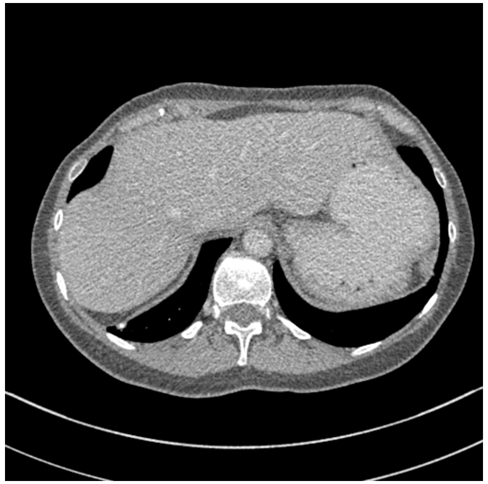
Transfer Learning for Low-Dose CT Denoising

Hongming Shan, Yi Zhang, Qingsong Yang, Uwe Kruger, Wenxiang Cong and Ge Wang
Biomedical Imaging Center, CBIS/BME/SoE
Rensselaer Polytechnic Institute
SHANH@RPI.EDU

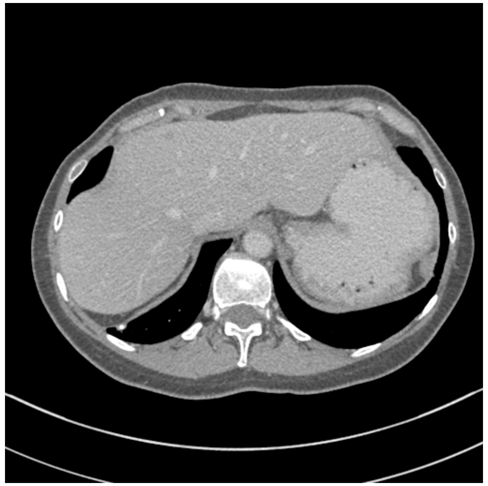
November 19, 2017

Low-Dose CT

- CT-associated high-dose x-ray radiation carries health risks for patients.
- Reduction of the radiation dose compromises CT image quality, and the resultant image noise can compromise diagnostic information.



Quarter-dose



Full-dose

Images are from *2016 NIH-AAPM-Mayo Clinic Low-Dose CT Grand Challenge*

Noise Reduction for Low-Dose CT

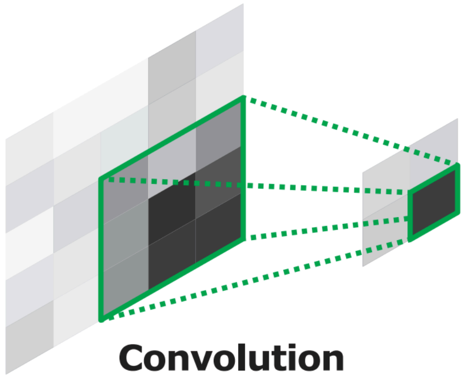
- **Sinogram filtration**
  - Perform on either raw data or log-transformed data
- **Iterative reconstruction**
  - Optimize an objective function that combines the statistical properties of data in the sinogram domain and prior information in the image domain together
- **Post-processing techniques**
  - Operate on an image directly which has been reconstructed from raw data.
  - Deep learning-based methods achieving impressive results.

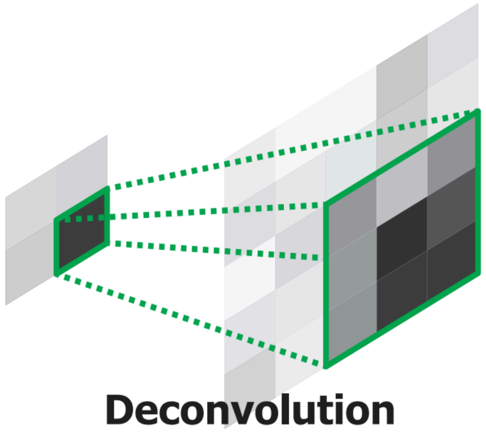
Deep Learning-based Denoising Method

- **Network architecture**: Complexity of model
  - Convolutional layer
  - Deconvolutional layer
  - Special connection
- **Objective function**: How to learn from image/data
  - Mean squared error (MSE), as well as L1 norm (Enhao’s talk)
  - Adversarial loss
  - Perceptual loss

Network architecture

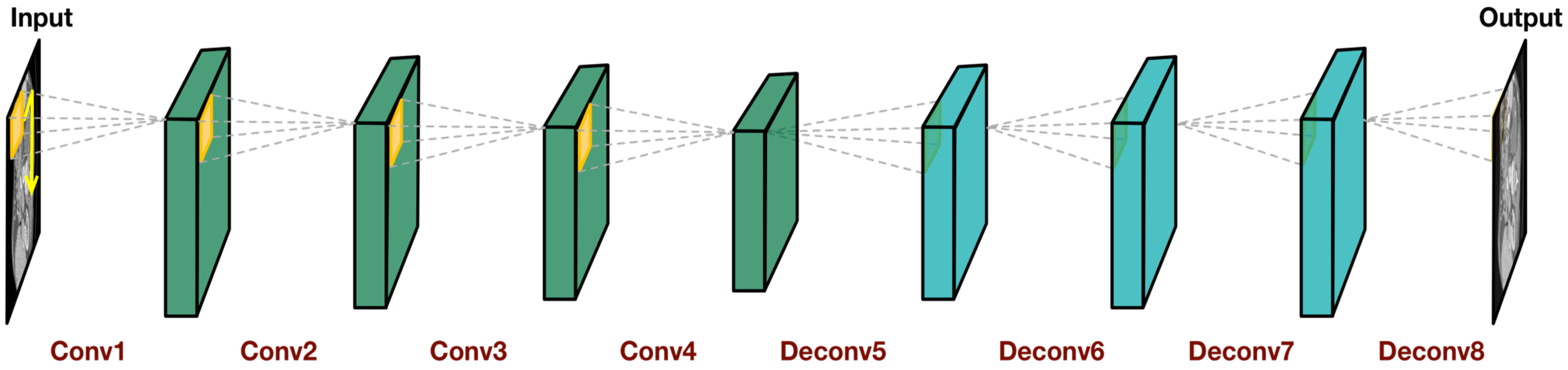
Methods	Network architecture		
	Conv. Layer	Deconv. Layer	Special Connection
CNN[1]	√	-	-
RED-CNN[2]	√	√	Shortcut
GAN-3D[3]	√	-	-
CNN-Cascade[4]	√	-	Cascade
WGAN-VGG[5]	√	-	-
Ours	√	√	Contracting





1. H. Chen, Y. Zhang, W. Zhang, P. Liao, K. Li, J. Zhou, and G. Wang, "Low-dose CT via convolutional neural network," Biomed. Opt. Express,, 2017.
2. H. Chen, Y. Zhang, M. K. Kalra, F. Lin, P. Liao, J. Zhou, and G. Wang, "Low-dose CT with a residual encoder-decoder convolutional neural network (RED-CNN)," IEEE Trans. Med. Imaging, 2017.
3. J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Isgum, "Generative adversarial networks for noise reduction in low-dose CT," IEEE Trans. Med. Imaging, 2017.
4. D. Wu, K. Kim, G. E. Fakhri, and Q. Li, "A cascaded convolutional nerual network for x-ray low-dose CT image denoising," arXiv preprint arXiv:1705.04267, 2017.
5. Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, and G. Wang, "Low dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss," arXiv preprint arXiv:1708.00961, 2017.

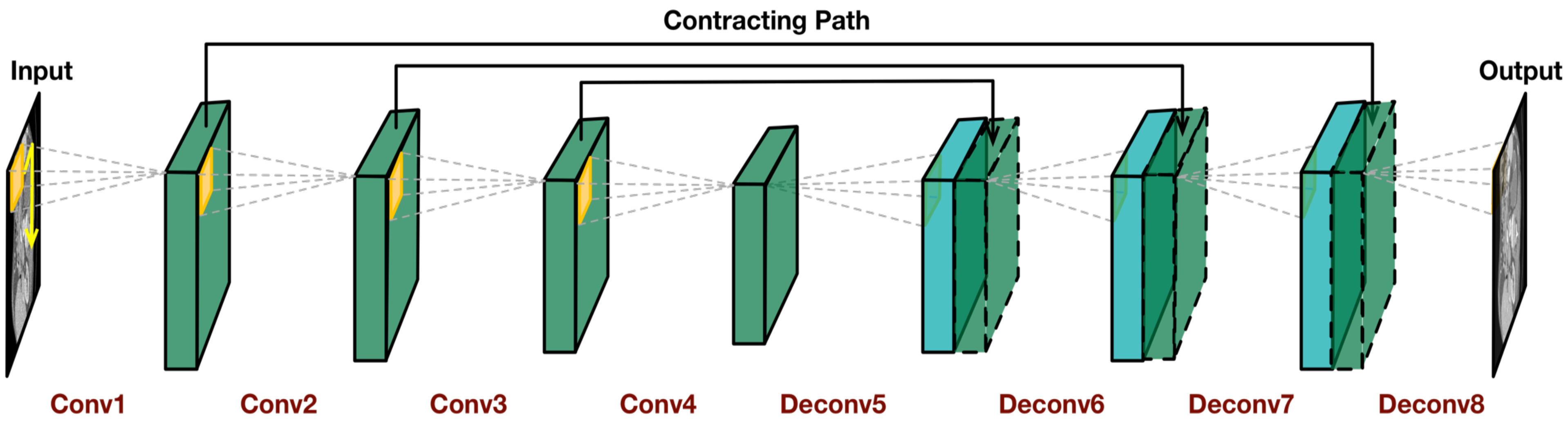
Convolutional Autoencoder (CA)



Traditional convolutional autoencoder includes convolutional layers and deconvolutional layers

- encoding low-dose CT image
- decoding to reconstruct normal-dose CT image

Contracting Path Convolutional Autoencoder (CPCA)



Contracting path copies the preceding feature maps and reuses them at later layers with the same feature-map sizes, preserving the details of the high resolution features.

- U-net[1]
- DenseNet[2]

1. O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in Int. Conf. Med. Image Comput. Comput. Assist. Interv, Springer, 2015.
2. G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten, “Densely connected convolutional networks,” in Proc. IEEE Conf. Comp. Vis. Patt. Recogn., 2017.

Objective function

Methods	Objective function		
	MSE	Adversarial Loss	Perceptual Loss
CNN[1]	√	-	-
RED-CNN[2]	√	-	-
GAN-3D[3]	√	√	-
CNN-Cascade[4]	√	-	-
WGAN-VGG[5]	-	√	√
Ours	-	√	√

MSE: Pixel-wise difference, Regression-to-Mean

Adversarial loss: Capture texture information, from same distribution, but samples are not matched very well

Perceptual loss: Measure similarity in feature space, parameters-fixed network

Objective Function

- Adversarial loss

$$\mathcal{L}_a = \mathbb{E}\Big[D(\boldsymbol{I}_{est})\Big] - \mathbb{E}\Big[D(\boldsymbol{I}_{ND})\Big] + \lambda\mathbb{E}\Big[\big(\|\nabla D(\bar{\boldsymbol{I}})\|_2 - 1\big)^2\Big]$$

- Perceptual loss

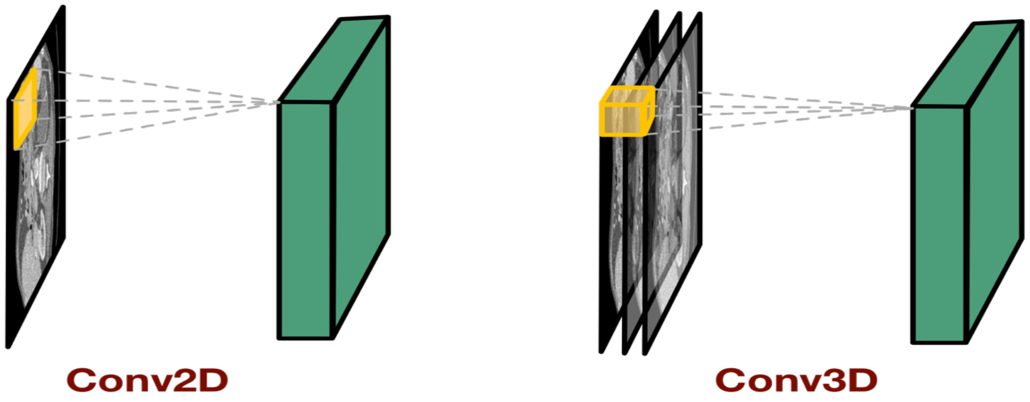
$$\mathcal{L}_p = \|\phi(\boldsymbol{I}_{est}) - \phi(\boldsymbol{I}_{ND})\|_2^2.$$

- Objective function

$$\mathcal{L} = \mathcal{L}_a + \lambda_p\mathcal{L}_p.$$

3D Denoising model

- Spatial information from adjacent LDCT slices
  - Most of the existing denoising networks focus on image denoising in 2D.
  - The adjacent image slices in a CT volume have strong correlative features that can potentially improve 2D-based image denoising.
- For example, we input one image with its 2 adjacent slices.
  - **Input**: Augment one LDCT image with three LDCT images;
  - **Filter**: Replace a 3×3 convolutional filter with a $3\times3\times3$ convolutional filter



Training 3D Denoising Model

~~Training from scratch?~~

Do transfer learning from a trained 2D model

2D filter to 3D filter

- We proposed a simple yet effective way to do transform from 2D filter to 3D filter
- Assume we have 2D filter $\boldsymbol{H} \in \mathbb{R}^{3\times3}$, then corresponding 3D filter $\boldsymbol{B} \in \mathbb{R}^{3\times3\times3}$ is

$$\boldsymbol{B}_{(0)} = \mathbf{0}_{3\times3}, \ \boldsymbol{B}_{(1)} = \boldsymbol{H}_{3\times3}, \ \boldsymbol{B}_{(2)} = \mathbf{0}_{3\times3}.$$

- In this way, the 2D neural network and 3D neural network have same performance, then do fine-tuning to learn spatial information from adjacent slices.
- Spatial information is unknown for network, let it learn from data
  - Suitable for any thickness in CT

Interpretation

- Under GAN framework, Generator **G** and Discriminator **D** are against each other.
  - **D** tells differences between fake samples and real samples
  - **G** fools **D** by generating more similar samples
  - **D** depends on **G**
  - **G** depends on **D**

Balance between G and D is very important.
Do not try to break it.

Experimental Data

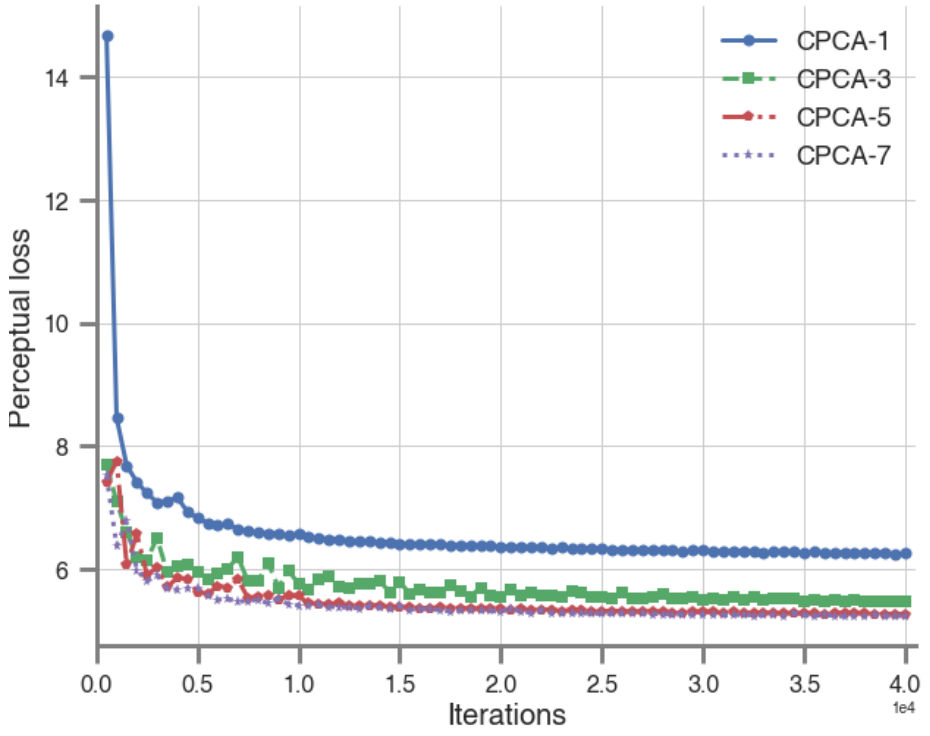
- Experimental data from Mayo Clinic Low-Dose CT Grand Challenge
- Input: Quarter-dose CT images
- Output: Full-dose CT images
- Training data: 128K patches of size 64×64
- Validation data: 64K patches of size 64×64

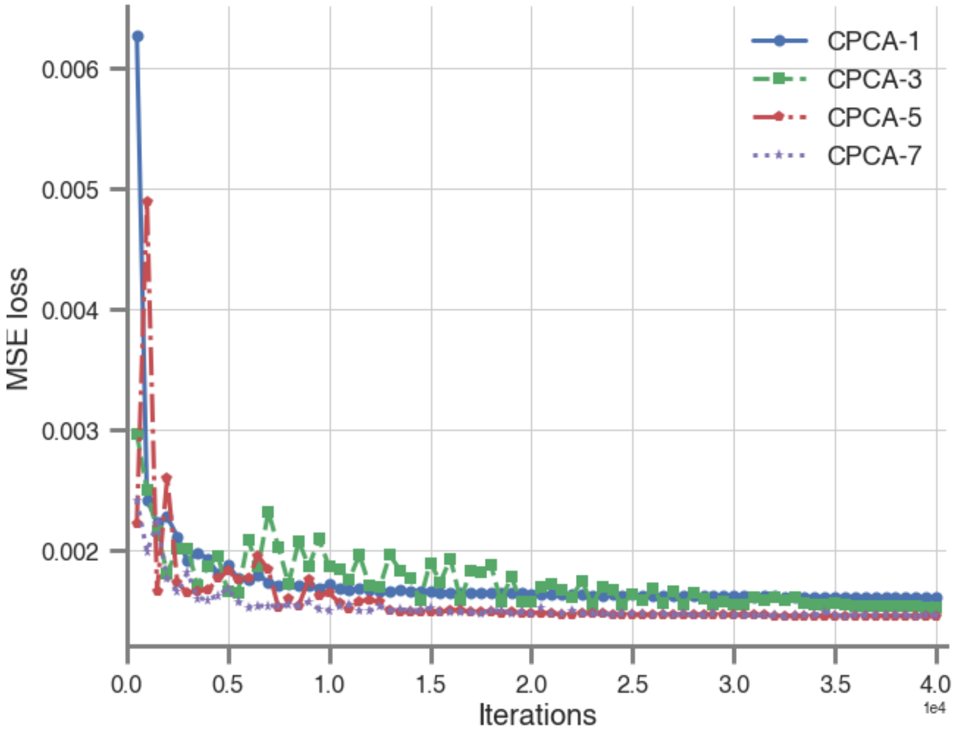
Network Parameters

- No. of feature maps is 32 except for last layer which has only 1.
- Filter size: 3×3, stride is 1.
- ReLU is used after each convolutional layer.
- 1×1 convolutional layer is used to reduce number of feature maps from 64 to 32 after each contracting path.
- Hyperparameter $\lambda_p = 0.1$ via cross-validation
- Learning rate for training from scratch: 1.0×10^{-4}.
- Learning rate for transfer learning from 2D: 0.5×10^{-4}. (fine-tuning)
- Learning rate decays as epoch goes.
- Adam is used for optimization

Comparison: Training from Scratch

- CPCA-i denotes i slices are fed into CPCA.
  - $i = 1$: 2D NN
  - $i = 3, 5, 7$: 3D NN in our experiments.
- Validation results



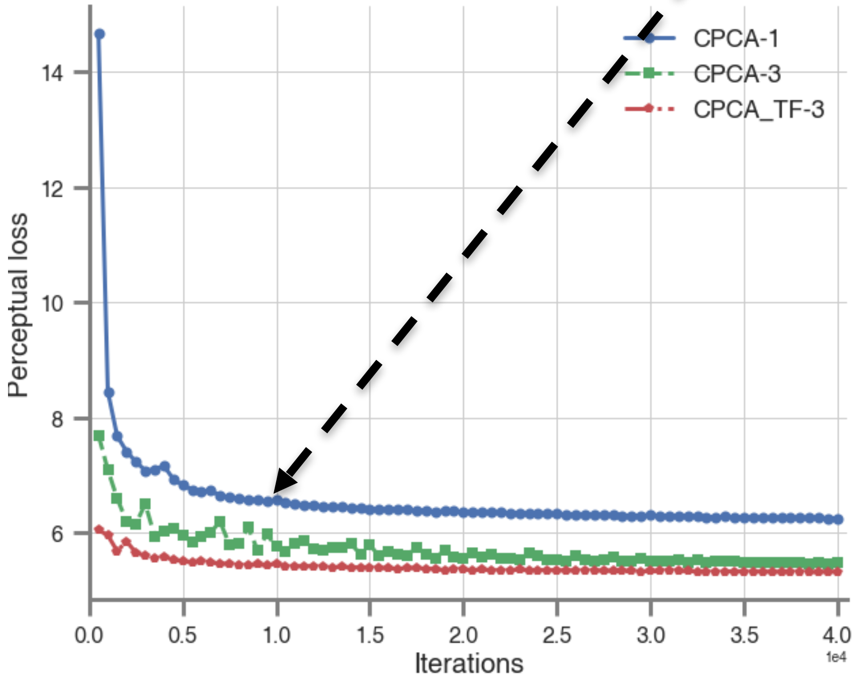


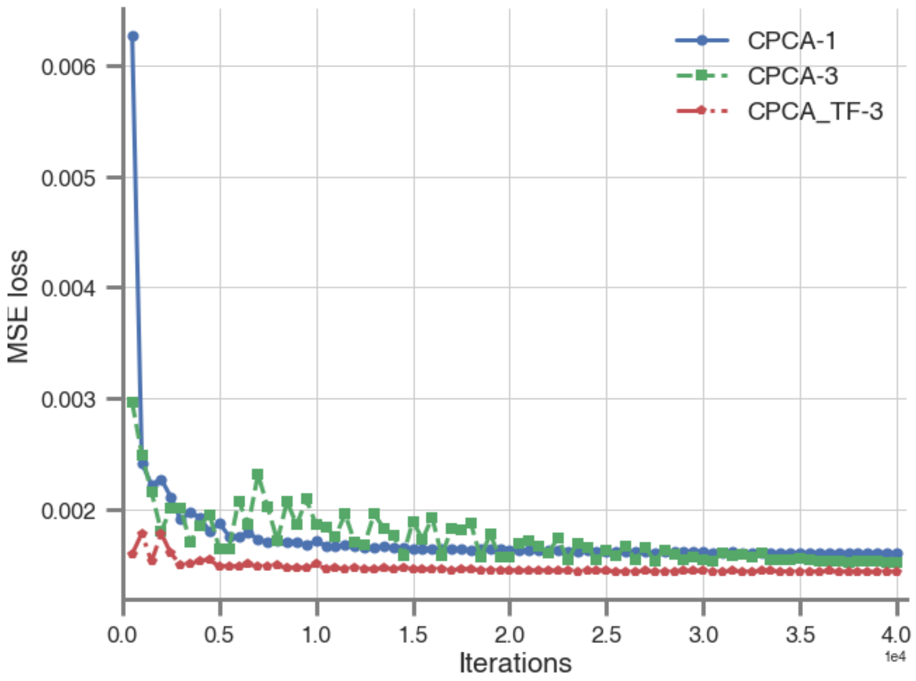
Transfer Learning v.s. Training from Scratch

Transfer learning from a trained 2D model at epoch 10

Input : 3 slices

Transferred from this point



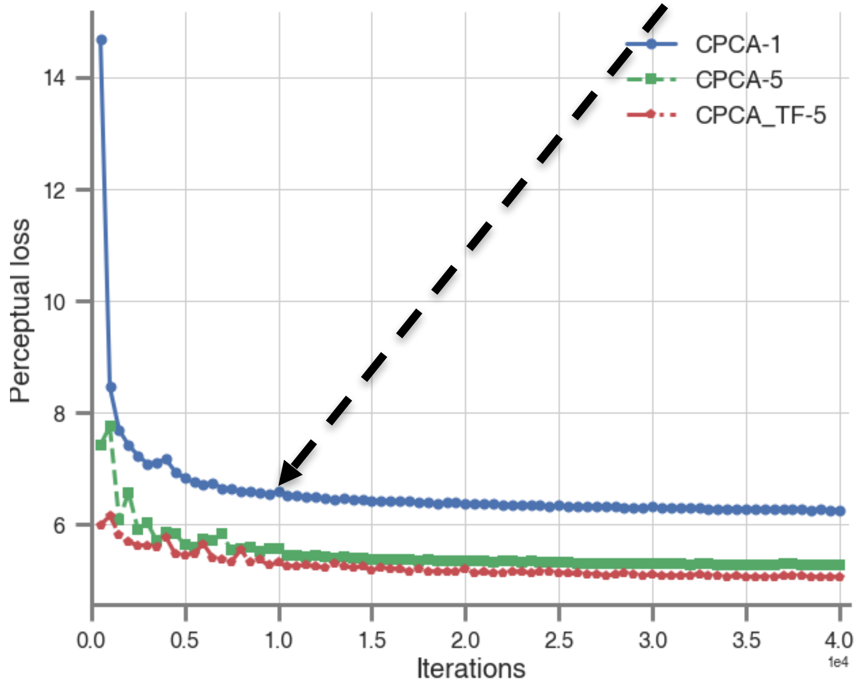


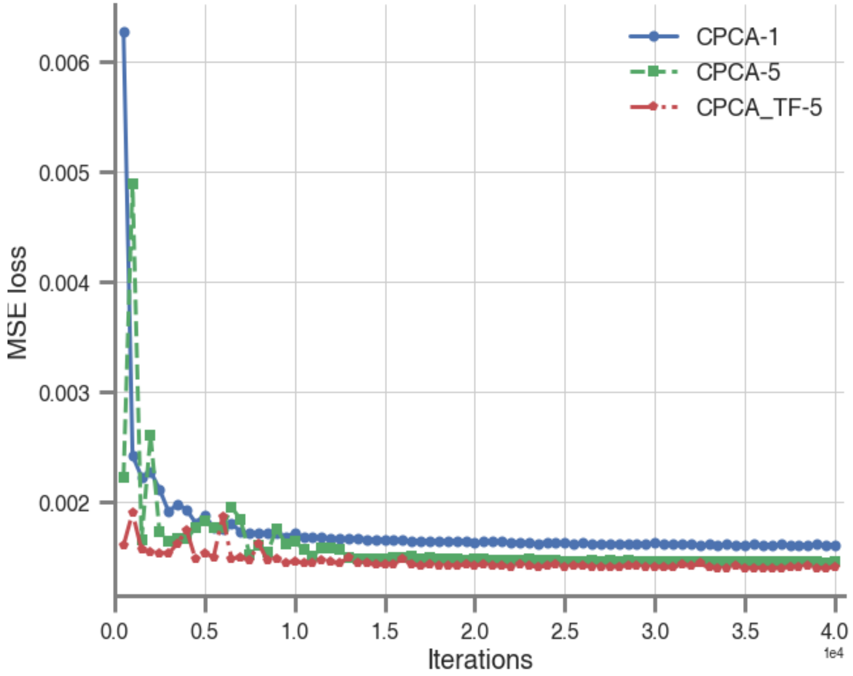
Transfer Learning v.s. Training from Scratch

Transfer learning from a trained 2D model at epoch 10

Input : 5 slices

Transferred from this point



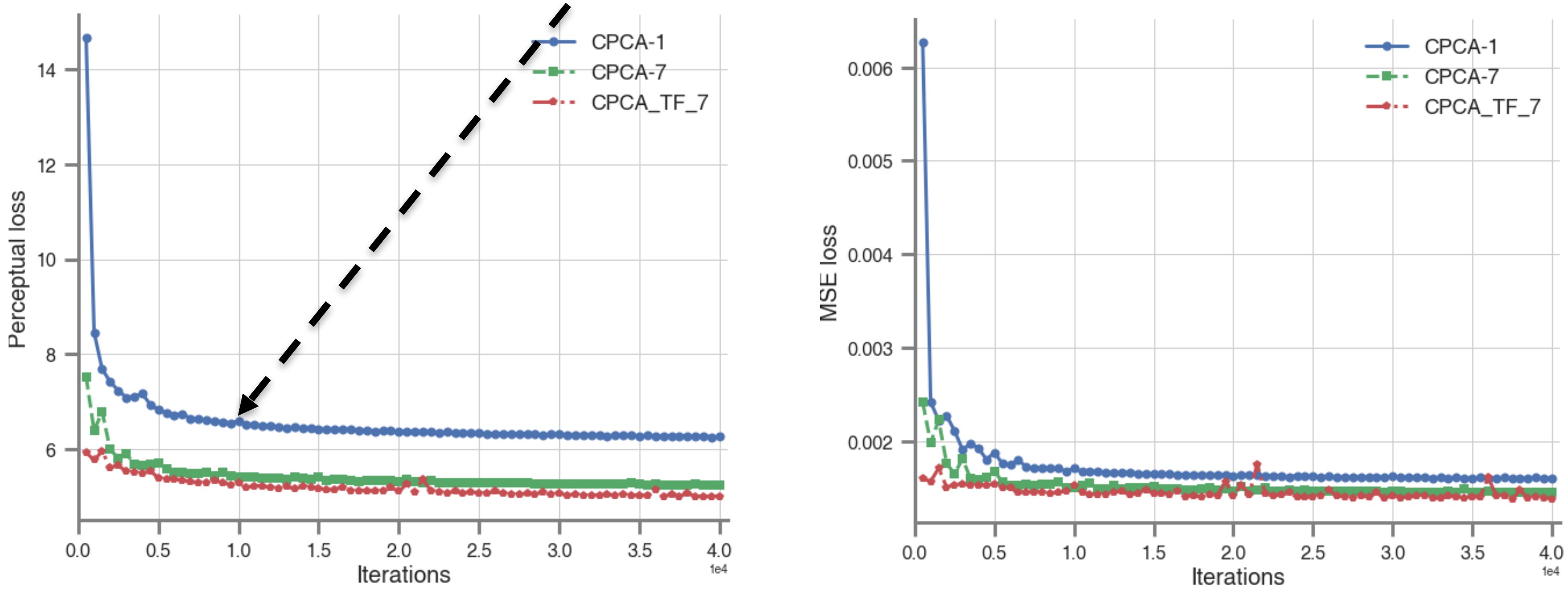


Transfer Learning v.s. Training from Scratch

Transfer learning from a trained 2D model at epoch 10

Input : 7 slices

Transferred from this point



Comparison with State-of-the-Art

- Testing the trained denoising model on full-size CT image (1300 of size 512x512 in total)
- Comparing with recently published methods
  - REDCNN[1]
  - WGAN-VGG[2]

1. H. Chen, Y. Zhang, M. K. Kalra, F. Lin, P. Liao, J. Zhou, and G. Wang, “Low-dose CT with a residual encoder-decoder convolutional neural network (RED-CNN),” IEEE Trans. Med. Imaging, 2017.
2. Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, and G. Wang, “Low dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss,” arXiv preprint arXiv:1708.00961, 2017.

Quantitative Analysis

	PSNR	SSIM	Perceptual Loss
Quarter-Dose	26.07	0.8340	4.81
RED-CNN	31.39	0.9194	4.31
WGAN-VGG	28.88	0.8957	2.55
CPCA-1	29.62	0.8976	2.37
CPCA-3	29.84	0.9004	2.06
CPCA-5	30.00	0.9023	1.99
CPCA-7	**30.01**	**0.9029**	**1.96**

RED-CNN: optimization using MSE loss leads to blurry output images due to regression-to-mean problem.

Quantitative Analysis

	PSNR	SSIM	Perceptual Loss
Quarter-Dose	26.07	0.8340	4.81
RED-CNN	31.39	0.9194	4.31
WGAN-VGG	28.88	0.8957	2.55
CPCA-1	29.62	0.8976	2.37
CPCA-3	29.84	0.9004	2.06
CPCA_TF-3	30.00	0.9031	2.01
CPCA-5	30.00	0.9023	1.99
CPCA_TF-5	30.04	0.9032	1.90
CPCA-7	30.01	0.9029	1.96
CPCA_TF-7	**30.14**	**0.9045**	**1.87**

Case Study: [-180, 200]HU

PSNR:24.99
SSIM: 0.792
P.Los.:5.33

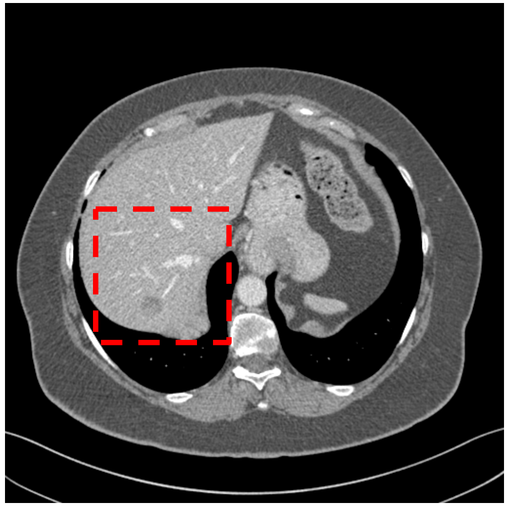
Quarter-Dose

PSNR:**30.67**
SSIM: **0.901**
P.Los.:4.76

RED-CNN

PSNR:28.62
SSIM:0.783
P.Los.:2.76

WGAN-VGG



Full-Dose

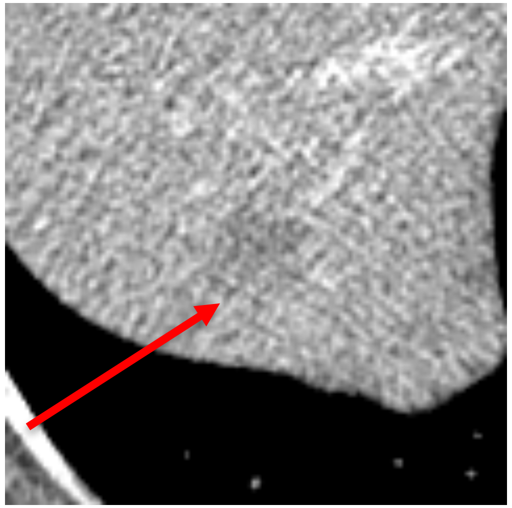
PSNR:28.73
SSIM: 0.870
P.Los.:2.43

CPCA-1

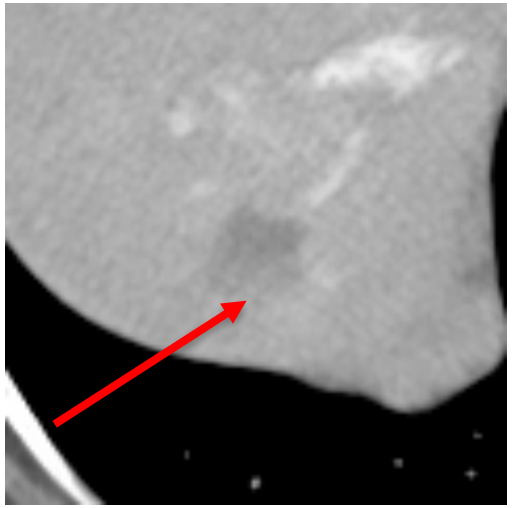
PSNR:**29.20**
SSIM: **0.878**
P.Los.:**2.29**

CPCA_TF-7

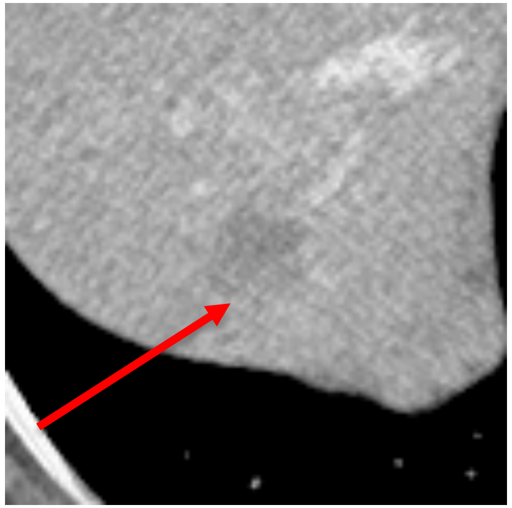
ROI: Metastasis



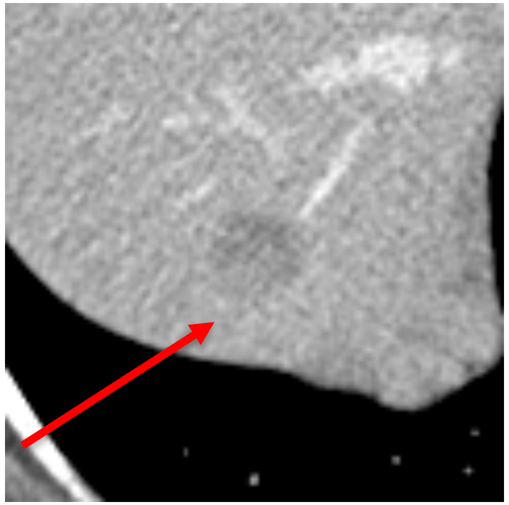
Quarter-Dose



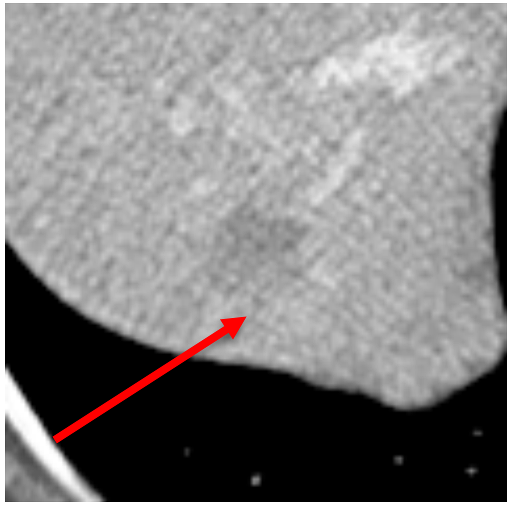
RED-CNN



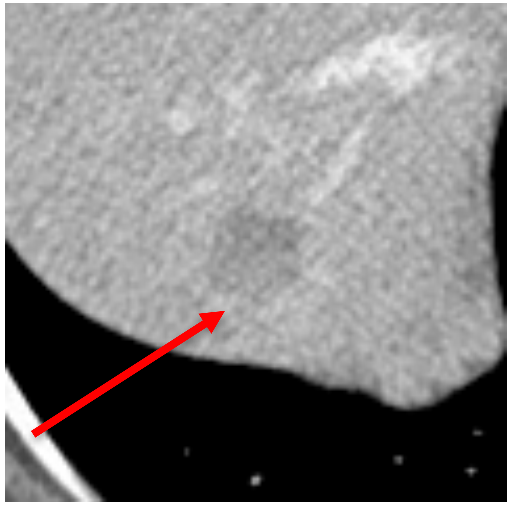
WGAN-VGG



Full-Dose



CPCA-1



CPCA_TF-7

Case Study: [-160, 240]HU

Quarter-Dose

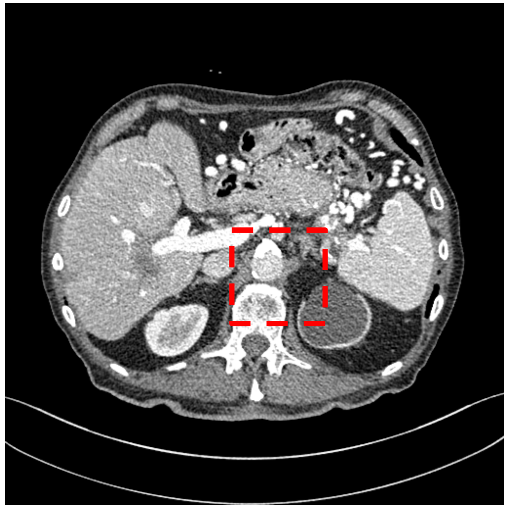
PSNR:22.82
SSIM: 0.799
P.Los.:6.25

RED-CNN

PSNR:**28.28**
SSIM: **0.886**
P.Los.:5.08

WGAN-VGG

PSNR:26.28
SSIM: 0.863
P.Los.: 2.82



Full-Dose

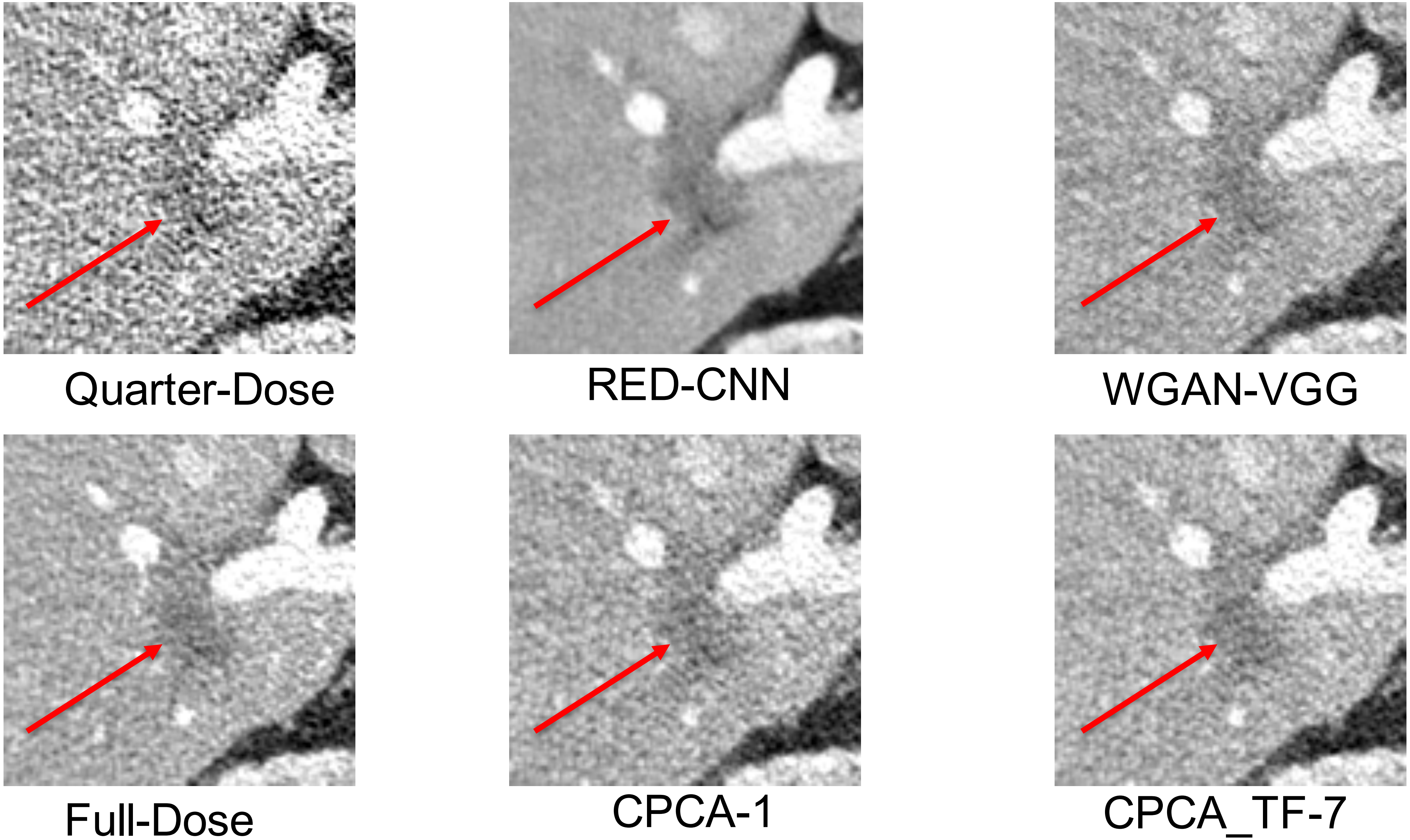
CPCA-1

PSNR:26.67
SSIM: 0.867
P.Los.:2.60

CPCA_TF-7

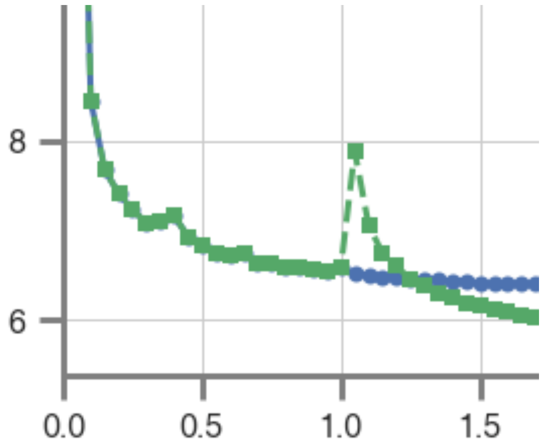
PSNR:**27.12**
SSIM: **0.872**
P.Los.:**2.17**

ROI: Metastasis

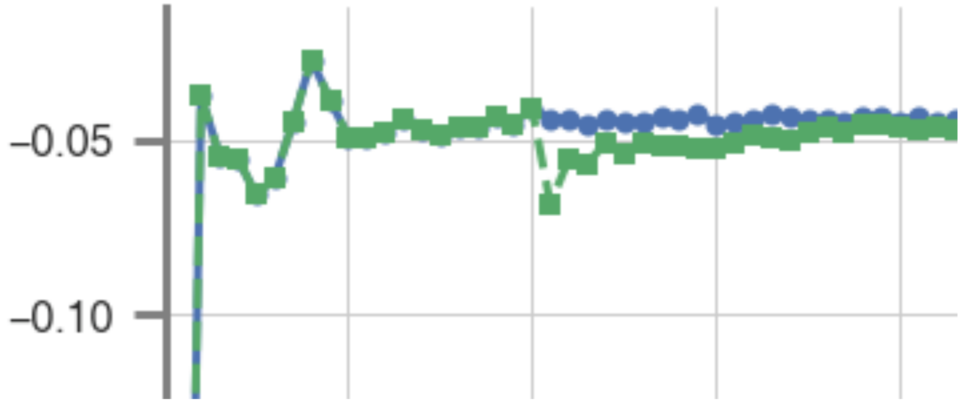


Discussion

- How do curves look like if we initialize 3D filter using random initialization or closed-form extension from a trained 2D filter to a 3D counterpart based on svmmetric consideration?



Perceptual loss



Wasserstein Distance

- What if the 2D model was not trained in the GAN framework?
    - Doesn’t matter. Train a discriminator from scratch to converge, then do transfer learning and fine-tuning.

Conclusion

- We have introduced contracting path convolutional autoencoder (CPCA) for low-dose CT denoising
- Optimized denoising model under WGAN framework
- Proposed a simple yet effective way of transfer learning from a 2D trained model to a 3D counterpart, advoiding 3D training from scratch
- Our work can be extended to higher dimensionality in other tomographic imaging scenarios

Future work

- Inspired by Prof. Quanzheng Li's talk
  - Evaluating denoising model via detection/classification task

Thank You!

Biomedical Imaging Center

Principal Investigator: Ge Wang

Wenxiang Cong	Qingsong Yang
Guang Li	Matthew Getzin
Hongming Shan	Lars Gjesteby
Ruibin Feng	Fenglei Fan
Tao Xu	Qing Lyu

Rensselaer