Experiments

To further evaluate the single-view reconstruction for the same patient, we have conducted additional testing with a single lateral view for the abdominal CT and lung CT. The left panel of Supplementary Fig. 7 shows the reconstruction results using deep learning model, together with the ground truth images for the abdominal CT. The averaged MAE/RMSE/SSIM/PSNR values over all testing samples for the lateral view reconstruction are 0.015, 0.140, 0.947, and 32.517, respectively, for the abdominal CT. The indices for lung CT are found to be 0.022, 0.387, 0.838, and 27.146, respectively. These values are comparable with that listed in Table 1, suggesting that our deep learning model is capable of reconstructing volumetric images with a single projection from either AP or lateral direction for the same patient.

Moreover, we note that the proposed method works not only for 4D-CT, but also for 3D-CT, primarily because of its ability of incorporating *a priori* knowledge into the reconstruction at training stage with various data augmentation. The essence of our deep learning method is that, different from the existing approaches such as the PCA-based method, it does not rely on a presumed motion model during reconstruction, making it a more general approach for patient-specific tomographic image reconstruction. To illustrate this, we conducted 3D image reconstruction experiments for a head-and-neck case. Specifically, we collected two sets of radiation therapy treatment planning CTs of a head-and-neck patient. The two CTs were acquired with five weeks of time interval. In this study, the first CT image and variant images obtained by data augmentation (a total of 2000 images were generated by translation, rotation and deformation within clinically relevant limits) were used to train the deep network described in this work, whereas the second CT data (and the corresponding augmentation data with this CT) were used for validation. Figure 5a shows the cross-sectional 3D CT images of the head-and-neck case used for model training. In Figure 5b, the testing sample and its difference images with respect to (w.r.t) the training sample are displayed in the left panel. The right panel shows the deep learning-predicted images and their differences w.r.t. the ground truth images in transverse, sagittal, coronal planes, respectively. This study testifies that our method is capable of using a single projection view to reconstruct 3D CT images for the same patient, thus supports our conclusion that the proposed method is valid for patient-specific CT reconstruction beyond abdominal and lung.

Comparison study with PCA-based method

In general, the PCA-based method is valid only for the reconstruction of 4D-CT images under two assumptions: (i) the voxels within the body move correlatively and their motions can be characterized by a few principal components; and (ii) the principal components remain unchanged over time from that of the previous scan (the reference scan) - the only thing that changes from scan to scan is the eigenvalues of the principal components, meaning that a 4D-CT scan at a subsequent time can be obtained from the reference scan by adjusting the eigenvalues. It is interesting that, when the above two assumptions are satisfied, the PCA coefficients of the 4D-CT scan can be obtained from a single projection measurement in such a way that the computed projection (which is a function of the PCA coefficients) matches the measured one [55-57]. In reality, none of the above two assumptions is trivial. For 3D-CT reconstruction (e.g., lung CT case in our manuscript) or, more generally, when the anatomical motion is not describable by principal components, PCA-based method becomes invalid. Deep learning-based model does not have such limitation. In reality, the principal components may change from scan to scan because of inter-scan variation in patient positioning. As will be seen later, when patient position changes slightly, inaccuracy may occur in image reconstruction with the use of PCA components of the reference scan.

For illustration, in Supplementary Fig. 7 we show the reconstructed images obtained by using the two methods (the left and right panels show the results of deep learning and PCA methods, respectively) when a small rotational error of 2.5^{o} occurs with respect to the reference scan. Quantitative comparison of the two methods for the testing samples shown in Supplementary Fig. 7 is summarized in Supplementary Table 1, which further confirms our conclusion above. Different from the model-based method, which assumes that the anatomy variations can be described by a few PCA components, our deep learning method is data-driven, which requires no presumed motion model for image reconstruction. Noteworthy, the manifold mapping function in the data-driven approach is extracted directly from the training datasets during the learning process, instead of relying on any *ad hoc* form of motion trajectory. Because of this important feature, the deep learning approach presents a possible strategy for patient-specific image reconstruction.

Ablative study and discussion

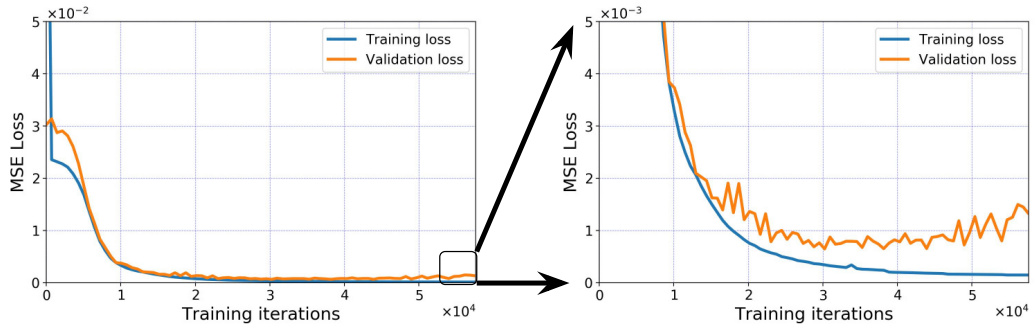
To better understand the proposed model, we conduct ablative studies to analyze the influence of different architecture choices such as residual learning and network depth.

Supplementary Fig. 8 shows the model performance when network depth varies. The value of x-axis stands for the number of convolution (deconvolution) blocks used to construct the representation (generation) network. The results indicate that the depth of network is an important factor affecting the reconstruction performance. In general, a deeper model leads to less prediction error in MAE, RSME and SSIM, and higher image quality as measured by PSNR. This conclusion could be explained through the representation learning process, since the generation of an accurate 3D volumetric image from the learned representations is possible only if the extracted features are sufficiently informative to describe the underlying object or scene.

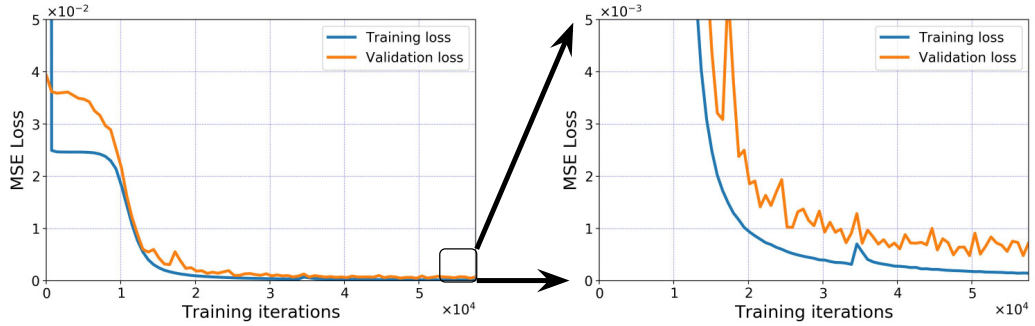
In Supplementary Table 2, we show experiment results of single-view reconstruction for abdominal CT with variant model architecture. For brevity, the “2D-Res” denotes utilizing 2D convolution residual block in representation network and “3D-Res” denotes using 3D deconvolution residual block in generation network. “2D” and “3D” stands for 2D representation network and 3D generation network without residual shortcuts in each block. By comparing the results, we notice that the residual learning is possible to help extract features representation in the model. We also deploy the architecture “2D + 3D” and “2D-Res + 3D-Res” in more experiments and find that, while they perform well in this single-view experiment, the performance gets worse when it comes to 5-view and 10-view reconstruction. All considered, we conclude that the architecture “2D-Res + 3D” provides a viable choice for this patient-specific image reconstruction problem with ultra-sparse views.

Supplementary Fig. 1 | Training loss curves of the image reconstruction in the study of abdominal CT. Blue and orange curves denote training and validation loss respectively. **a-d,** Image reconstructed using 1, 2, 5, and 10 views, respectively.

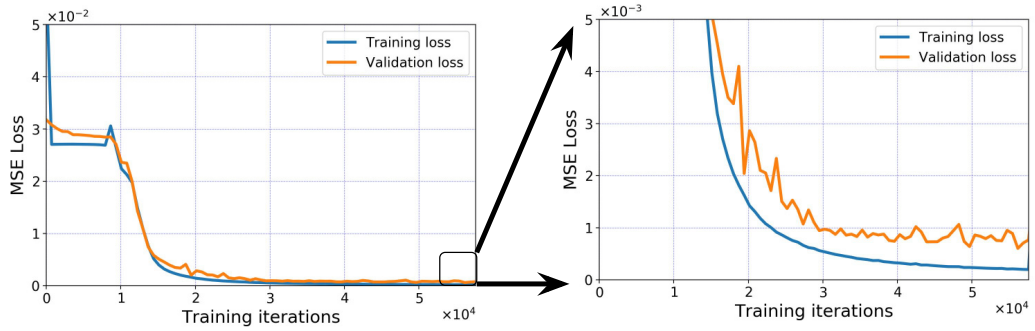
(a) 1 View



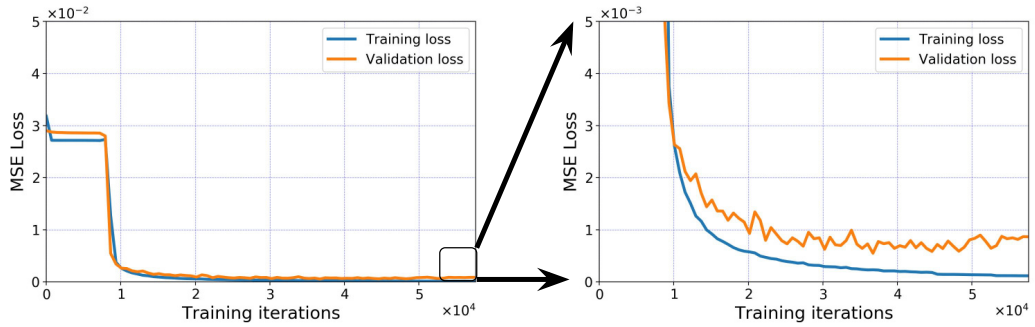
(b) 2 Views



(c) 5 Views

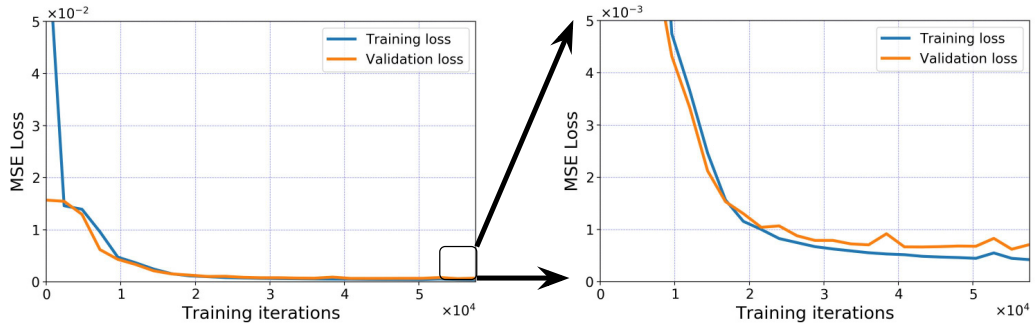


(d) 10 Views

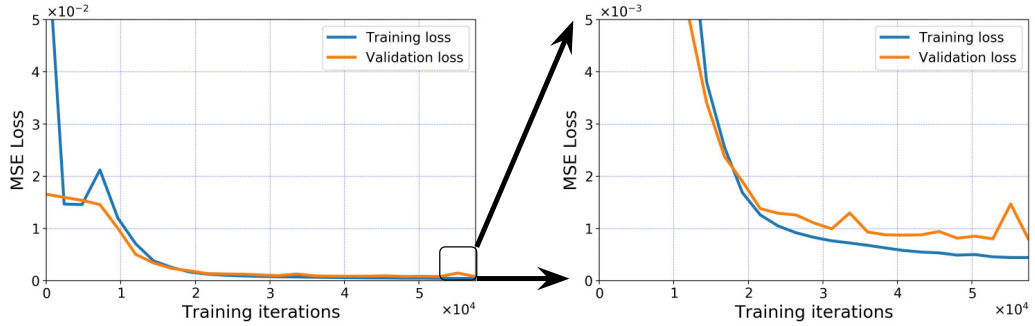


Supplementary Fig. 2 | Training loss curves of the image reconstruction in the study of lung CT. Blue and orange curves denote training and validation loss respectively. **a-d,** Image reconstructed using 1, 2, 5, and 10 views, respectively.

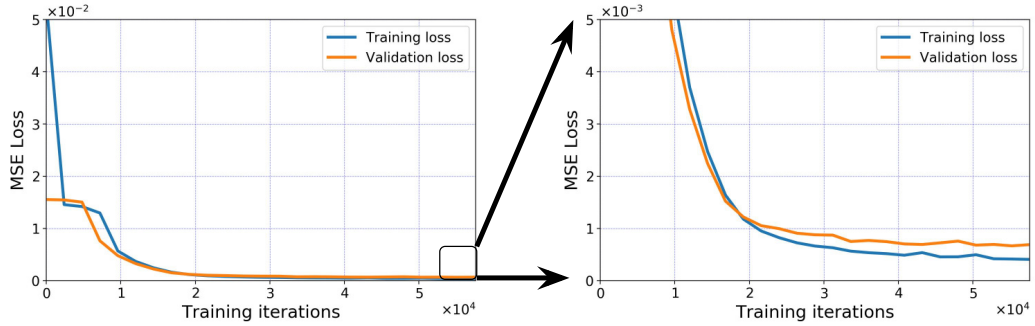
(a) 1 View



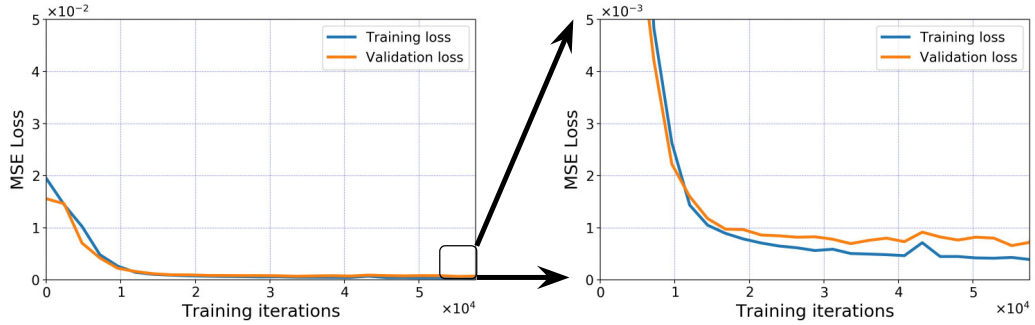
(b) 2 Views



(c) 5 Views



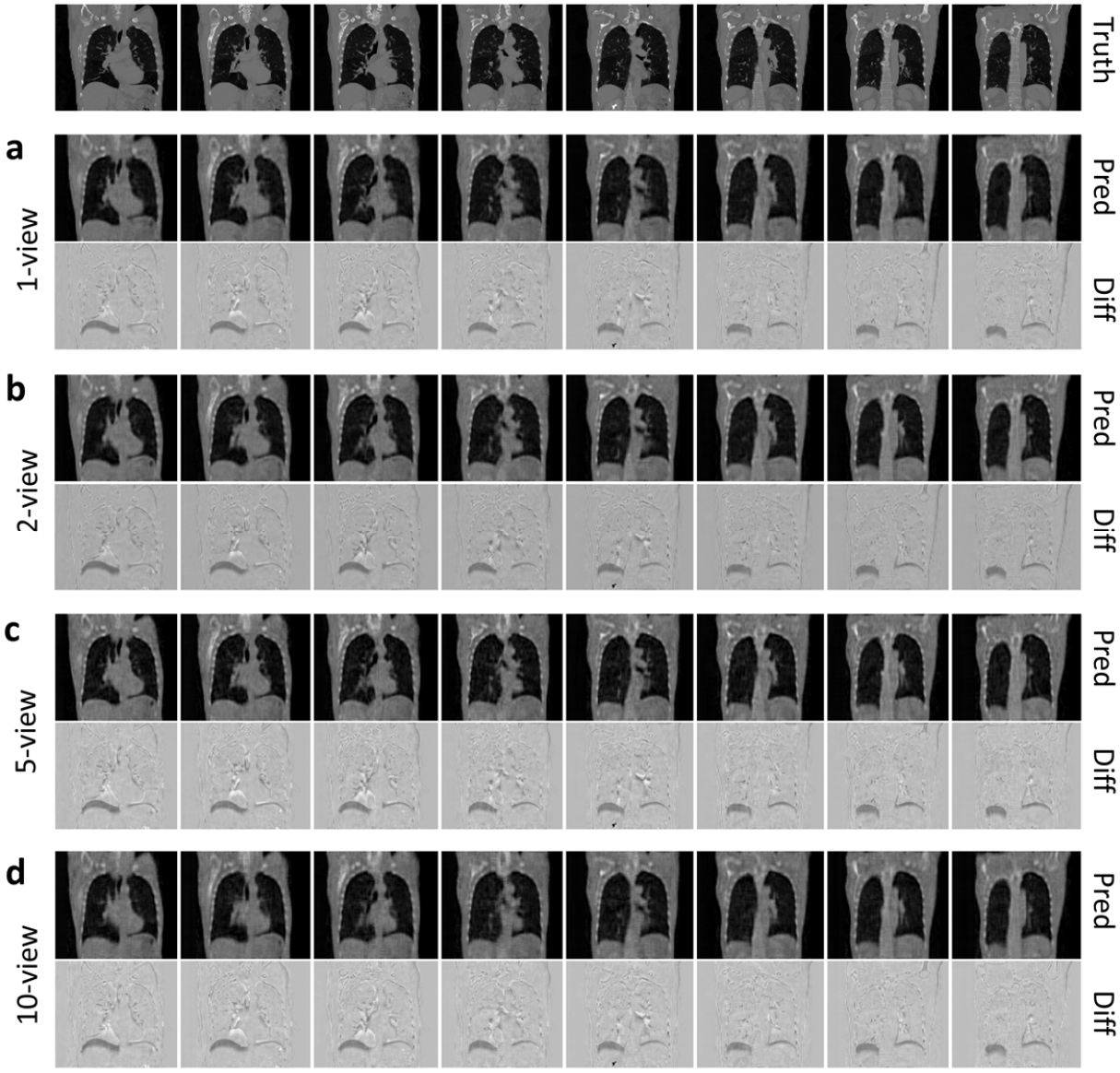
(d) 10 Views



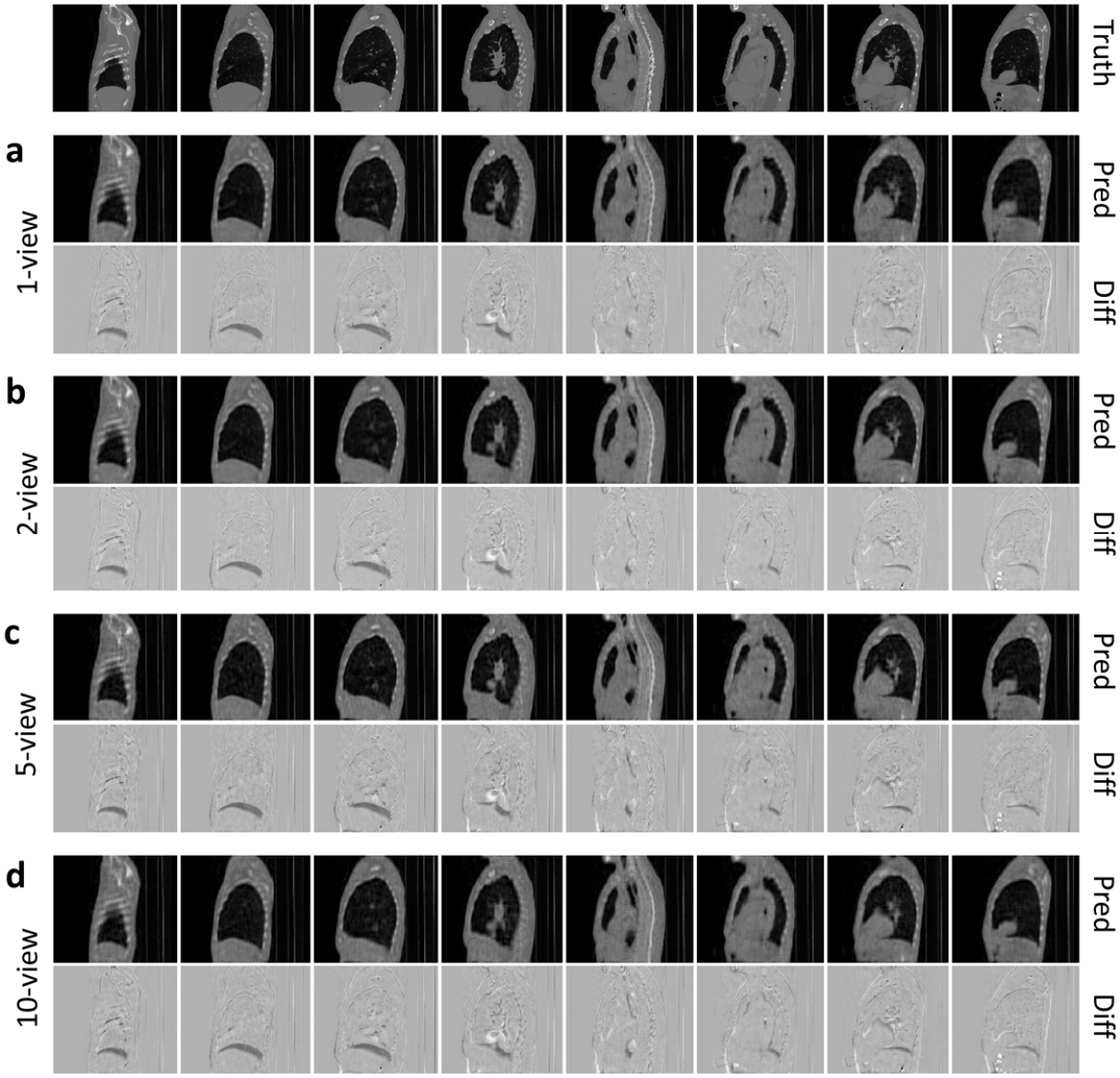
Supplementary Fig. 3 | Coronal images of abdominal CT with different number of 2D projections. Both predicted images (Pred) and difference images (Diff) between the prediction and the corresponding ground truth (Truth) are shown. **a**-**d,** Image reconstructed using 1, 2, 5, and 10 views, respectively.

Supplementary Fig. 4 | Sagittal images of abdominal CT with different number of 2D projections. Both predicted images (Pred) and difference images (Diff) between the prediction and the corresponding ground truth (Truth) are shown. **a**-**d,** Image reconstructed using 1, 2, 5, and 10 views, respectively.

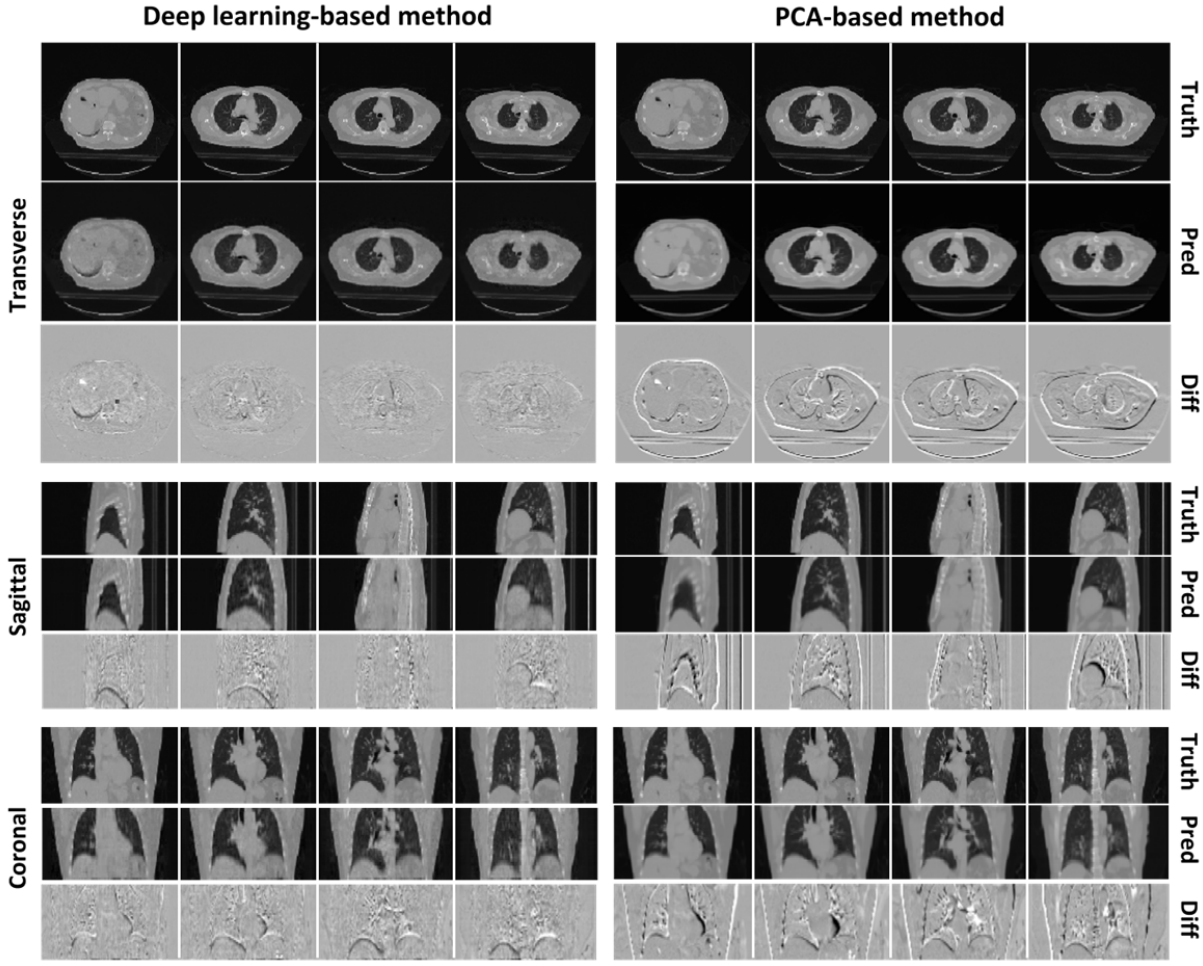
Supplementary Fig. 5 | Coronal images of lung CT with different number of 2D projections. Both predicted images (Pred) and difference images (Diff) between the prediction and the corresponding ground truth (Truth) are shown. **a**-**d,** Image reconstructed using 1, 2, 5, and 10 views, respectively.



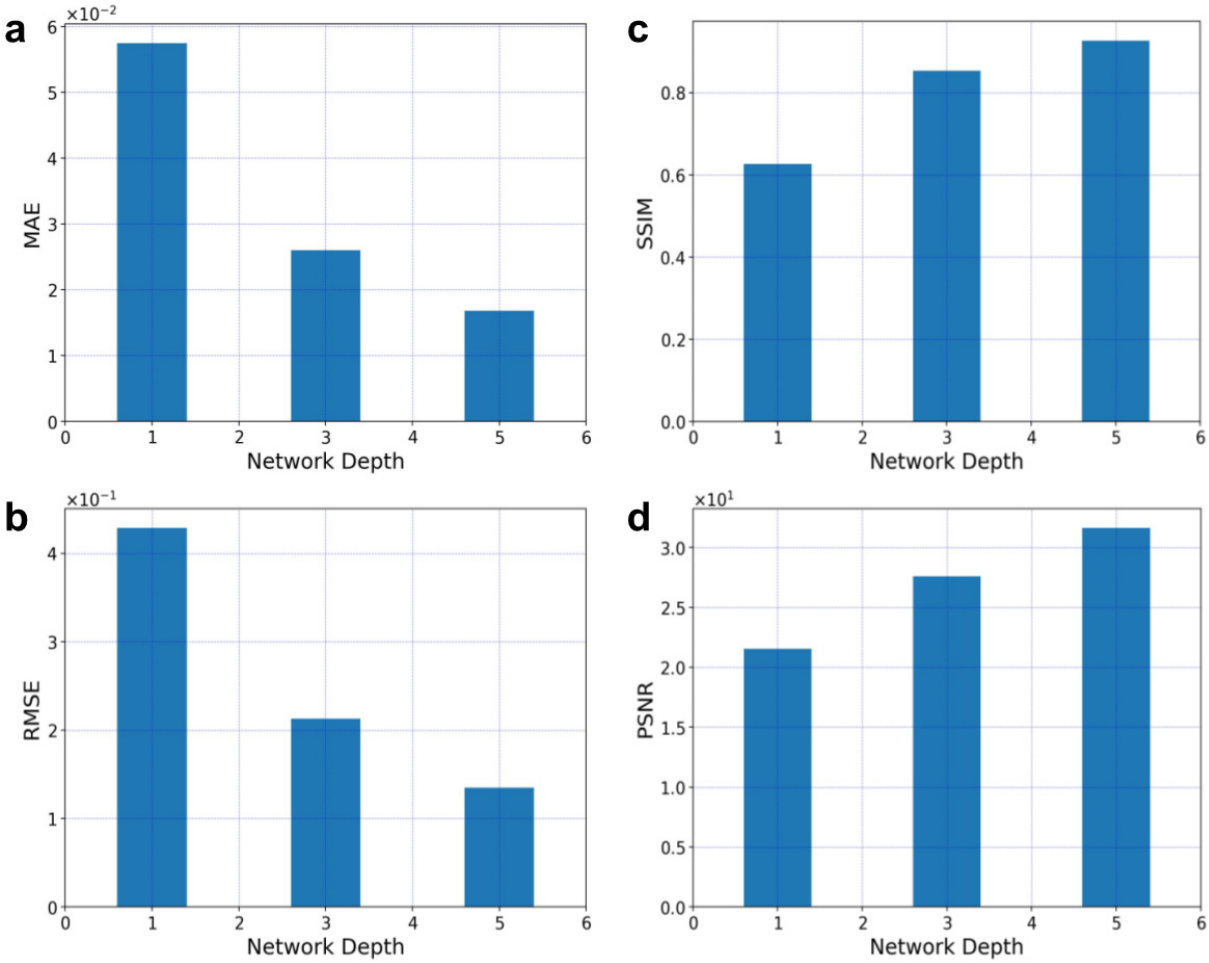
Supplementary Fig. 6 | Sagittal images of lung CT with different number of 2D projections. Both predicted images (Pred) and difference images (Diff) between the prediction and the corresponding ground truth(Truth) are shown. **a**-**d,** Image reconstructed using 1, 2, 5, and 10 views, respectively.



Supplementary Fig. 7 | Deep learning-based reconstruction with a single lateral-view projection for abdominal CT and comparison with PCA-based method. Left and right panels show the images reconstructed by using deep learning and PCA methods, respectively. These results and quantitative evaluation show that, in the presence of patient positioning inaccuracy, deep learning model outperforms the PCA-based method.



Supplementary Fig. 8 | Comparison of network depth in single-view reconstruction of abdominal CT. The network depth denotes the number of convolutional blocks used in representation and generation network. We evaluate the model performance on testing dataset through four metrics: **a**, MAE, **b**, RMSE, **c**, SSIM, and **d**, PSNR.



Supplementary Table 1 | Comparison of deep learning and principal component analysis.

Reconstruction View	Deep-learning-based method				PCA-based method			
	MAE	**RMSE**	**SSIM**	**PSNR**	**MAE**	**RMSE**	**SSIM**	**PSNR**
Lateral view	0.010	0.097	0.966	35.829	0.014	0.213	0.915	28.967

MAE, mean absolute error; RMSE, root mean squared error; SSIM, structural similarity; PSNR, peak signal noise ratio.

Supplementary Table 2 | Comparison of network architecture in single-view reconstruction of abdominal CT. "2D-Res" denotes utilizing 2D convolution residual block in representation network and "3D-Res" denotes utilizing 3D deconvolution residual block in generation network. "2D" and "3D" stands for 2D representation network and 3D generation network without residual shortcuts in each block.

Architecture Choice	Abdominal CT (single-view)			
	MAE	**RMSE**	**SSIM**	**PSNR**
2D + 3D	0.017	0.161	0.936	31.495
2D-Res + 3D	0.018	0.177	0.929	30.523
2D + 3D-Res	0.019	0.196	0.921	29.659
2D-Res + 3D-Res	0.016	0.157	0.939	31.603

MAE, mean absolute error; RMSE, root mean squared error; SSIM, structural similarity; PSNR, peak signal noise ratio.