Image quality evaluation in deep-learning-based CT noise reduction using virtual imaging trial methods: Lesion characterization

Zhongxing Zhou, Hao Gong, Scott Hsieh, Shuai Leng, Joel G. Fletcher, Cynthia H. McCollough, Lifeng Yu

Department of Radiology, Mayo Clinic, Rochester MN USA 55905-0001

INTRODUCTION

Deep learning-based image reconstruction and noise reduction (DLR) techniques are being increasingly adopted in clinical CT to improve image quality at reduced radiation doses. While prior studies have demonstrated the benefits of DLR for lesion detection tasks [1-4], its performance in more challenging lesion characterization tasks—specifically differentiating malignant from benign lesions based on morphological features such as spiculated boundaries—remains largely unexplored across varying lesion contrasts, dose levels, and denoising strengths. Accurate lesion characterization is clinically critical for diagnostic confidence and patient management. This study applies a patient-data-based virtual imaging trial (VIT) framework to objectively evaluate DLR performance for lesion characterization.

DLR can degrade lesion characterization performance at low contrasts and strong denoising strengths, despite improvements in lesion detection tasks.

METHODS

VIT-Based Data Generation

A patient-data-based virtual imaging trial (VIT) framework (**Figure 1**) was previously developed [5]. This framework was used to generate realistic ensembles for lesion characterization. Cylindrical lesions (15 mm diameter, 5 mm height) with or without spiculated boundaries were digitally inserted into real patient liver CT projection data to simulate benign (round/smooth) and malignant (spiculated) lesions. Ensembles were created at four lesion contrasts (−20, −30, −50, and −100 HU) and three dose levels (12.5%, 25%, and 50% of routine dose) by adding realistic quantum noise to the projection data. Images were reconstructed using filtered backprojection (FBP), iterative reconstruction (IR), and an in-house developed image-domain DLR algorithm at weak, medium, and strong denoising strengths[6].

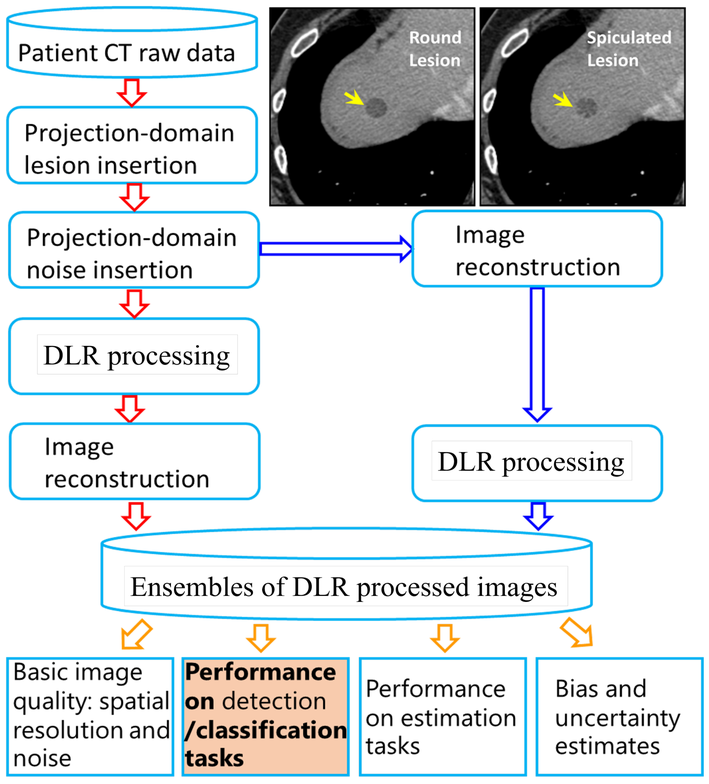

Figure 1. A patient-data-based virtual imaging trial (VIT) framework for evaluating deep-learning reconstruction (DLR). This work is focused on lesion classification (characterization) tasks. Round and spiculated lesions of varying contrast levels and sizes were inserted into projection data for the lesion characterization task.

Channelized Hotelling Observer (CHO) Evaluation

Lesion characterization performance was assessed using a Channelized Hotelling Observer (CHO) [7-9]. For each condition, 3,000 image pairs (5 central slices × 600 noise realizations) were analyzed. The CHO template was computed from the intraclass scatter matrix and the difference between the channelized ensemble averages of the spiculated and round lesion classes. This model-observer approach was adapted from standard lesion detection tasks to evaluate boundary morphology discrimination (spiculated vs. round). We refer to the resulting index of class separability as characterizability (c') to differentiate it from the conventional detectability (d') used for lesion detection tasks.

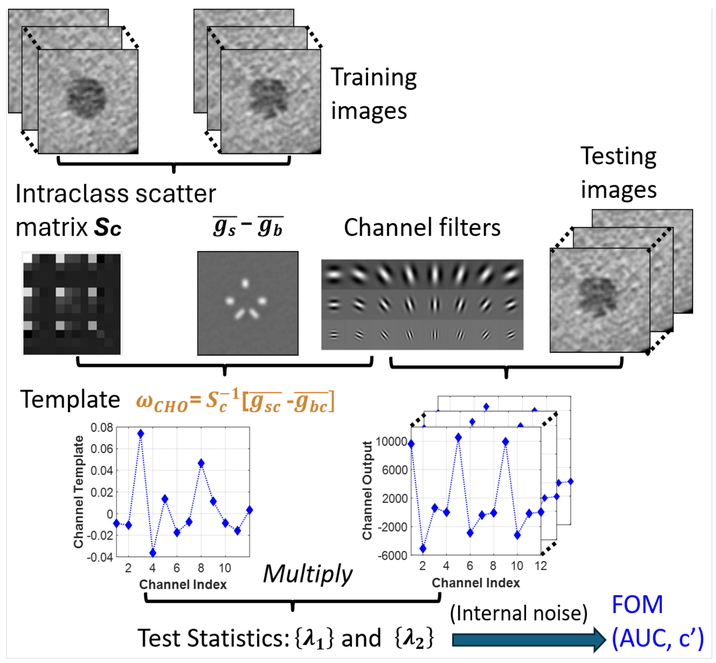

Figure 2. The CHO-based detectability estimation framework for lesion characterization tasks, a concept we refer to as characterizability (c').

RESULTS

For lesion characterization, c' decreased with greater denoising strength and at reduced contrast (-20 to -50 HU). DLR-Strong reduced c' relative to both FBP and IR at lower contrasts (-20 to -50 HU) across all doses and at 12.5% dose for -100 HU contrast, but improved c' at -100 HU with 25–50% doses. In contrast, for lesion detection, DLR consistently improves d' over FBP, with gains rising as denoising strength increased.

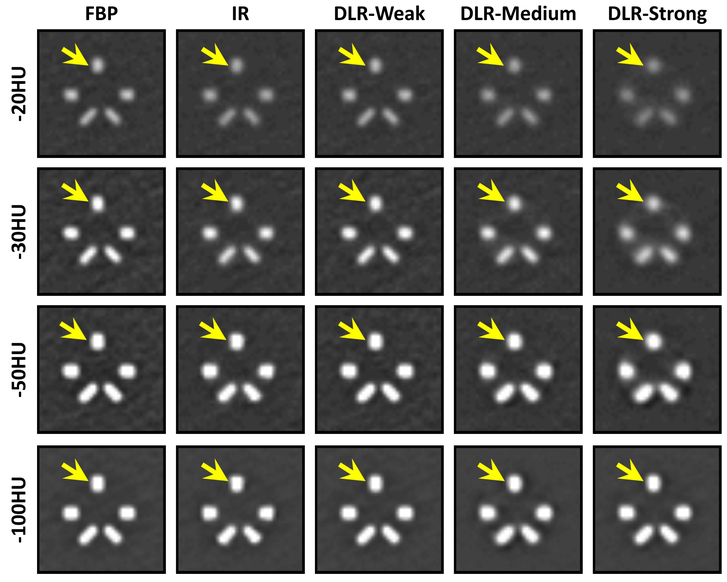

Figure 3. Ensemble-averaged boundary signals (yellow arrows) from FBP, IR, and DLR-Weak/Medium/Strong reconstructions across four contrast levels (−20 HU, −30 HU, −50 HU, and −100 HU) at 50% of routine dose. Display settings (window level, window width) are (30, 110) HU for the −100 HU boundary signals and (10, 35) HU for the others.

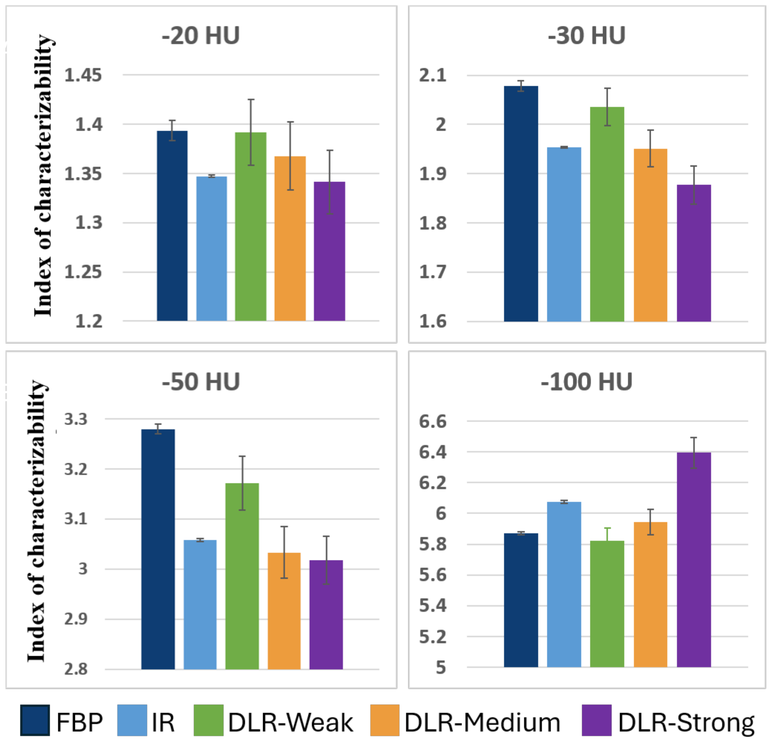

Figure 4. CHO characterizability index (c') across different reconstruction algorithms (FBP, IR, DLR-Weak, DLR-Medium, and DLR-Strong) at different lesion contrasts (-20, -30, -50 and -100 HU) using 50% of routine dose. Error bars indicate standard deviation.

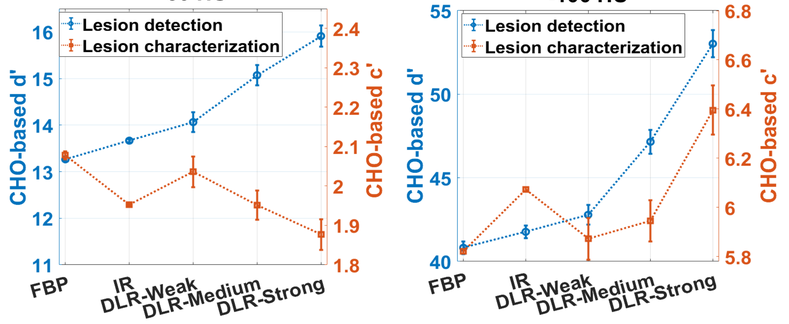

Figure 5. Comparison of CHO characterizability (c') and detectability (d') between lesion characterization and lesion detection tasks at −30 HU and −100 HU contrast, 50% dose, across five reconstruction methods (FBP, IR, DLR-Weak, DLR-Medium, DLR-Strong)..

CONCLUSIONS

Patient-data-based virtual imaging trials with CHO evaluation demonstrated that DLR can degrade lesion characterization performance at low contrasts and very low doses, despite improvements in lesion detection, warranting caution when applying strong DLR under such challenging conditions.

ACKNOWLEDGEMENTS

This work was supported in part by NIH award (R01 EB036541). Dr. McCollough receives industry grant support from Siemens unrelated to this work. No other potential conflicts of interest were declared.

REFERENCES

1. Lyu, P., et al., Is it possible to use low-dose deep learning reconstruction for the detection of liver metastases on CT routinely? European Radiology, 2023. 33(3): p. 1629-1640.
2. Lee, D.H., et al., Image Quality and Diagnostic Performance of Low-Dose Liver CT with Deep Learning Reconstruction versus Standard-Dose CT. Radiology-Artificial Intelligence, 2024. 6(2).
3. Singh, R., et al., Image Quality and Lesion Detection on Deep Learning Reconstruction and Iterative Reconstruction of Submillisievert Chest and Abdominal CT. American Journal of Roentgenology, 2020. 214(3): p. 566-573.
4. Jiang, B.B., et al., Deep Learning Reconstruction Shows Better Lung Nodule Detection for Ultra-Low-Dose Chest CT. Radiology, 2022. 303(1): p. 202-212.
5. Zhou, Z.X., et al. Image quality evaluation in deep-learning-based CT noise reduction using virtual imaging trial methods: Contrast-dependent spatial resolution. Medical Physics, 2024. 51(8): p. 5399-5413.
6. Zhou, Z.X., et al. Multislice input for 2D and 3D residual convolutional neural network noise reduction in CT. Journal of Medical Imaging, 2023. 10(1).
7. Yu, L.F., et al. Correlation between a 2D channelized Hotelling observer and human observers in a low-contrast detection task with multislice reading in CT. Medical Physics, 2017. 44(8): p. 3990-3999.
8. Fan, M.D., et al. Accurate and efficient measurement of channelized Hotelling observer-based low-contrast detectability on the ACR CT accreditation phantom. Medical Physics, 2023. 50(2): p. 737-749.
9. Fan, M.D., et al. Channelized hotelling observer-based low-contrast detectability on the ACR CT accreditation phantom: Part II. Repeatability study. Medical Physics, 2024. 51(3): p. 1714-1725.

Contact information: Zhou.Zhongxing@mayo.edu