DOLCE: A Model-Based Probabilistic Diffusion Framework for Limited-Angle CT Reconstruction

Jiaming Liu[1] Rushil Anirudh[2] Jayaraman J. Thiagarajan[2] Stewart He[2]
K. Aditya Mohan[2] Ulugbek S. Kamilov[1*] Hyojin Kim[2*]
[1]Washington University in St. Louis [2]Lawrence Livermore National Laboratory

A. CT Reconstruction Formulation

In our experiments, the object to be imaged is placed in between a source of parallel beam x-rays and a planar detector array. The x-rays get attenuated as they propagate through the object and the intensity of attenuated x-rays exiting the object is measured by the detector. To perform tomographic imaging, the object is rotated along an axis and repeatedly imaged at regular angular intervals of rotation. Assume that the object is stationary in the Cartesian coordinate system represented by the axes (x, y, z), at each rotation angle θ of the object, we are interested in reconstructing 2D slice images, denoted as $\rho(x, y, z)$ of object linear attenuation coefficient (LAC) values along the propagation path. The projection at a distance of r on the detector is given by

$$S_\theta(r, z) = \int\int \rho(x, y, z)\delta(x\cos(\theta) + y\sin(\theta) - r)\, dxdy, \tag{1}$$

where δ is the indicator function and $S_\theta(r, z)$ is known as sinogram. Note that equation (1) is separable in the z coordinate. Hence, the projection relation is essentially a 2D function in the $x - y$ plane that is repeatedly applied along the z-axis. The reconstruction of $\rho(x, y, z)$ from incomplete sinogram can be formulated as image inverse problem described in the main paper (See [5, 7] for more references).

FBP Reconstruction Filtered back-projection (FBP) is an analytic algorithm for reconstructing the sample $\rho(x, y, z)$ for the projections $S_\theta(r, z)$ at all the rotation angles θ. In FBP, we first compute the filtered projection measurement of each slice $\widehat{S}_\theta(r) = \int \mathcal{F}[S_\theta](\omega)|\omega|e^{j2\pi\omega r}d\omega$, where $\mathcal{F}$ denotes the Fourier transform and $|\omega|$ is the frequency response of the filter. The filtered back projection reconstruction is then given by [5]

$$f_{\text{FBP}}(x, y) = \int_0^\pi \widehat{S}_\theta(x\cos(\theta) + y\sin(\theta))d\theta. \tag{2}$$

According to equation (2), we know that a filtered version of S_θ is smeared back on the $x - y$ plane along the direction $(\pi/2 - \theta)$. The FBP reconstruction thus consists of the cumulative sum of the smeared contributions from all the projections ranging in $0 < \theta < \pi$.

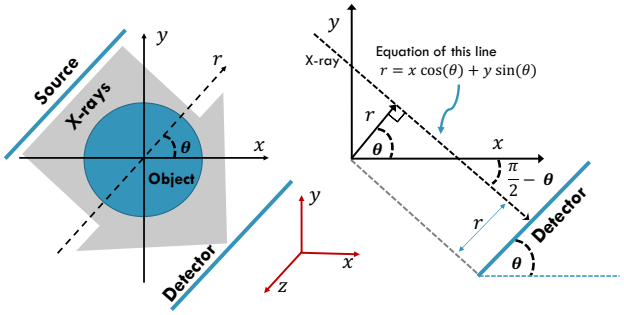

Figure 1. Implementation of x-ray CT. An object is rotated along an axis and exposed to a parallel beam of x-rays. The intensity of attenuated x-rays exiting the object is measured by the detector at regular angular intervals. The projection at an angle of θ measured at a distance of r on the detector is the line integral of LAC values along the line perpendicular to the detector.

LACT Reconstruction Artifacts by FBP. If the projections are acquired over a limited angular range, then the integration in (2) will be incomplete in the angular space. Since each projection $S_\theta(r)$ contains the cumulative sum of the LAC values at a rotation angle of θ, it also contains information about the edges that are oriented along the angular direction $(\pi/2 - \theta)$ as shown in Fig. 1. Now, suppose data acquisition starts at $\theta = 0$ and stops at an angle of $\theta = \theta_{\max} < \pi$. Then, the edge information contained in the projections at the angles $\theta \in [\theta_{\max}, \pi]$ will be missing in the final reconstruction.

*Corresponding authors.

Metric	PSNR ↑			SSIM ↑		
Angle	**60°**	**90°**	**120°**	**60°**	**90°**	**120°**
FBP	14.95	17.28	20.97	0.464	0.543	0.603
RLS	22.72	26.19	30.42	0.699	0.833	0.888
DOLCE (FBP, w/o prox)	33.72	38.00	40.63	0.927	0.945	0.954
DOLCE (FBP, w/ prox)	34.05	38.73	42.00	0.938	0.960	0.972
DOLCE (RLS, w/o prox)	34.91	38.65	41.25	0.936	0.951	0.959
DOLCE (RLS, w/ prox)	35.15	39.27	42.52	0.945	0.963	0.974
DOLCE-SA (FBP, w/o prox)	34.31	38.75	41.49	0.936	0.952	0.961
DOLCE-SA (FBP, w/ prox)	34.58	39.28	42.55	0.943	0.963	0.975
DOLCE-SA (RLS, w/o prox)	35.55	39.31	42.12	0.944	0.954	0.965
DOLCE-SA (RLS, w/ prox)	**35.78**	**39.75**	**43.11**	**0.949**	**0.969**	**0.977**

Table 1. Average PSNR and SSIM results comparing test slices with the ground truth from medical body CT dataset.

Metric	PSNR ↑			SSIM ↑		
Angle	**60°**	**90°**	**120°**	**60°**	**90°**	**120°**
FBP	26.08	28.34	32.18	0.668	0.713	0.752
RLS	28.05	31.01	35.61	0.775	0.860	0.914
DOLCE (FBP, w/o prox)	33.13	37.43	42.83	0.928	0.953	0.974
DOLCE (FBP, w/ prox)	33.44	38.17	44.12	0.928	0.959	0.983
DOLCE (RLS, w/o prox)	33.70	38.64	43.76	0.933	0.959	0.978
DOLCE (RLS, w/ prox)	33.98	39.28	45.18	0.935	0.967	0.987
DOLCE-SA (FBP, w/o prox)	33.85	38.26	43.74	0.932	0.959	0.978
DOLCE-SA (FBP, w/ prox)	34.11	38.91	44.87	0.931	0.963	0.985
DOLCE-SA (RLS, w/o prox)	34.41	39.40	44.52	**0.941**	0.966	0.981
DOLCE-SA (RLS, w/ prox)	**34.68**	**39.88**	**45.68**	**0.941**	**0.971**	**0.988**

Table 2. Average PSNR and SSIM results comparing test slices with the ground truth from checked-in luggage dataset.

This is the reason behind the edge blur in the FBP reconstructions shown in this paper.

B. Additional Implementation Details

CTNet [1] is an end-to-end DL method, designed to predict the invisible sinogram data by incorporating a GAN into the neural network architecture. Note that the original CTNet was developed on 128×128 images. To make it work on 512×512 images, given the pre-trained CTNet on 128×128 images, we additionally train a super-resolution diffusion model presented in [8] to super-resolve the low-resolution outputs of CTNet to the same 512×512 dimension as other methods.

DPIR [10] refers to the SOTA PnP methods using deep denoiser as prior for solving various ill-posed image inverse problems. We modify the publicly available implementation to train the deep denoiser [1] on each dataset separately and follow the similar implementation settings [2] at inference. Since our CT images are naturally in smaller intensity range, we train the DPIR denoiser for the AWGN removal within noise level of $\sigma \in [0, 5]$.

ILVR [2] and **DPS** [3] refer to recently developed conditioning methods based on unconditionally trained DDPM for solving versatile ill-posed inverse problems. We modify the publicly available implementation of ILVR [3] and DPS [4] in order to incorporate our LACT forward-model. We use the similar grid search as DOLCE for fine-tuning the hyper-parameters within ILVR and DPS, respectively.

We train all the diffusion models used in this paper, modified based on the publicly available PyTorch implementation [5]. To indicate the high quality of our pre-trained diffusion models used within ILVR and DPS, we present the random samples from our two unconditionally trained denoising diffusion 512×512 models in Fig. 2 for luggage and medical dataset, respectively.

[1] https://github.com/cszn/KAIR

[2] https://github.com/cszn/DPIR

Metric	PSNR ↑			SSIM ↑		
Angle	**45°**	**50°**	**55°**	**45°**	**50°**	**55°**
FBP	14.49	14.65	14.83	0.321	0.397	0.441
RLS	19.01	19.98	20.95	0.559	0.596	0.632
ILVR [2]	23.85	24.94	26.75	0.815	0.839	0.872
DPS [3]	24.50	25.84	26.71	0.833	0.848	0.870
DOLCE	24.97	27.47	31.02	0.838	0.884	0.934
DOLCE-SA	**25.68**	**27.88**	**31.51**	**0.841**	**0.889**	**0.939**

Table 3. Average PSNR and SSIM results for several methods on body CT dataset. **Best values** for each metric are highlighted.

Metric	PSNR ↑			SSIM ↑		
Angle	**45°**	**50°**	**55°**	**45°**	**50°**	**55°**
FBP	24.59	24.77	25.64	0.653	0.658	0.661
RLS	26.62	26.95	27.31	0.723	0.748	0.758
ILVR [2]	28.99	29.15	29.85	0.856	0.867	0.871
DPS [3]	29.42	29.85	30.40	0.862	0.871	0.878
DOLCE	30.46	31.45	32.39	0.884	0.899	0.921
DOLCE-SA	**30.98**	**31.79**	**32.98**	**0.890**	**0.906**	**0.925**

Table 4. Average PSNR and SSIM results on luggage dataset.

C. Additional Numerical Results

Comparison of RLS and FBP as Conditional Input. In Table 1 and Table 2, we present additional numerical evaluations for FBP and RLS as conditional input of our DOLCE models cross various angular ranges (*e.g.*, $\theta_{\max} \in \{60°, 90°, 120°\}$). We use the same randomly selected 300 images from test luggage and medical dataset as in the main paper, respectively. While DOLCE using FBP as conditional input provides substantial improvements, using RLS input further boosts the overall performance.

Incorporation of Data-Consistency. We also report additional numerical validations on the incorporation of forward-model at inference stage in Table 1 and Table 2. We observer that DOLCE using the data-consistency provided by the proximal-mapping produces better quality reconstruction samples, which highlights the potential of enforcing forward-model within sampling step.

Behavior of DOLCE to Model Mismatch. In Table 3 and Table 4, we demonstrate the behavior of our proposed approach for varying number of views during testing. Specifically, we consider the DOLCE for forward-model mismatch scenarios, where the pre-trained DOLCE models in the main paper are tested for $\theta_{\max} \in \{45°, 50°, 55°\}$ limited-angle data. For references, we compare DOLCE and DOLCE-SA (sample average) to FBP, RLS, ILVR, and DPS. We can see that our DOLCE consistently outperforms baseline methods even under model mismatch cases.

Additional Visual Evaluation. In Fig. 3, we compare the visual results of DOLCE on medical CT test images to FBP, U-Net, ILVR, and DPS for $\theta_{\max} = 60°$. Fig. 4-6 present additional visual comparison for several methods on the luggage dataset. In Fig. 7, we compare DOLCE and DOLCE-SA to DPIR and U-Net on each test dataset, respectively. We also provide *video* comparisons of our DOLCE reconstruction results in the supplement material.

Comparison of DOLCE and Pix2Pix [4] GAN method. We in addition report numerical comparison of DOLCE and Pix2Pix [6], a well know image-to-image translation GAN, on the medical CT dataset in Table C below. The training and testing dataset of Pix2Pix are the same as DOLCE and prepared following its official implementation. Clearly, DOLCE outperforms PiX2Pix by a large margin for different number of views.

Method	**60°**	**90°**	**120°**
Pix2Pix [4]	28.03	31.16	37.55
DOLCE (Ours)	35.11	39.04	42.16

More Results for Uncertainty Quantification. Fig. 8-10 shows additional numerical validation that

[3] https://github.com/jychoi118/ilvr_adm
[4] https://github.com/DPS2022/diffusion-posterior-sampling
[5] https://github.com/openai/guided-diffusion
[6] https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix

DOLCE is able to quantify uncertainty by estimating the variances directly. Since a well-calibrated model indicates larger variance in areas of larger absolute error, variance can be used as a proxy for reconstruction error in the absence of ground truth.

D. 3D Segmentation Results

We presents additional segmentation results on the reconstruction 2D slices obtained from our DOLCE in Fig 11- 15. The purpose of these experiments are to evaluate how quality affects object segmentation. In specific, we use a popular region growing segmentation similar to the method used in [6], which is a simplified version of the method in [9], with a randomly chosen starting position and a fixed kernel size. The luggage dataset contains segmentation labels of objects of interest, and the evaluation focuses on how well each segmentation extracts the labeled object. We reconstruct all slices of each bag through the proposed method and combine them into a single bag in 3D. Then we run the region growing in 3D at multiple, hand-tuned parameter settings (intensity threshold ranging from 0.0001 to 0.02), and reported the results from the best performing setting. This is done as some reconstruction results are in poor quality and sensitive to the threshold. We compare the segmentations obtained using our method to the segmentation labels as reference, and those obtained using full-view ground truth, FBP, TV and DPS, respectively. It is evident to observe that our proposed reconstruction segments the objects of interest very similar to the ground truth images, than compared to using baseline methods for reconstruction.

References

[1] R. Anirudh, H. Kim, J. J Thiagarajan, K. A. Mohan, K. Champley, and T. Bremer. Lose the views: Limited angle ct reconstruction via implicit sinogram completion. In *Proc. CVPR*, pages 6343–6352, 2018. 2

[2] J. Choi, S. Kim, Y. Jeong, Y. Gwon, and S. Yoon. Ilvr: Conditioning method for denoising diffusion probabilistic models. In *Proc. ICCV*, pages 14347–14356, 2021. 2, 3, 5

[3] H. Chung, J. Kim, M. T. Mccann, M. L. Klasky, and J. C. Ye. Diffusion posterior sampling for general noisy inverse problems. *arXiv preprint arXiv:2209.14687*, 2022. 2, 3, 5

[4] P. Isola, J. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *Proc. CVPR*, 2017. 3

[5] A. C. Kak and M. Slaney. *Principles of computerized tomographic imaging*. SIAM, 2001. 1

[6] H. Kim, J. J. Thiagarajan, and P.T. Bremer. A randomized ensemble approach to industrial ct segmentation. In *Proc. ICCV*, pages 1707–1715, 2015. 4

[7] J. Radon. On the determination of functions from their integral values along certain manifolds. *IEEE transactions on medical imaging*, 5(4):170–176, 1986. 1

[8] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi. Image super-resolution via iterative refinement. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2022. 2

[9] D. F. Wiley, D. Ghosh, and C. Woodhouse. Automatic segmentation of ct scans of checked baggage. In *Proceedings of the 2nd International Meeting on Image Formation in X-ray CT*, pages 310–313, 2012. 4

[10] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, and R. Timofte. Plug-and-play image restoration with deep denoiser prior. *IEEE Trans. Patt. Anal. and Machine Intell.*, pages 1–1, 2021. 2

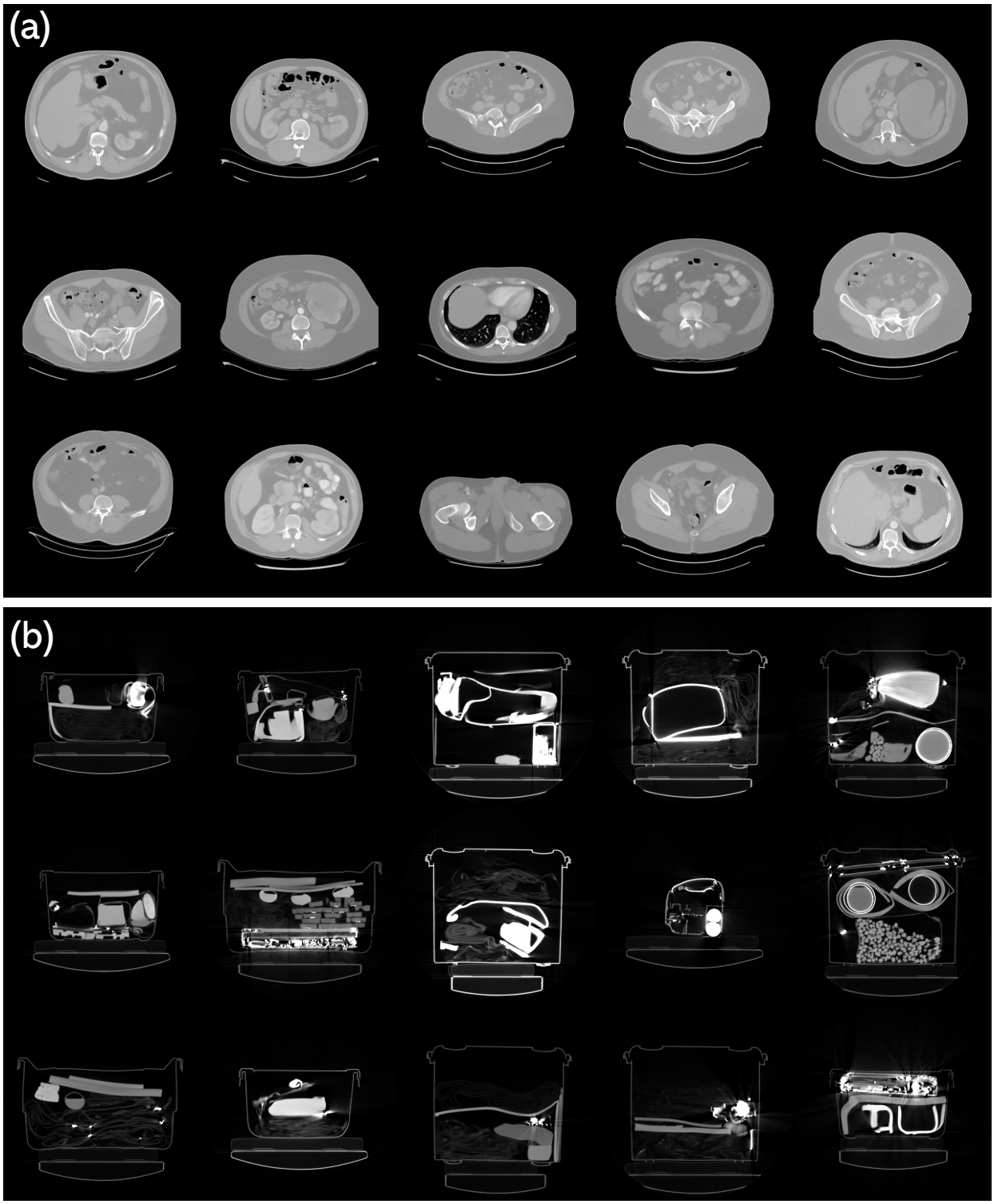

Figure 2. Random samples from our two unconditionally trained denoising diffusion 512×512 models, respectively. **(a):** diffusion model trained on human body CT images; **(b):** diffusion model trained on checked-in luggage dataset. These models are used in ILVR [2] and DPS [3] as baseline methods. Images are normalized for better visualization.

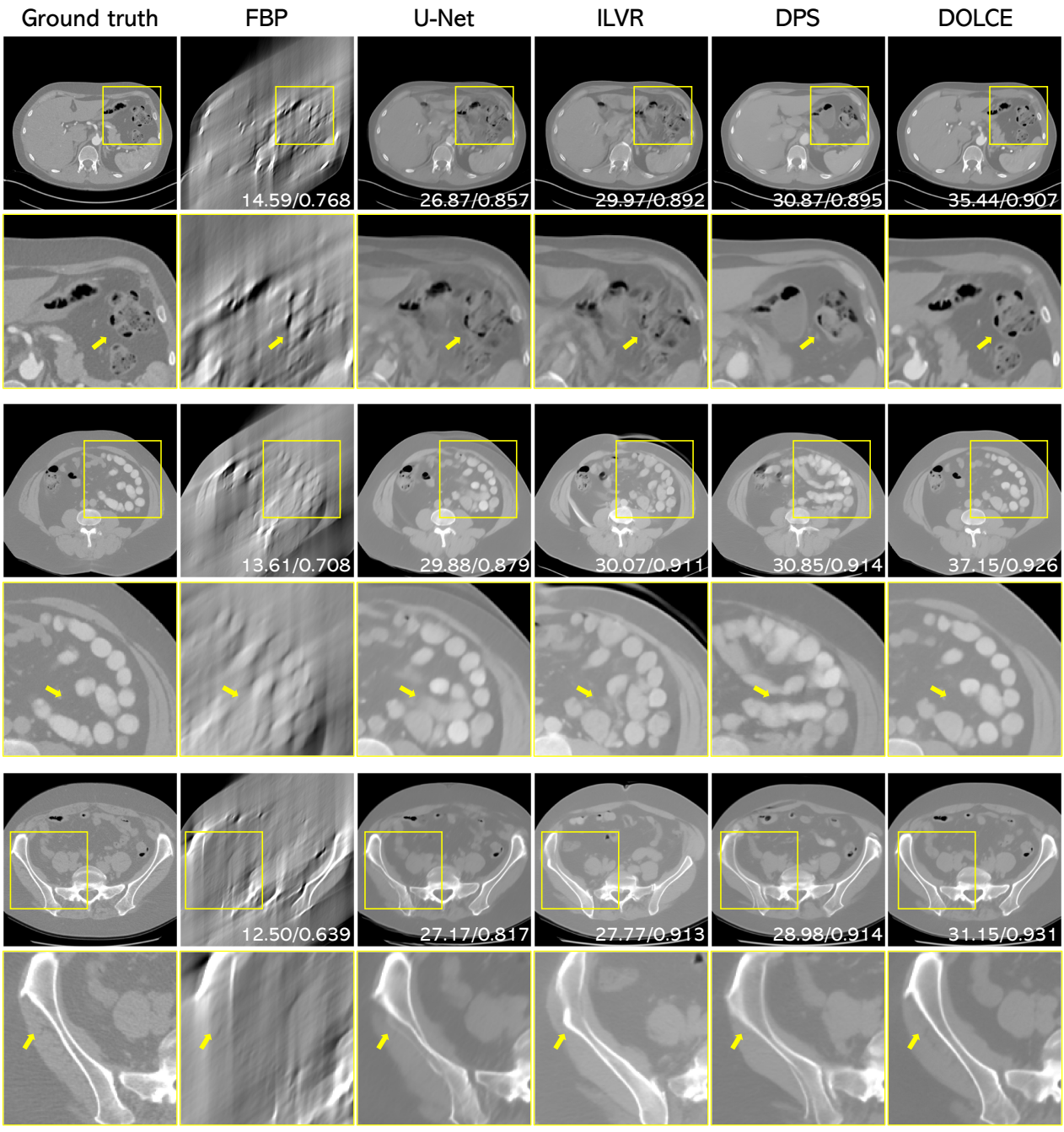

Figure 3. Visual evaluation of limited angle tomographic reconstruction in body CT images, where the input measurements are captured respectively from an angular coverage of 60°. Images are normalized for better visualization.

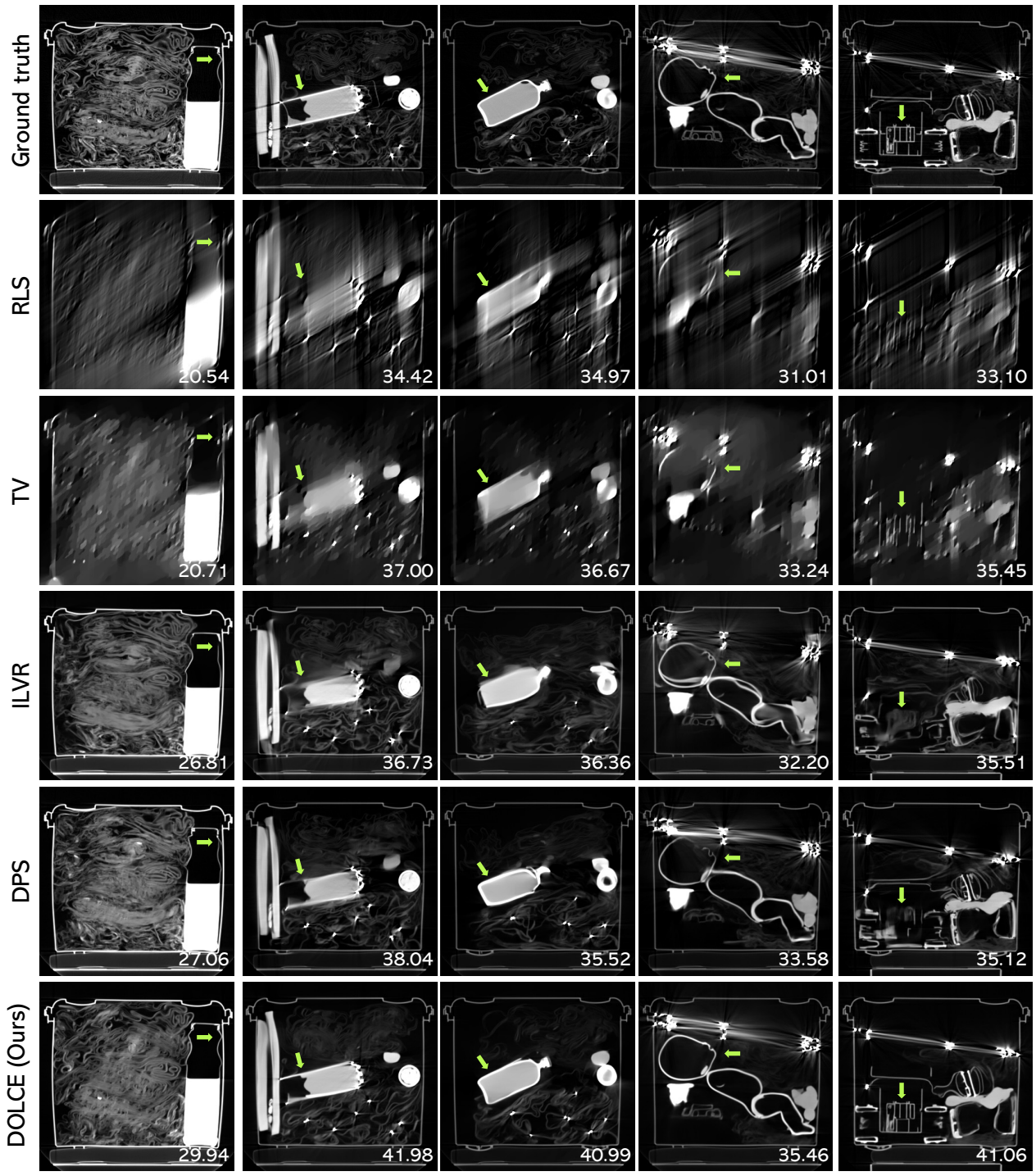

Figure 4. Visual evaluation of limited angle tomographic reconstruction in checked-in luggage, where the input measurements are captured respectively from an angular coverage of 60°. Images are normalized for better visualization.

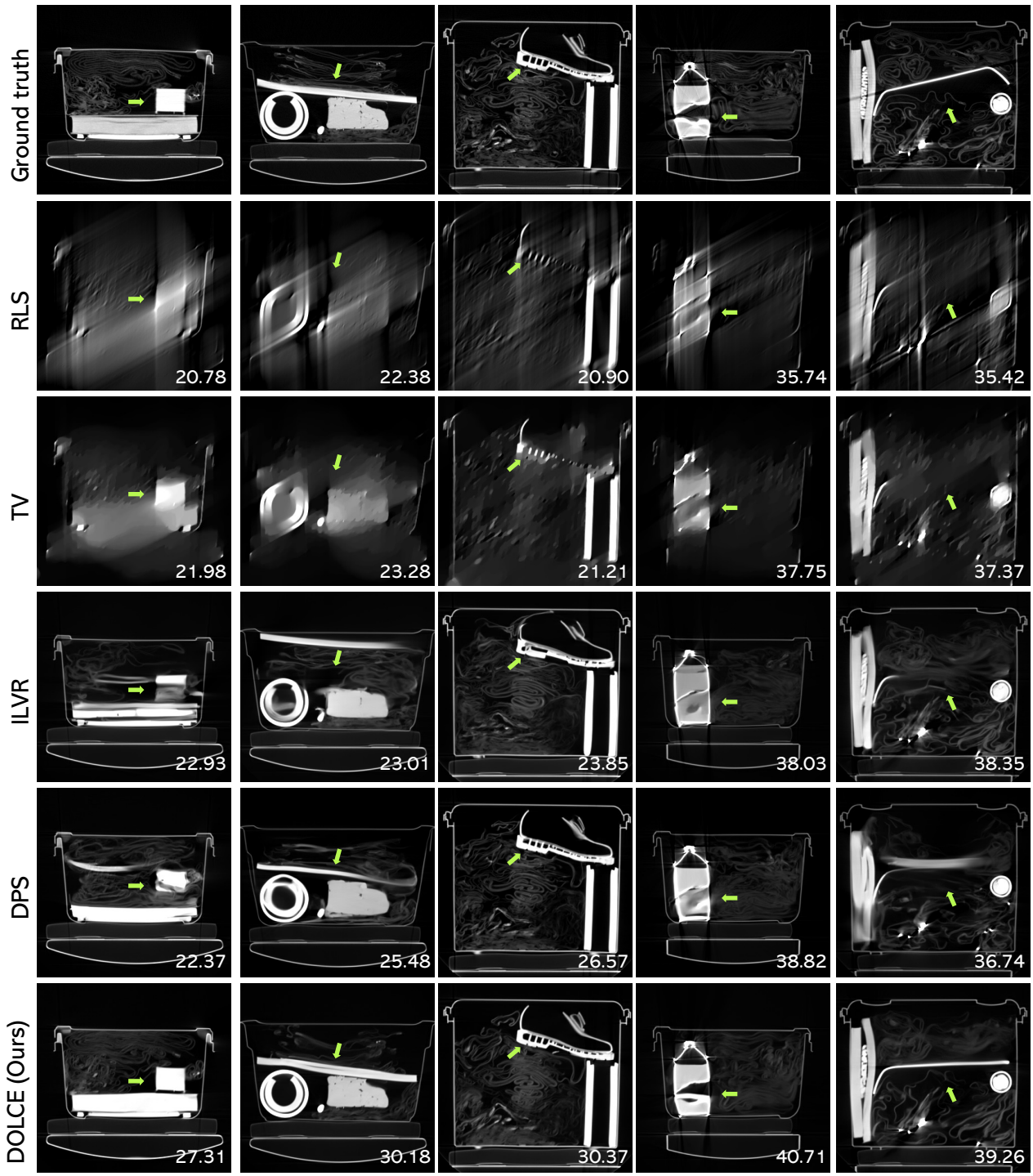

Figure 5. Visual evaluation of limited angle tomographic reconstruction in checked-in luggage, where the input measurements are captured respectively from an angular coverage of 60°. Images are normalized for better visualization.

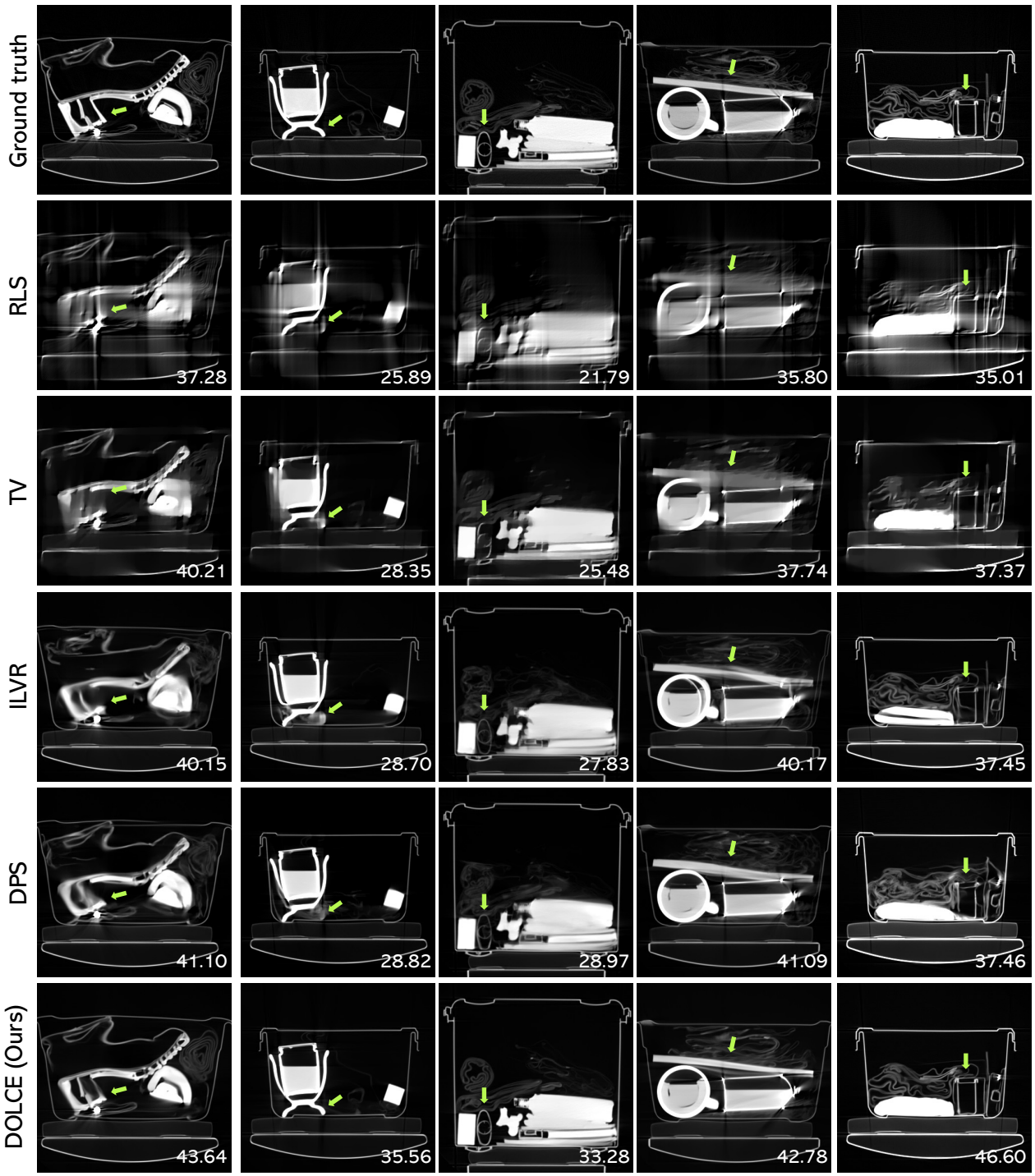

Figure 6. Visual evaluation of limited angle tomographic reconstruction in checked-in luggage, where the input measurements are captured respectively from an angular coverage of 90°. Images are normalized for better visualization.

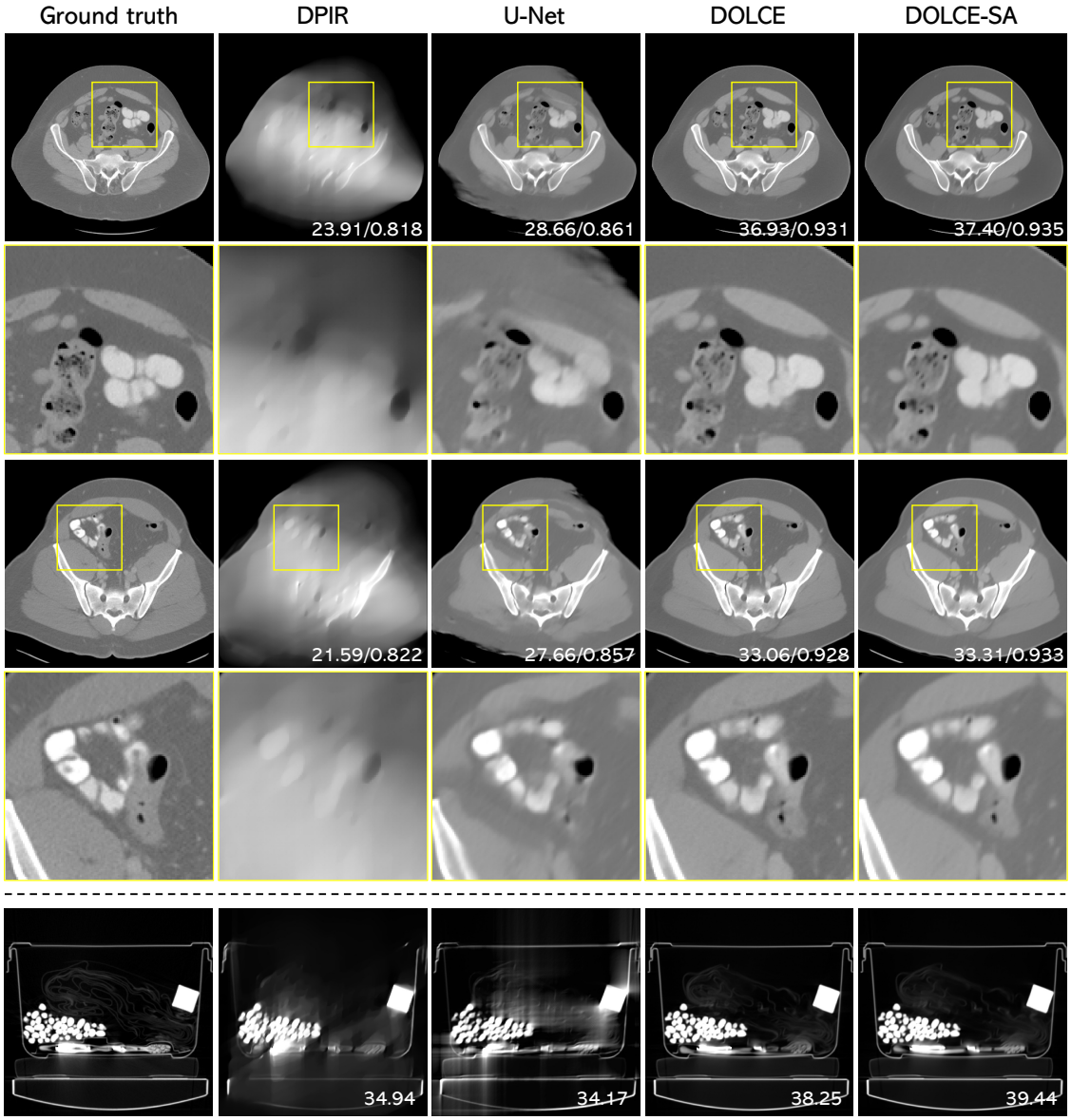

Figure 7. Additional visual evaluation of limited angle tomographic reconstruction in body CT scan (**top**) and checked-in luggage (**bottom**), where the input measurements are captured respectively from an angular coverage of 60°. Images are normalized for better visualization.

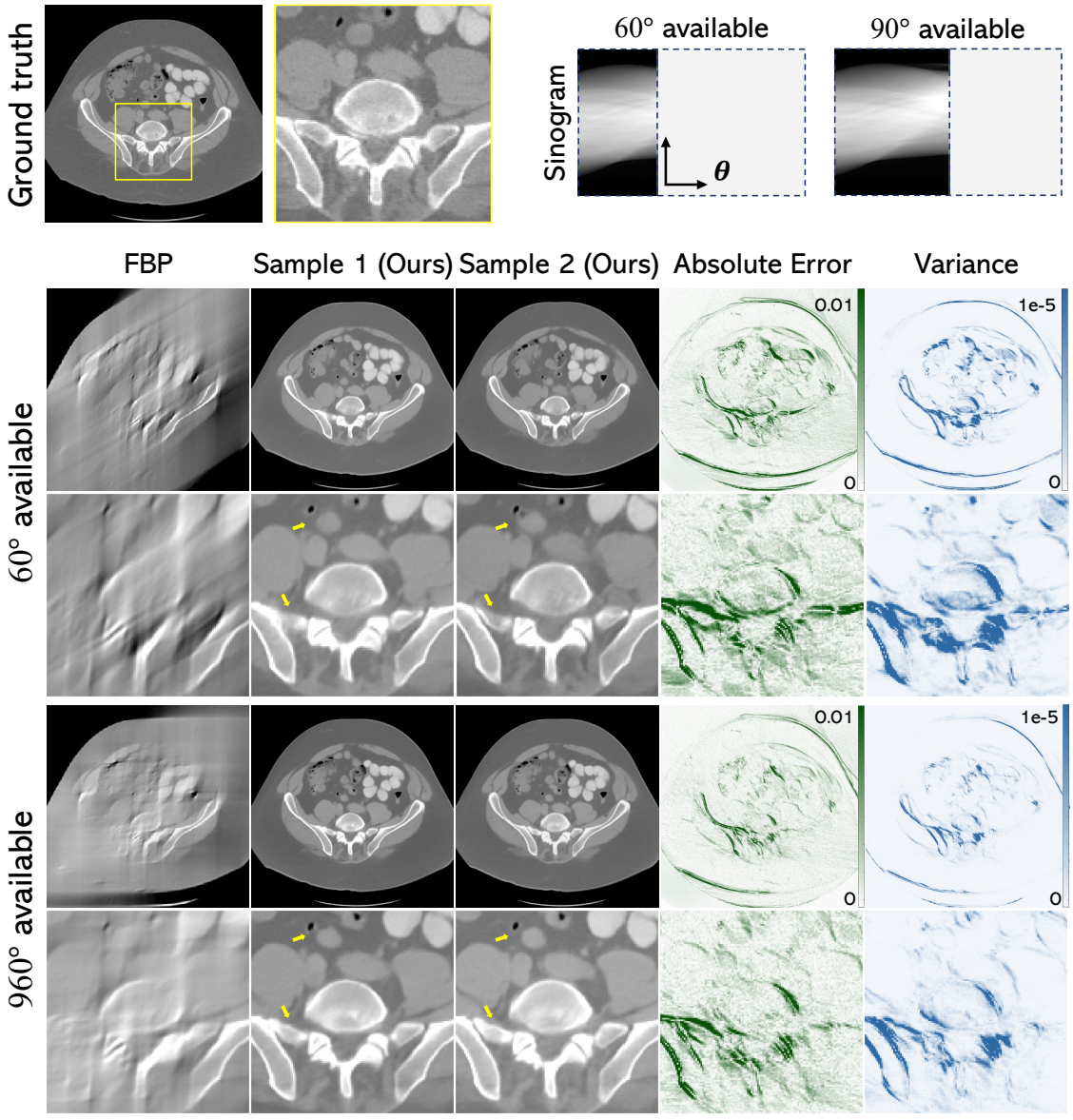

Figure 8. More visual results on body CT images. The error to the ground truth is computed using the conditional mean $\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}]$, and the variance corresponds to per-pixel standard deviation. It is evident that the ill-posed nature of the reconstruction task has a direct impact on the diversity of the generated samples, and the variances are highly correlated with the reconstruction errors.

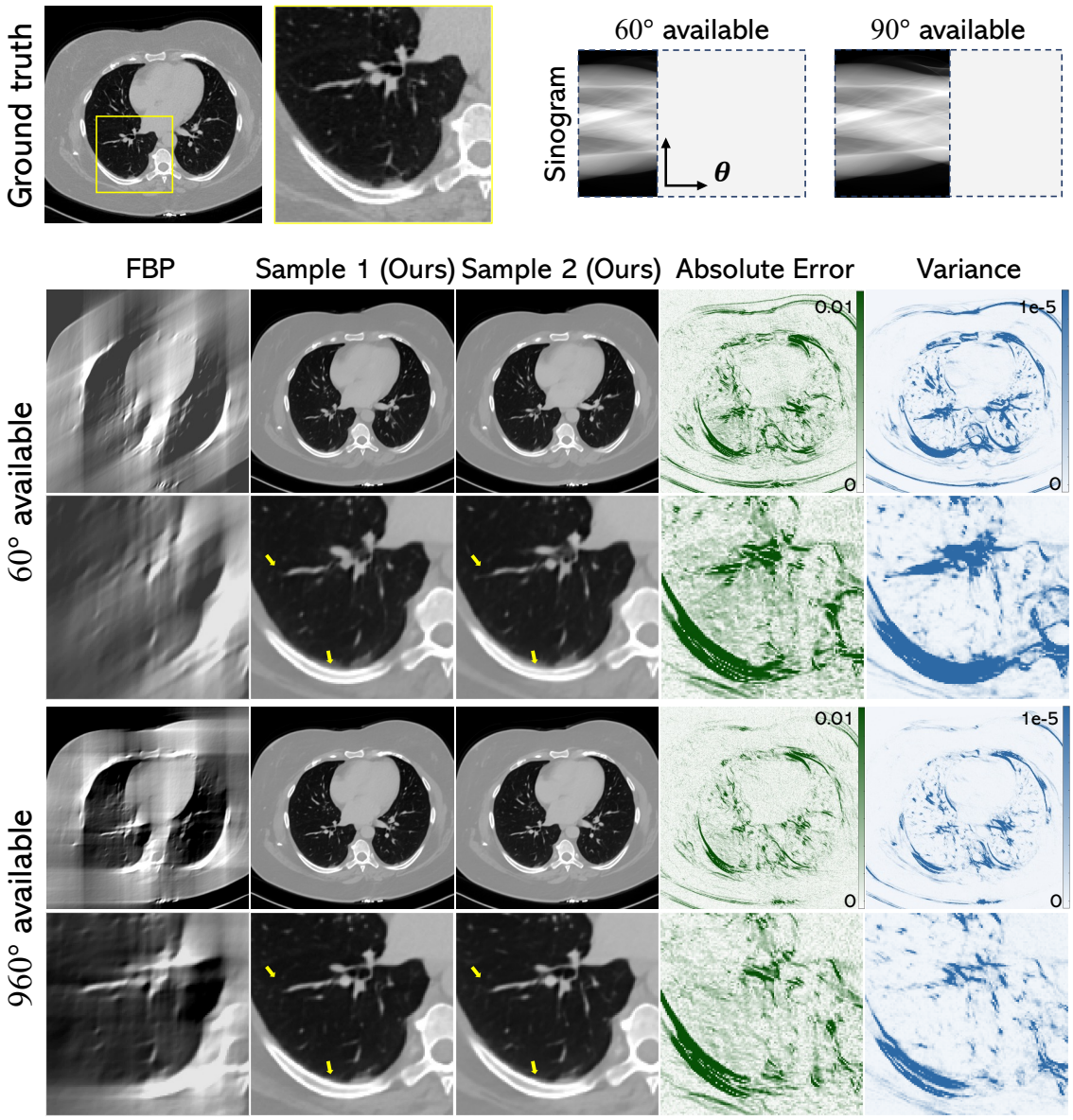

Figure 9. More visual results on body CT images. The error to the ground truth is computed using the conditional mean $\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}]$, and the variance corresponds to per-pixel standard deviation. It is evident that the ill-posed nature of the reconstruction task has a direct impact on the diversity of the generated samples, and the variances are highly correlated with the reconstruction errors.

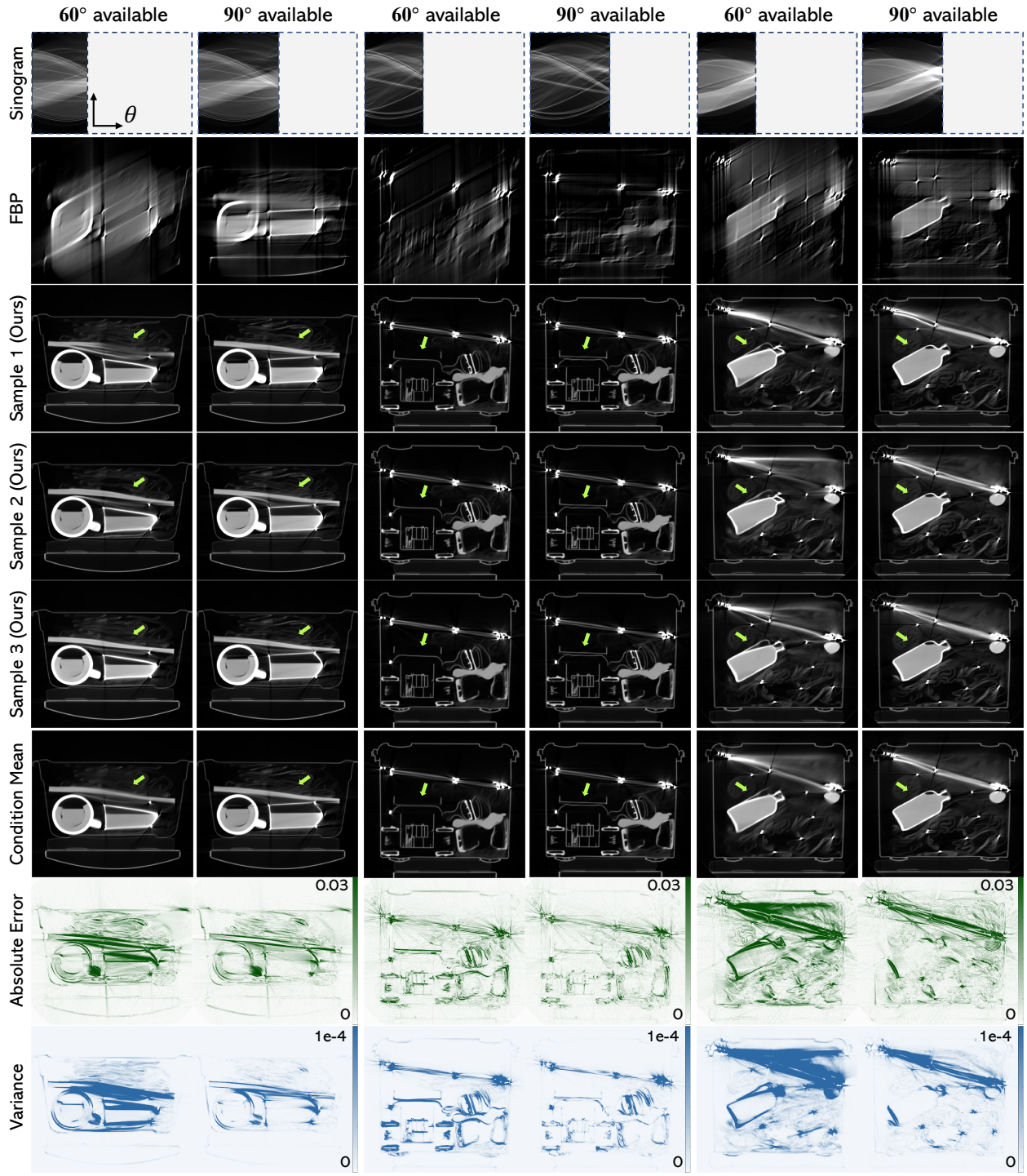

Figure 10. More visual results on luggage images. The error to the ground truth is computed using the conditional mean $\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}]$, and the variance corresponds to per-pixel standard deviation. It is evident that the ill-posed nature of the reconstruction task has a direct impact on the diversity of the generated samples, and the variances are highly correlated with the reconstruction errors.

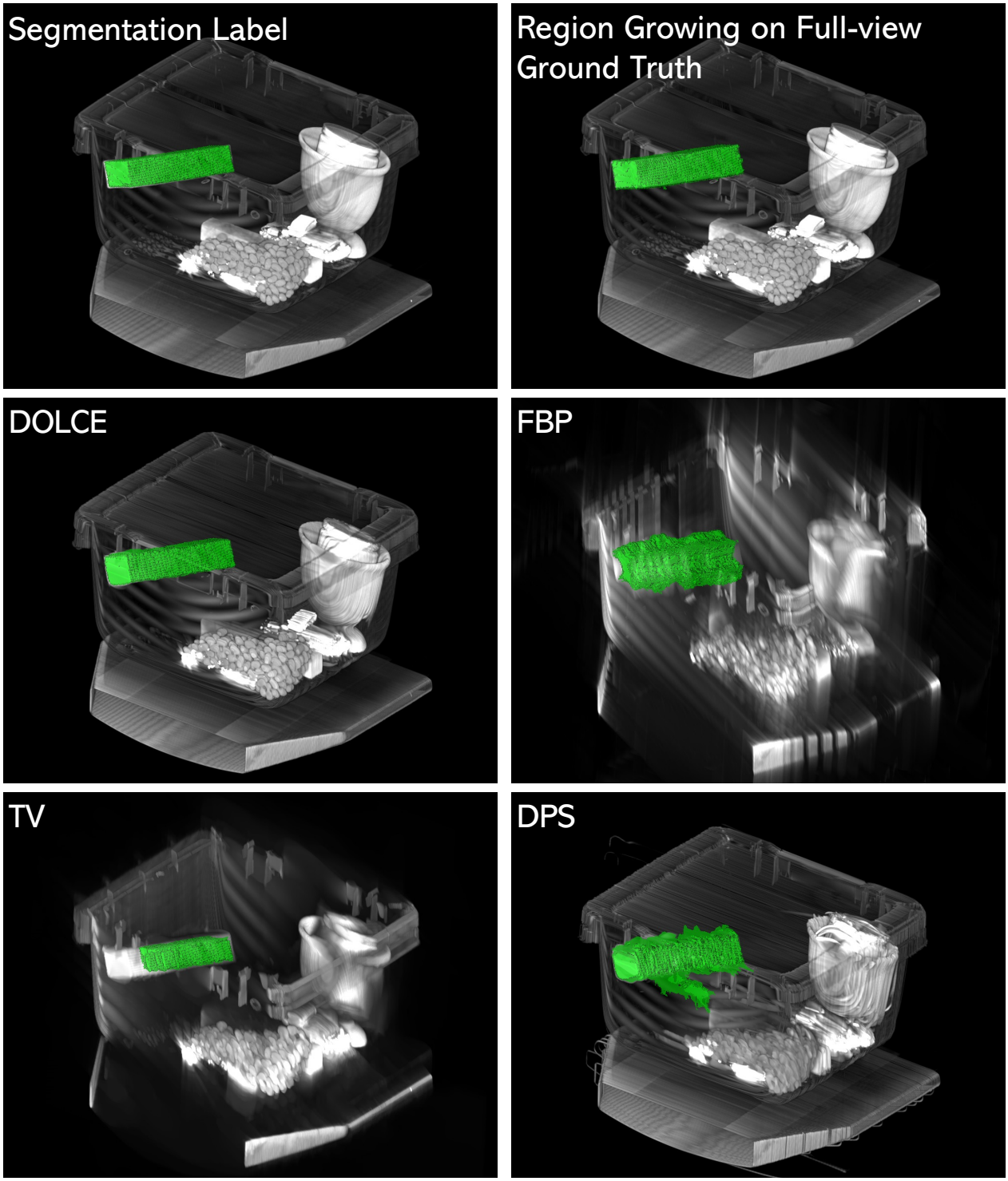

Figure 11. **More 3D Segmentation Results on Test Bag 2:** We use a region growing 3D segmentation in all cases and the resulting segmentations are highlighted in color, against a 3D rendering of the 274 reconstructed 2D slices using $\theta_{\max} = 60^{\circ}$.

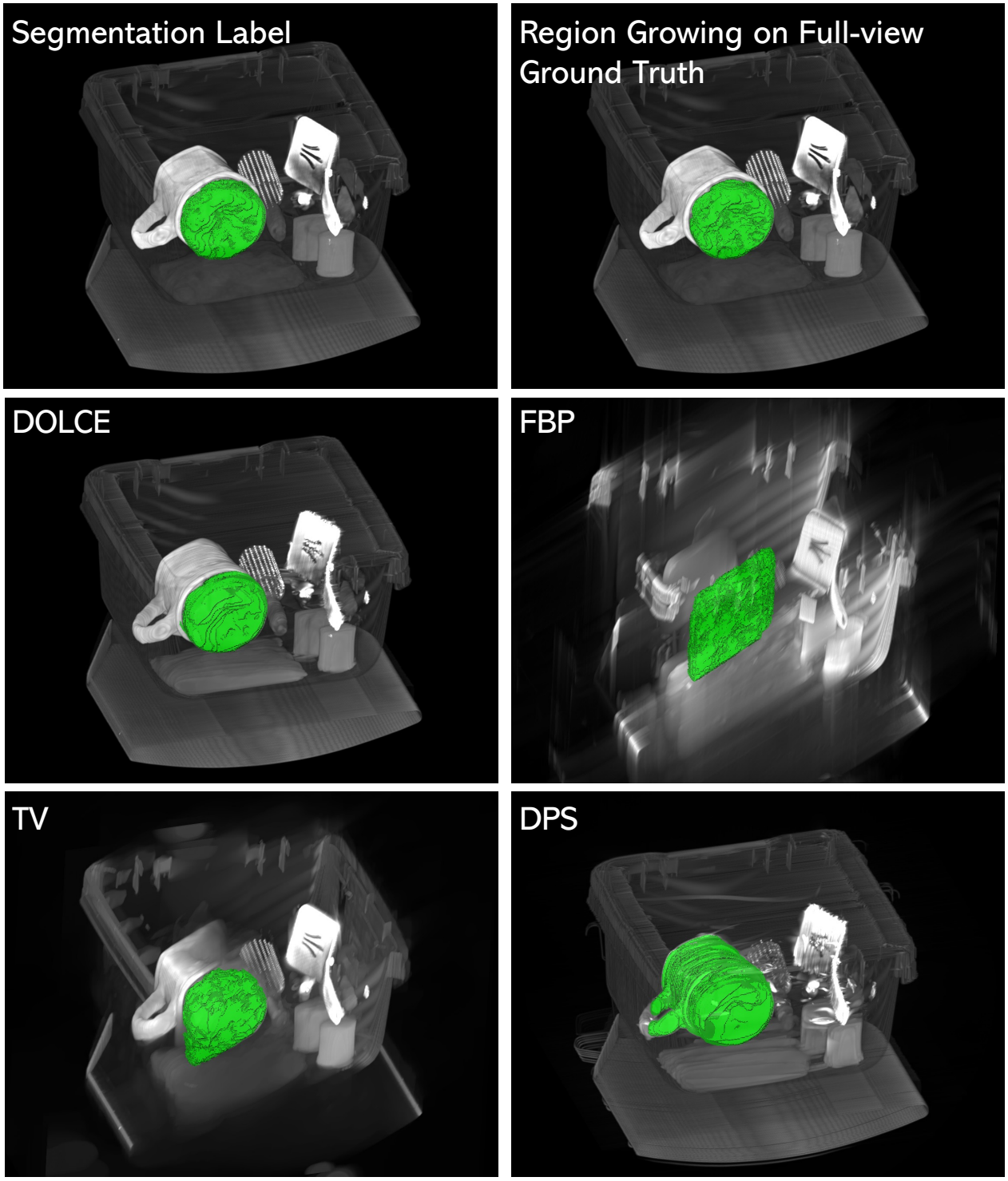

Figure 12. **More 3D Segmentation Results on Test Bag 3:** We use a region growing 3D segmentation in all cases and the resulting segmentations are highlighted in color, against a 3D rendering of the 274 reconstructed 2D slices using $\theta_{\max} = 60^{\circ}$.

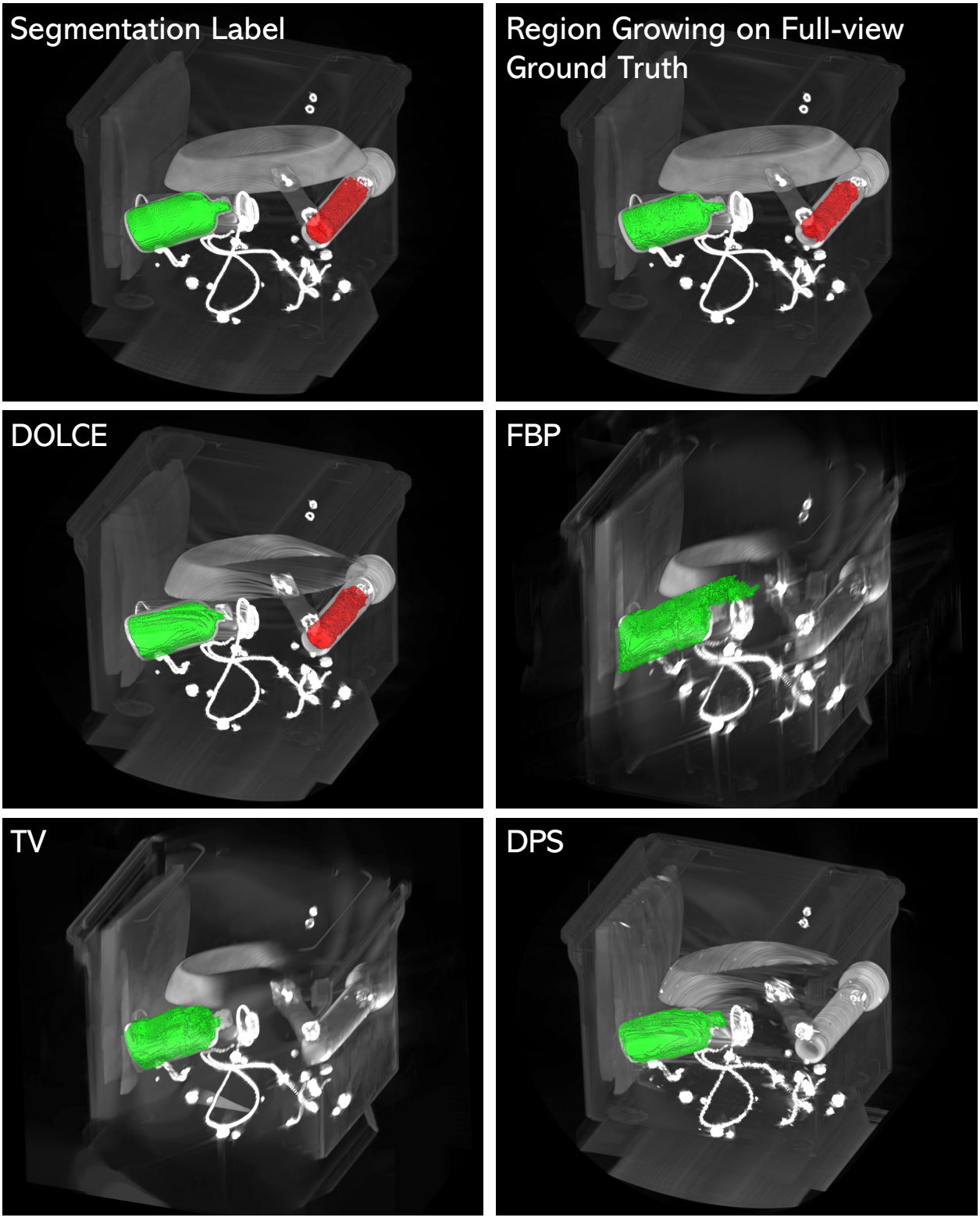

Figure 13. **More 3D Segmentation Results on Test Bag 4:** We use a region growing 3D segmentation in all cases and the resulting segmentations are highlighted in color, against a 3D rendering of the 268 reconstructed 2D slices using $\theta_{\max} = 60^{\circ}$.

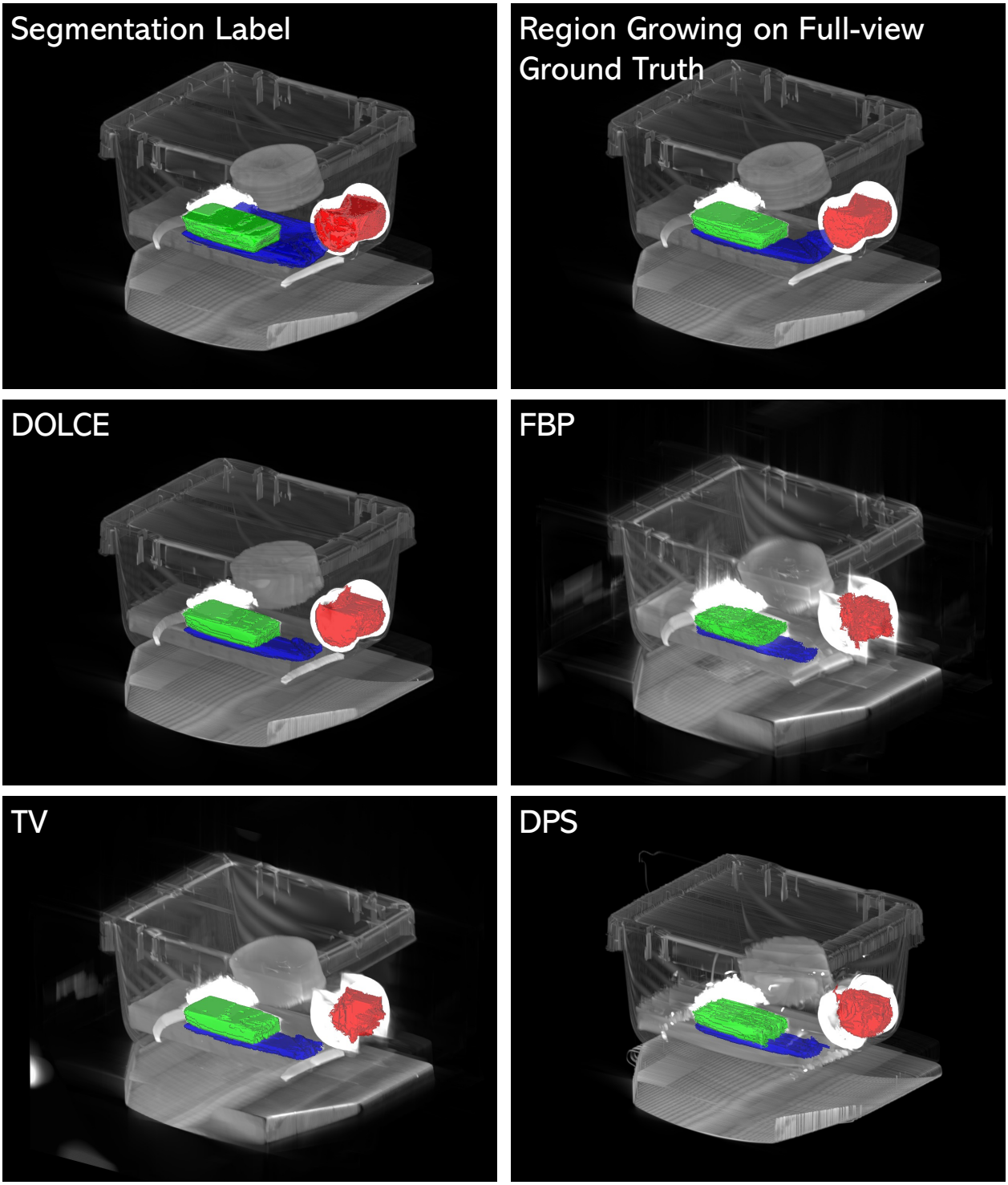

Figure 14. **More 3D Segmentation Results on Test Bag 5:** We use a region growing 3D segmentation in all cases and the resulting segmentations are highlighted in color, against a 3D rendering of the 274 reconstructed 2D slices using $\theta_{\max} = 90^{\circ}$.

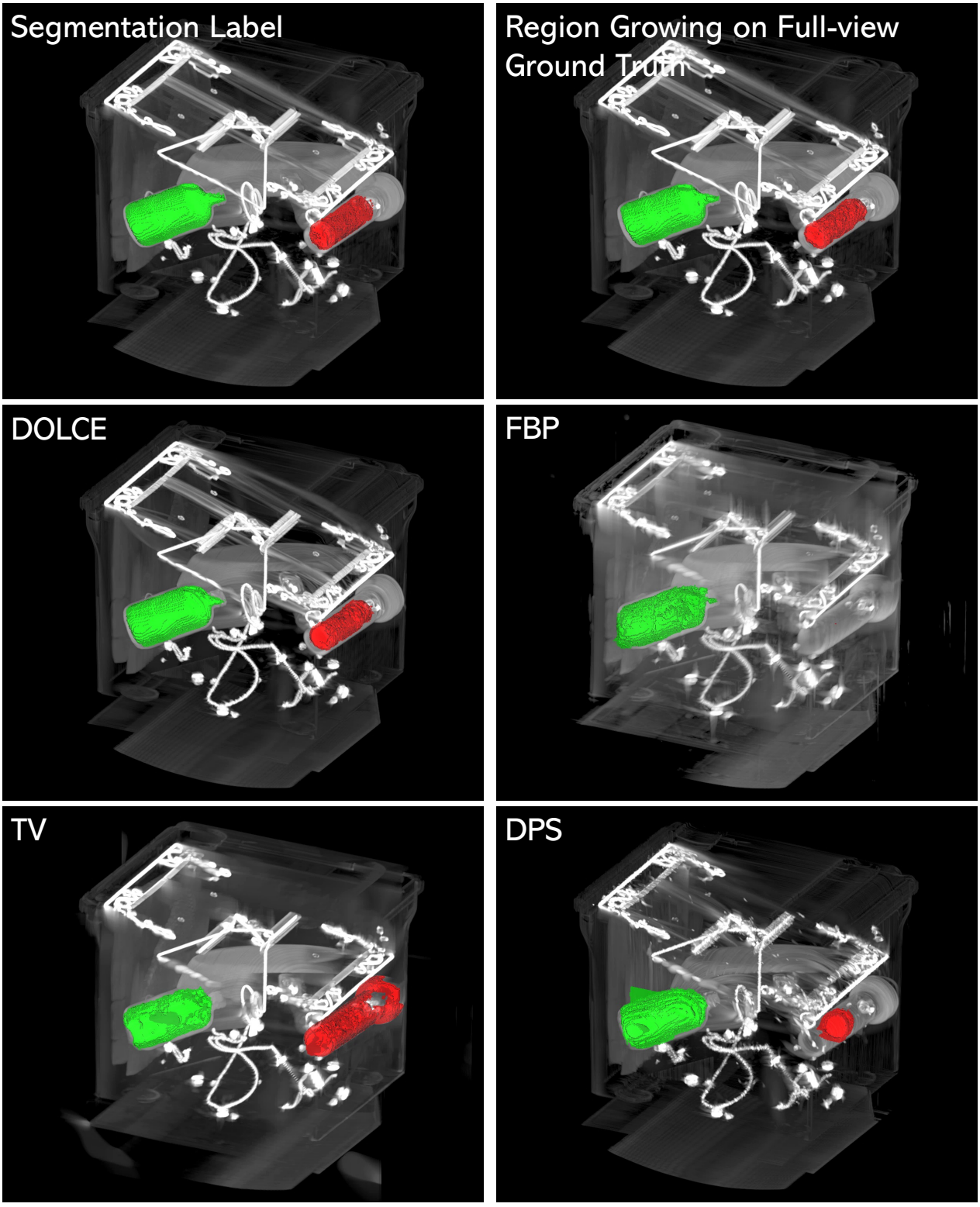

Figure 15. **More 3D Segmentation Results on Test Bag 6:** We use a region growing 3D segmentation in all cases and the resulting segmentations are highlighted in color, against a 3D rendering of the 268 reconstructed 2D slices using $\theta_{\max} = 90^{\circ}$.