Generalizability Test of Deep Learning Based Low-dose CT Denoising

Claire Lin[1], Rongping Zeng[2]

[1]Electrical Engineering and Computer Science, University of Michigan

[2]Division of Imaging, Diagnostics and Software Reliability, OSEL, CDRH, FDA

Introduction

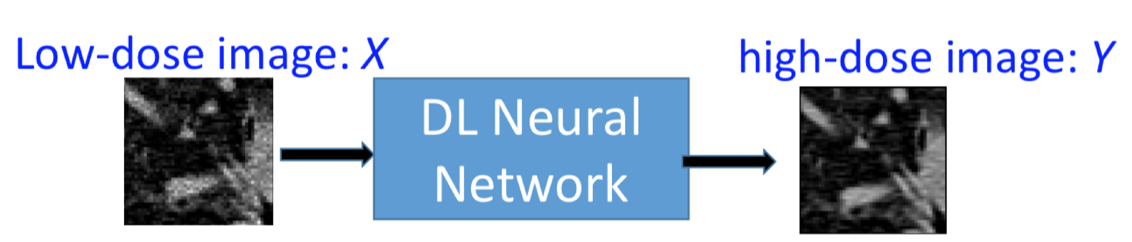
- **Deep learning** (DL) is increasingly explored in medical image reconstruction.
- While having potential to improve image quality and produce images quickly, **generalizaibility** of DL methods is a concern.
- CT data can be obtained with various acquisition **parameters** (e.g., kv, mA, filter, slice thickness, pitch).
- Our goal is to **explore** the generalizability of training dataset and **identify** important CT scan parameters that affect DL image performance.

Regulatory Impact

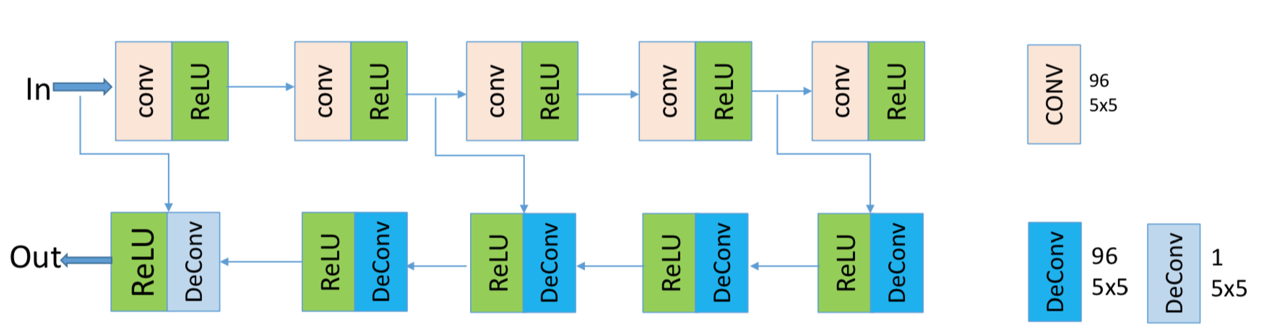
- DL reconstruction / denoising products are submitted to FDA for **premarket approval** (e.g., CT, MRI and PET imaging).
- Quality of DL-processed images directly affects clinical **decision accuracy**.
- We develop an assessment methodology to help the **evaluation** of DL-based imaging products.

Methods

DL denoising



- Residual Encoder-Decoder Convolutional Neural Network (RED-CNN) [1]:
- 5 convolutional and 5 deconvolutional layers with shortcuts



Training and testing data

- AAPM 2016 Low-dose CT Grand Challenge dataset [2]
- Data variability[a]:
  - **Reconstruction kernel**: sharp v.s. smooth
  - **Slice thickness**: 1mm v.s. 3mm
  - **Dose level**: constant (25%) v.s. mixed (10 - 60%)
- Use different body cross section slices
- Train from 7 patients, test from 1 patient

Evaluation metrics

- Mean squared error (MSE): average error
- Modulation transfer function (MTF): image sharpness
- Noise power spectrum (NPS): noise correlation
- Visual smoothess and detectability

[a] Unless otherwise noted: smooth kernel, 3mm slice, 25% dose

Results

1. MSE: training & testing data variability

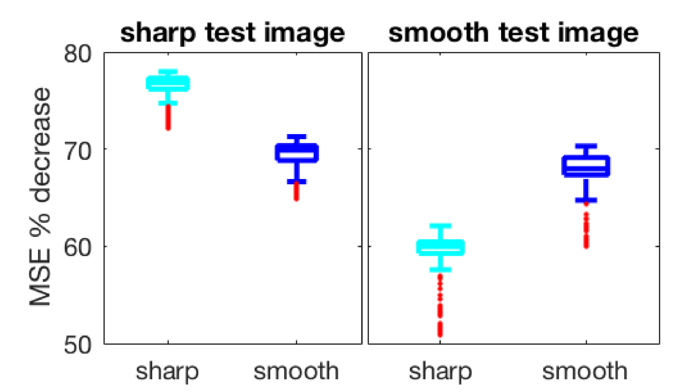


Fig. 1: Kernel is not generalizable.

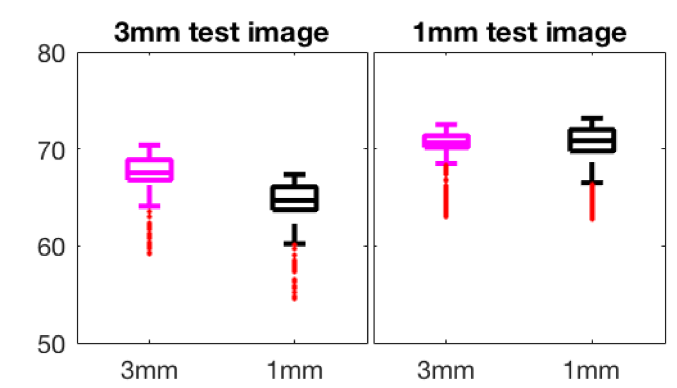


Fig. 2: Slice thickness is well generalized.

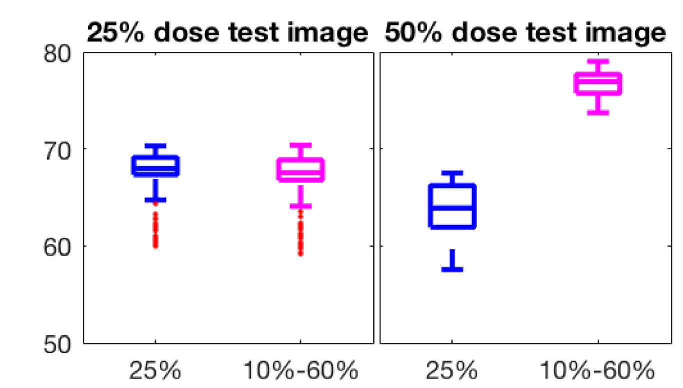


Fig. 3: Mixed-dose is more generalizable.

2. Object and MTF: contrast phantom

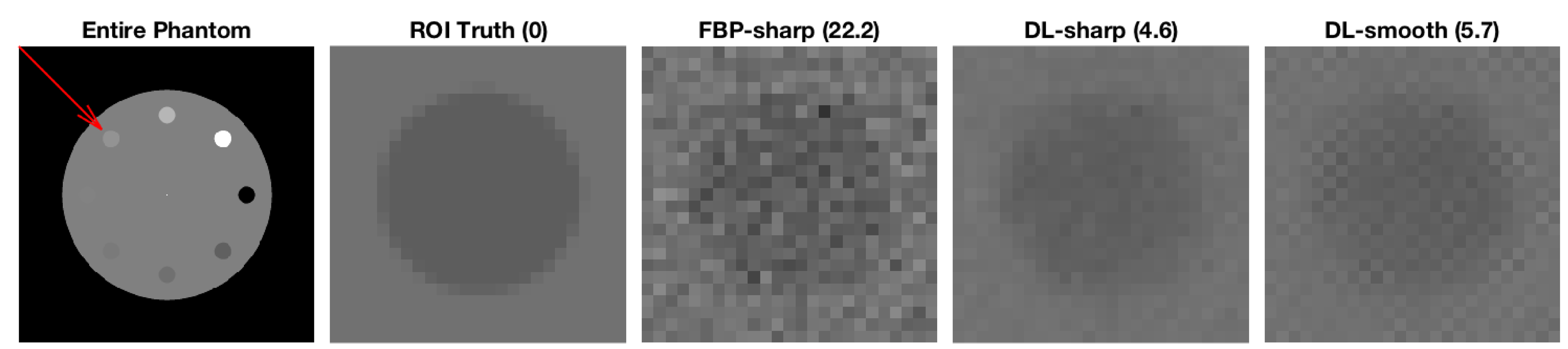


Fig. 4: Contrast phantom, zoomed-in -35 HU object, and its quarter-dose results, with standard deviations of region of interest (ROI).

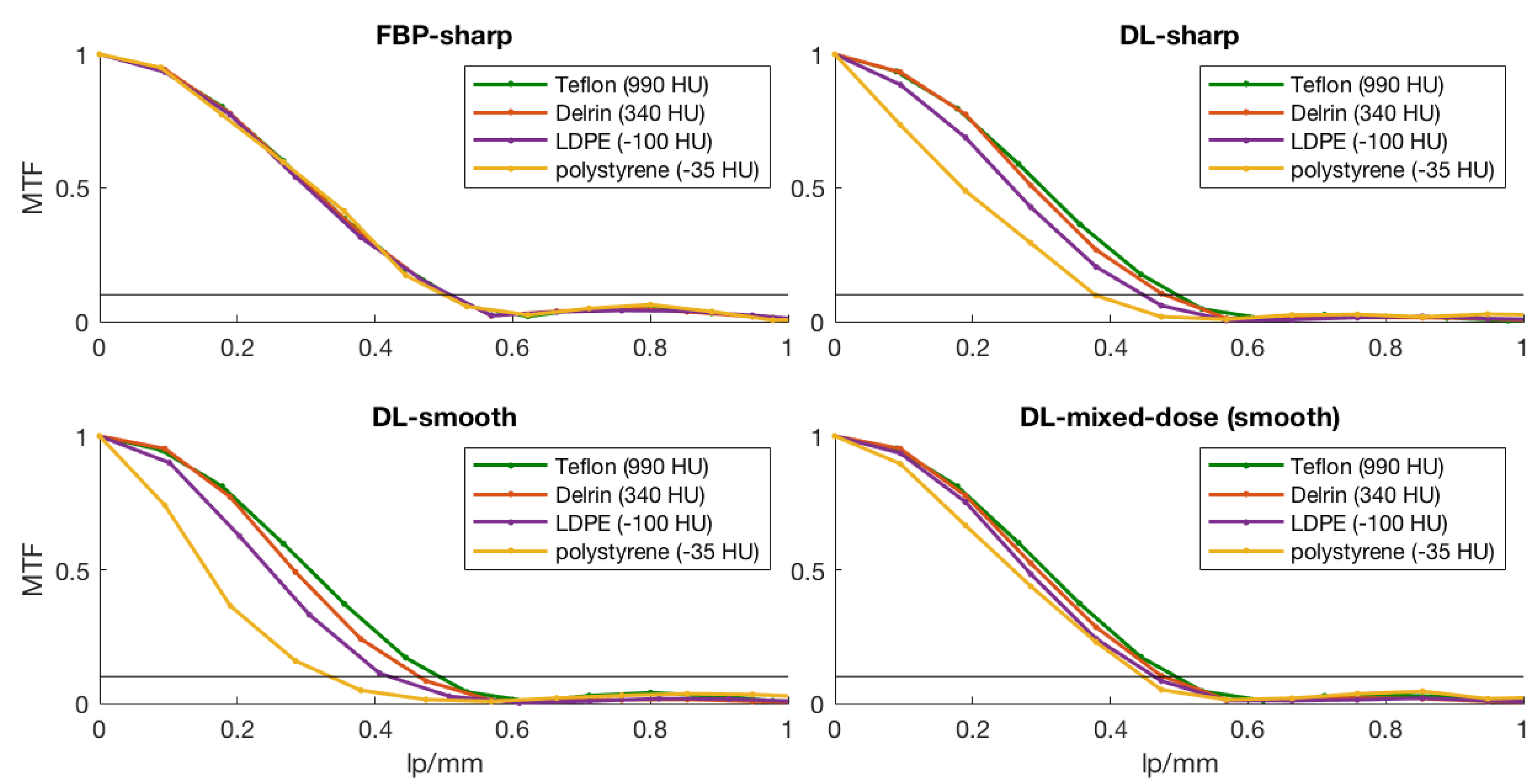


Fig. 5: MTF of noiseless phantom image is contrast-independent for FBP, but decreases with contrast for DL methods.

3. NPS: uniform water phantom

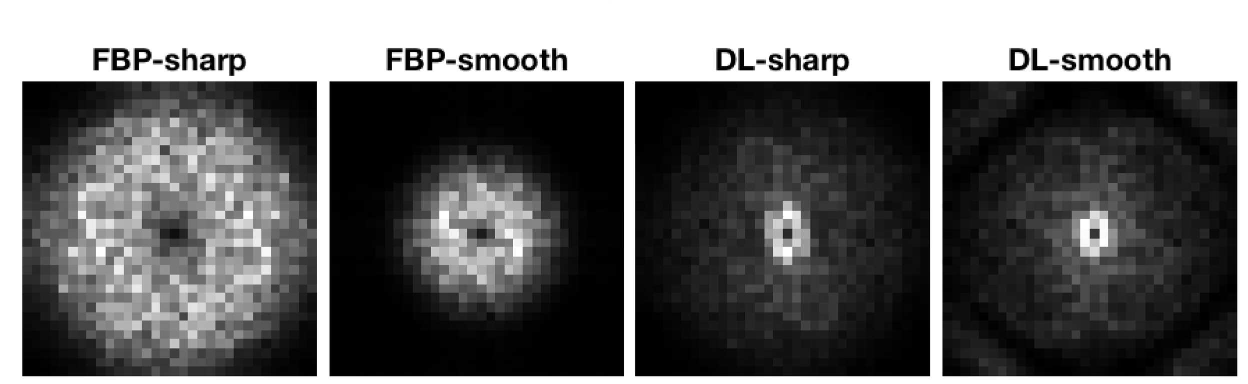


Fig. 6: NPS of half-dose phantom for 4 methods, over 20 noise realizations.

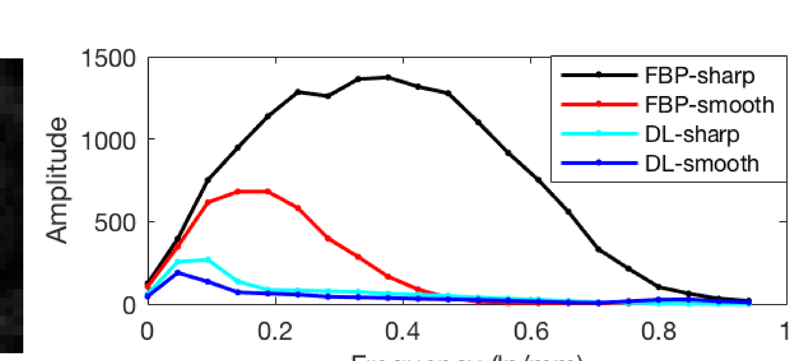


Fig. 7: Radial profiles of NPS.

4. Visual smoothness: clinical results

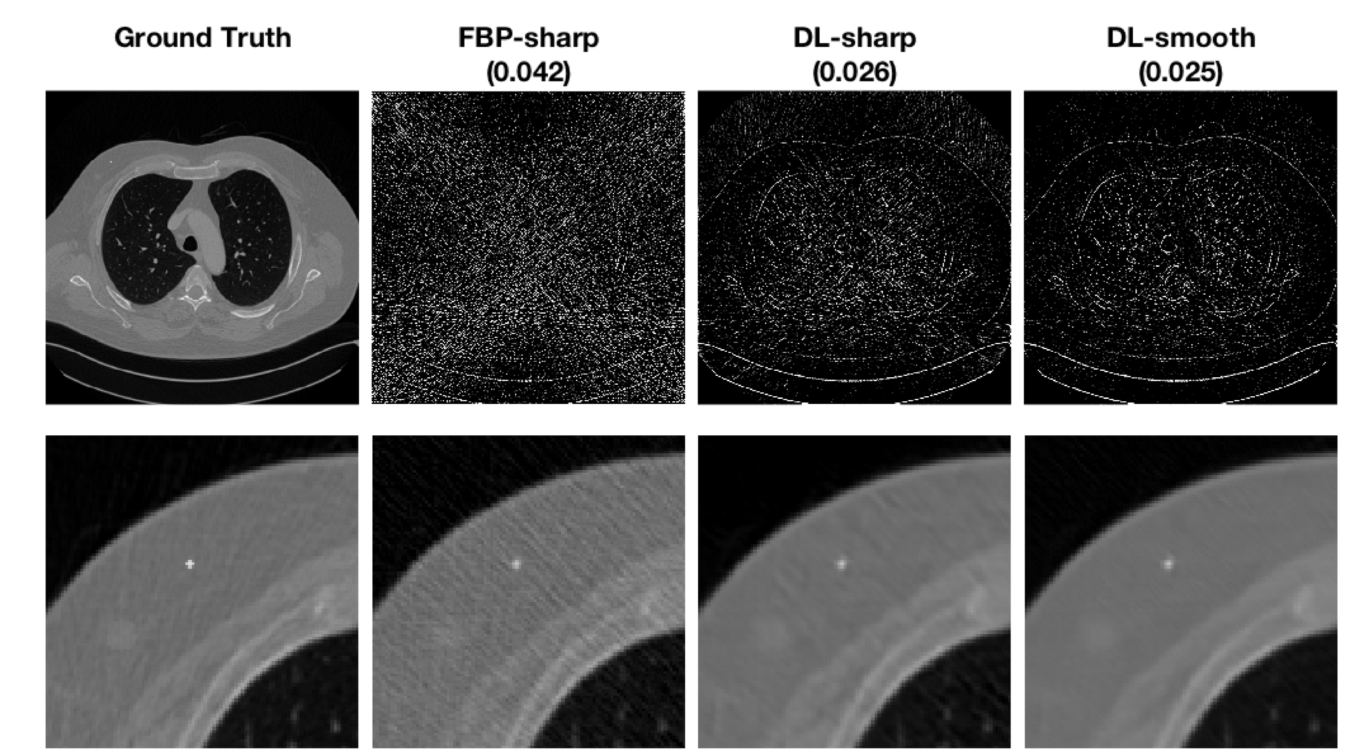


Fig. 8: Ground truth, difference, and denoised clinical image ROIs, with NRMSE.

Conclusion

Reconstruction kernel:
- Kernel choice in training affects test results.
- Network trained with smooth kernel may introduce artifacts when applied to FBP sharp-filtered images.

Slice thickness:
- Slice thickness choice in training has minor impact on test results.

Dose level:
- 25%-dose-trained network has similar performance on low- and higher-dose test images.
- Mixed-dose-trained network generalizes better to higher-dose test images.

Acknowledgment

This project was supported in part by an appointment to the Research Participation Program at CDRH, FDA, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the DOE and the FDA.

References

[1] H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang, "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE transactions on medical imaging*, vol. 36, no. 12, pp. 2524–2535, 2017.

[2] C. H. McCollough, A. C. Bartley, R. E. Carter, B. Chen, T. A. Drees, P. Edwards, D. R. Holmes III, A. E. Huang, F. Khan, S. Leng, *et al.*, "Low-dose CT for the detection and classification of metastatic liver lesions: results of the 2016 low dose CT grand challenge," *Medical physics*, vol. 44, no. 10, pp. e339–e352, 2017.