmentation benchmark for multi-detector CT images. Medical image analysis 73 (2021), 102166.

[23] Oguz Semerci, Ning Hao, Misha E. Kilmer, and Eric L. Miller. 2014. Tensor-based formulation and nuclear norm regularization for multifidelity computed tomography. IEEE transactions on Image Processing 23, 4 (2014), 1678–1693.

[24] Ziju Shen, Yufei Wang, Dafan Wu, Xu Yang, and Bin Dong. 2020. Learning to scan: a deep reinforcement learning approach for个性化scanning in ct imaging. arXiv preprint arXiv:2006.02420 (2020).

[25] Emil Y Sidky and Xiaochuan Pan. 2008. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. Physics in Medicine & Biology 53, 17 (2008), 4777.

[26] Jian Sun, Huibin Li, Zongheng Xu, et al. 2016. Deep ADMN-Net for compressive sensing MRI. Advances in neural information processing systems 29 (2016).

[27] Ce Wang, Zhaolun Chen, Kun Shang, and Huaming Wu. 2019. Label-removed generative adversarial networks incorporating with K-Means. Neurocomputing 361 (2019), 126–136.

[28] Ce Wang, Haimiao Zhang, Qian Li, Kun Shang, Yuanxuan Lyu, Bin Dong, and S Kevin Zhuo. 2021. Improving generalizability in limited-angle ct reconstruction with sinogram extrapolation. In International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 86–96.

[29] Ge Wang, J. C. Ye, and Bruno De Man. 2020. Deep learning for tomographic image reconstruction. Nature Machine Intelligence 2 (2020), 737–748.

[30] Wang, Zhong, McCann, Ozan, and Kilmer, Misha. 2014. Tensorized Learned Experts’ Assessment-Based Reconstruction Network for Sparse-Data CT. IEEE transactions on medical imaging 37, 6 (2018), 1333–1347.

[31] Zhou Wang, Chandan Misra, and Stephen J. Nowicki. 2016. Ego learned neural network for x-ray ct reconstruction. In Proceedings of the 2016 International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 1466–1473.

[32] Weiwen Wu, Dianlin Hu, Chuang Ni, Hengyong Yu, Varut Vandhanabbuti, and Ge Wang. 2021. DRONE: dual-domain residual-based optimization network for sparse-view CT reconstruction. IEEE Transactions on Medical Imaging 40, 11 (2021), 3092–3104.

[33] Bo Zhou, S Kevin Zhou, James S Duncan, and Chi Liu. 2021. Improving generalizability in limited-angle ct reconstruction with sinogram extrapolation. In International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 86–96.

[34] Emil Y Sidky and Xiaochuan Pan. 2008. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. Physics in Medicine & Biology 53, 17 (2008), 4777.

[35] Ge Wang, J. C. Ye, and Bruno De Man. 2020. Deep learning for tomographic image reconstruction. Nature Machine Intelligence 2 (2020), 737–748.
Algorithm 1: Training Strategy

**Input:** The ground truth CT image $u$, the involved $a_{\text{max}}$, $k_{\text{max}}$, $k_0$, $k_p$, $k_q$, the number of epochs $T$, batch size $b = 1$, learning rate $lr$, the network parameters $\theta$ of $R$, $\psi$ of $SA$, and $\beta_1, \beta_2$ of Adam.

**Output:** The optimized parameters $\hat{\theta}$, $\hat{\psi}$.

1. for $t = 1 : T$ do
2. Sample initial $k_0$ sinograms, denoted as $y_0$;
3. for $n = 0 : N$ ($N = (k_{\text{max}} - k_0)/k$) do
4. Reconstruct image $R_{\theta}(y_n)$;
5. Evaluate $R_{\hat{\theta}}(y_n)$ with $SA_{\hat{\psi}}$;
6. Recommend $k$ sampling positions in $(a_p, a_q)$;
7. Sample the $k$ sinograms, denoted as the set $K_n$;
8. Update $y_{n+1} = y_n \cup K_n$;
9. if $t \% 2 = 0$ then
10. Update reconstructor parameters $\theta$;
11. $\theta = \text{Adam}(\text{grad}_{\theta}, \theta, lr, \beta_1, \beta_2)$
12. else
13. Update evaluator parameters $\psi$;
14. $\psi = \text{Adam}(\text{grad}_{\psi}, \psi, lr, \beta_1, \beta_2)$

Algorithm 2: Active sampling and reconstruction

**Input:** The optimized parameters $\hat{\theta}$, $\hat{\psi}$, and the involved $a_{\text{max}}$, $k_{\text{max}}$, $k_0$, $k_p$, $a_p$, $a_q$.

**Output:** The reconstruction $\hat{u}$.

1. Sample initial $k_0$ sinograms, denoted as $y'_0$;
2. for $n = 0 : N$ ($N = (k_{\text{max}} - k_0)/k'$) do
3. Reconstruct image $R_{\hat{\theta}}(y'_n)$;
4. Evaluate $R_{\hat{\theta}}(y'_n)$ with $SA_{\hat{\psi}}$;
5. Recommend $k'$ sampling positions in $(a_p', a_q')$;
6. Sample the $k'$ sinograms, denoted as the set $K'_n$;
7. Update $y'_{n+1} = y'_n \cup K'_n$;
8. Reconstruct the final CT image $\hat{u}$ with $y'_{N-1}$.

A TRAINING AND TESTING ALGORITHMS

To make the work concrete and easy to reproduce, we exhibit the training strategy and testing process of our proposed framework in Algorithm 1 and Algorithm 2. As described in the main manuscript, we make the hyperparameters different in training and testing. In this way, we can release more freedom in training for the sampling position searching, and therefore, provide robustness for the model. Then, with tightening it in testing stage, we can further empirically and manually reduce the bad cases.