SF 18c - Clinical benefits and challenges of an expanding role of AI in radiology
March 7, 09:30 - 11:00 CET

AI in CT Image Formation: Maximizing Benefits, Minimizing Pitfalls

Marc Kachelrieß

German Cancer Research Center (DKFZ)
Heidelberg, Germany
www.dkfz.de/ct

Potential Pitfalls and their Clinical Risk

Pitfall	Likelihood	Clinical Risk	Comments
Lesions are removed or introduced: hallucinations	unkown	high	• May completely impair the diagnosis. • The user cannot detect the error (looks like a real lesion). • May easily happen with image-based AI. Very unlikely in projection-based AI.
Apparent spatial resolution is increased	high (for CT systems that provide super resolution AI)	medium	• Fabrication of information that was never measured • Illegitimate resolution recovery (not an MTF deconvolution) • May be useful when it recovers suppressed information. • Perceptual quality may matter more than strict quantitative fidelity: Human detectability may improve.
Artifacts (streaks, blurring, …) are introduced	unkown	low	• The user can easily detect the error (appears unnatural). • May occur with projection-based AI. Unlikely in image-based AI.

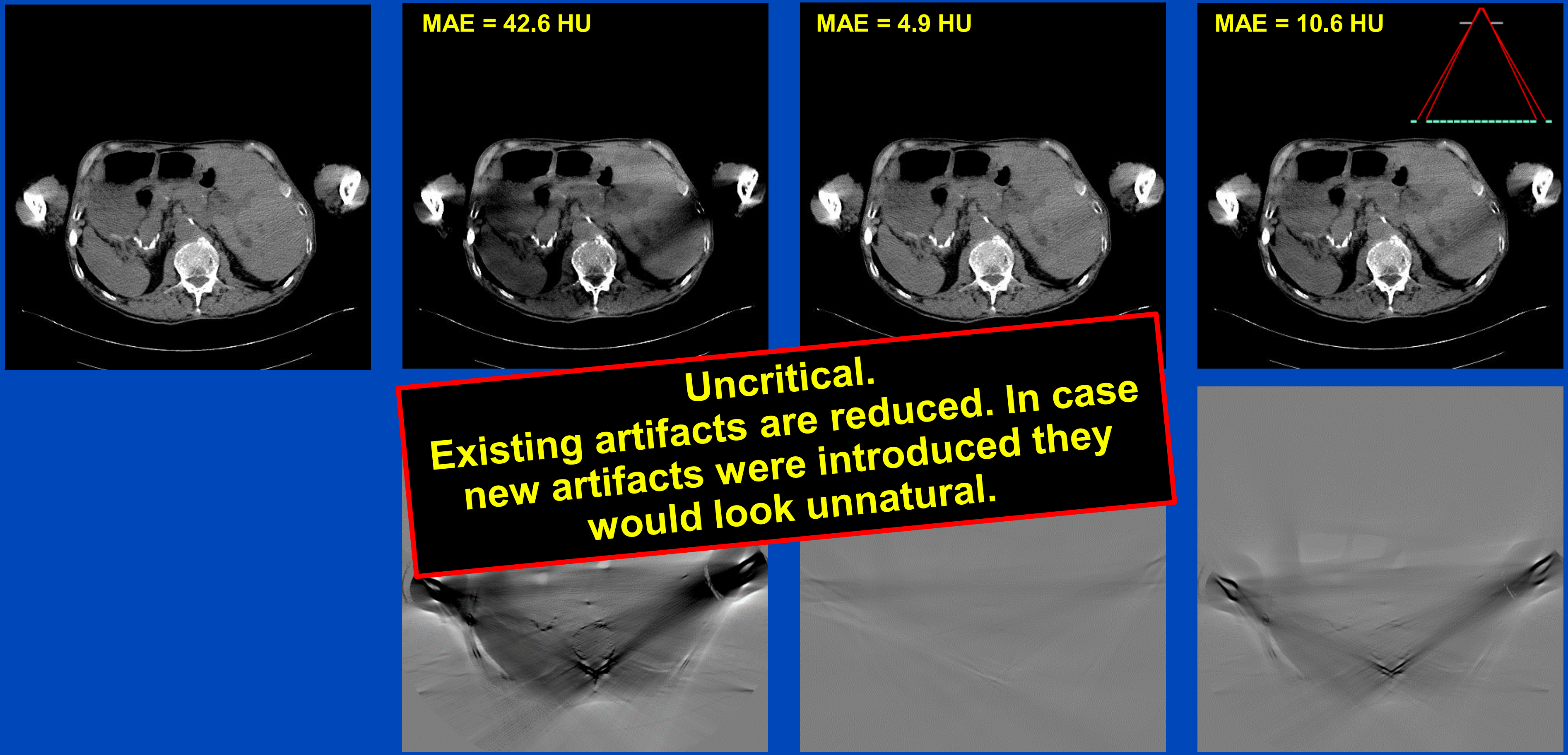
dkfz.

Examples for

AI IN CT IMAGE FORMATION

Scatter and Cross-Scatter: DSE and Cross-DSE

Ground Truth | Uncorrected | xDSE (2D, xSSE) | Measurement-based



Uncritical.
Existing artifacts are reduced. In case new artifacts were introduced they would look unnatural.

xDSE (2D, xSSE) maps

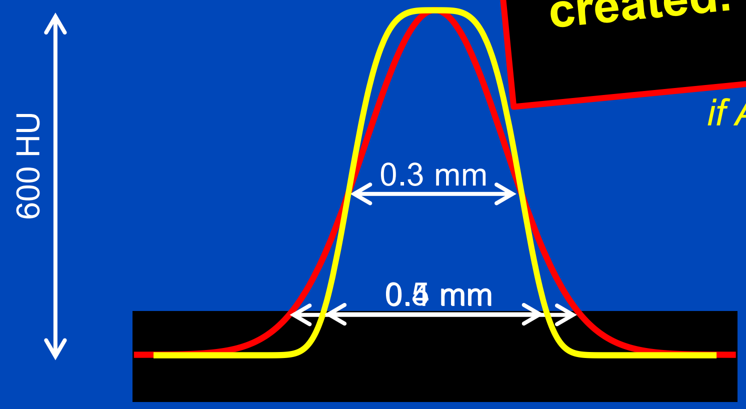
primary + forward scatter + cross-scatter + cross-scatter approximation → cross-scatter

Images C = 40 HU, W = 300 HU, difference images C = 0 HU, W = 300 HU

Canon PIQE

- Precise IQ Engine (PIQE).
- Trained on data from Canon’s Precision high spatial resolution CT
- Converts images from Canon’s standard spatial resolution scanners (e.g. Aquilion ONE / PRISM edition) to look like high spatial resolution images.

OK.
Apparent spatial resolution is increased, but no new information is created. Human detectability may be improved.



calcium score,
if Agatston scoring is used.

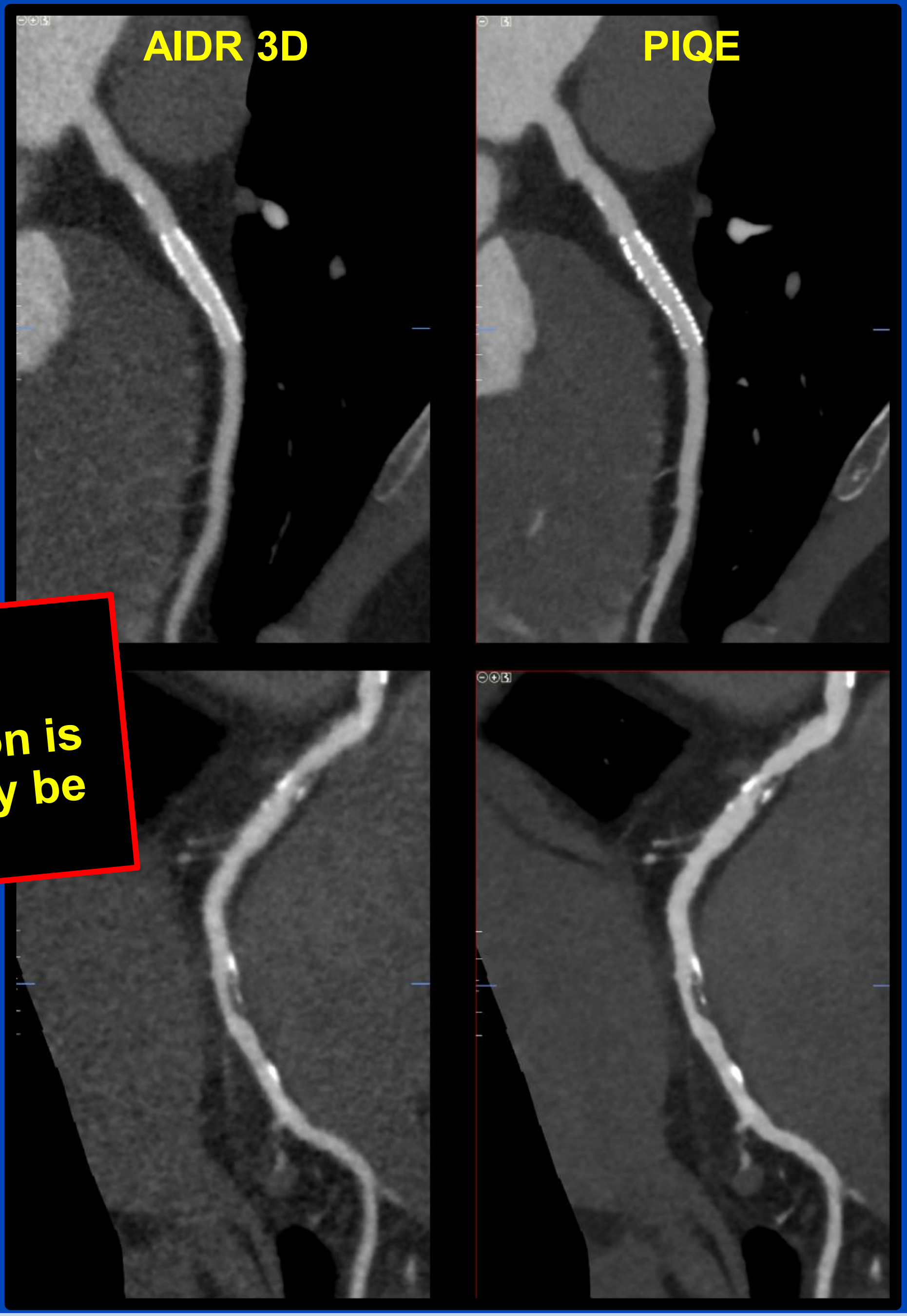


Image courtesy of Canon Medical Systems

Fake Contrast Enhancement

[1] G. Santini, L. M. Zumbo, N. Martini, G. Valvano, A. Leo, A. Ripoli, F. Avogliero, D. Chiappino, D. D. Latta, "Synthetic contrast enhancement in cardiac CT with deep learning," arXiv 1807:01779, 2018.

[2] J. Liu, Y. Tian, A. M. Ağıldere, K. M. Haberal, M. Coşkun, C. Duzgol, and O. Akin, "DyeFreeNet: Deep virtual contrast CT synthesis," Lecture Notes in Computer Science. Springer International Publishing, pp. 80–89, 2020.

[3] A. Chandrashekar, A. Handa, N. Shivakumar, P. Lapolla, V. Grau, R. Lee, "A deep learning approach to generate contrast-enhanced computerised tomography Angiography without the use of intravenous contrast agents," arXiv 2003.01223, 2020.

[4] J. W. Choi, Y. J. Cho, J. Y. Ha, S. B. Lee, S. Lee, Y. H. Choi, J.-E. Cheon, and W. S. Kim, "Generating synthetic contrast enhancement from non-contrast chest computed tomograph[redacted]erative adversarial network," Scientific Reports, vol. 11, no. 1, [redacted]

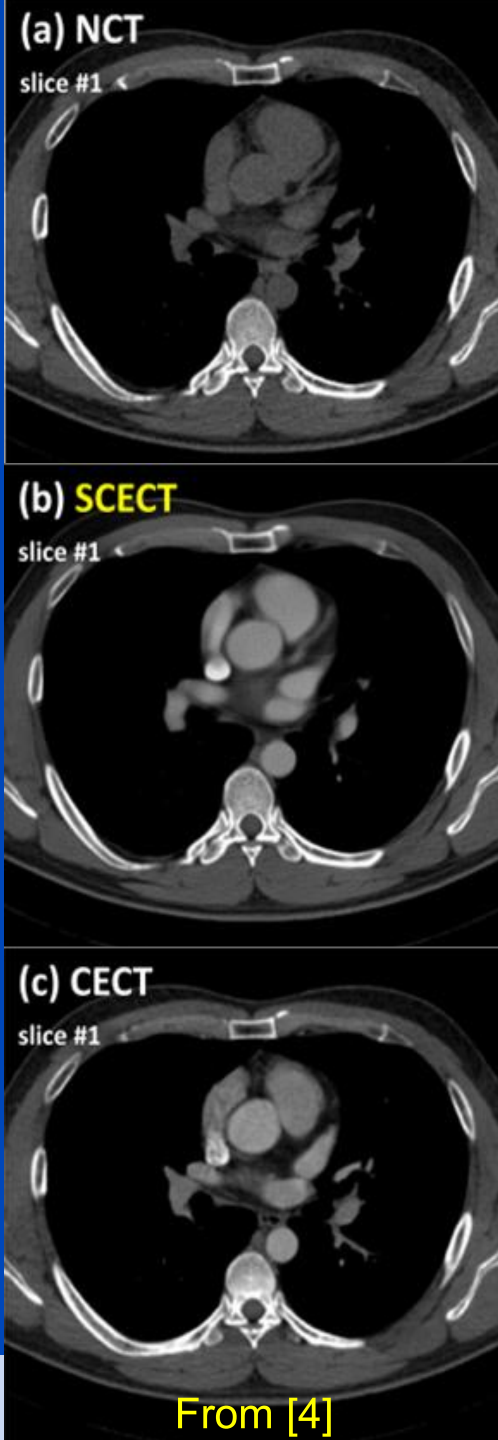
[5] S. W. Kim, J. H. Kim, S. Kwak, M. Se[redacted]eon, "The feasibility of deep learning-based sy[redacted]emergency department patients with acute abd[redacted]

[6] J. Chun, J. S. Chang, C. Oh, I. Park, M[redacted]Moon, S. Y. Chung, Y. J. Suh, and J. S. Kim, "Synthetic contra[redacted]tomography generation using a deep convolutional neural network for cardiac substructure delineation in breast cancer radiation therapy: a feasibility study," Radiation Oncology, vol. 17, no. 1, 2022.

[7] Y. Gao, H. Xie, C. Chang, J. Peng, S. Pan, R. L. J. Qiu, T. Wang, B. Ghavidel, J. Roper, J. Zhou, and X. Yang, "CT-based synthetic iodine map generation using conditional denoising diffusion probabilistic model," Medical Physics, vol. 51, no. 9, pp. 6246–6258, 2024.

[8] S. Han, J.-M. Kim, J. Park, S. W. Kim, S. Park, J. Cho, S.-J. Park, H.-J. Chung, S.-M. Ham, S. J. Park, and J. H. Kim, "Clinical feasibility of deep learning based synthetic contrast-enhanced abdominal CT in patients undergoing non-enhanced CT scans," Scientific Reports, vol. 14, no. 1, 2024.

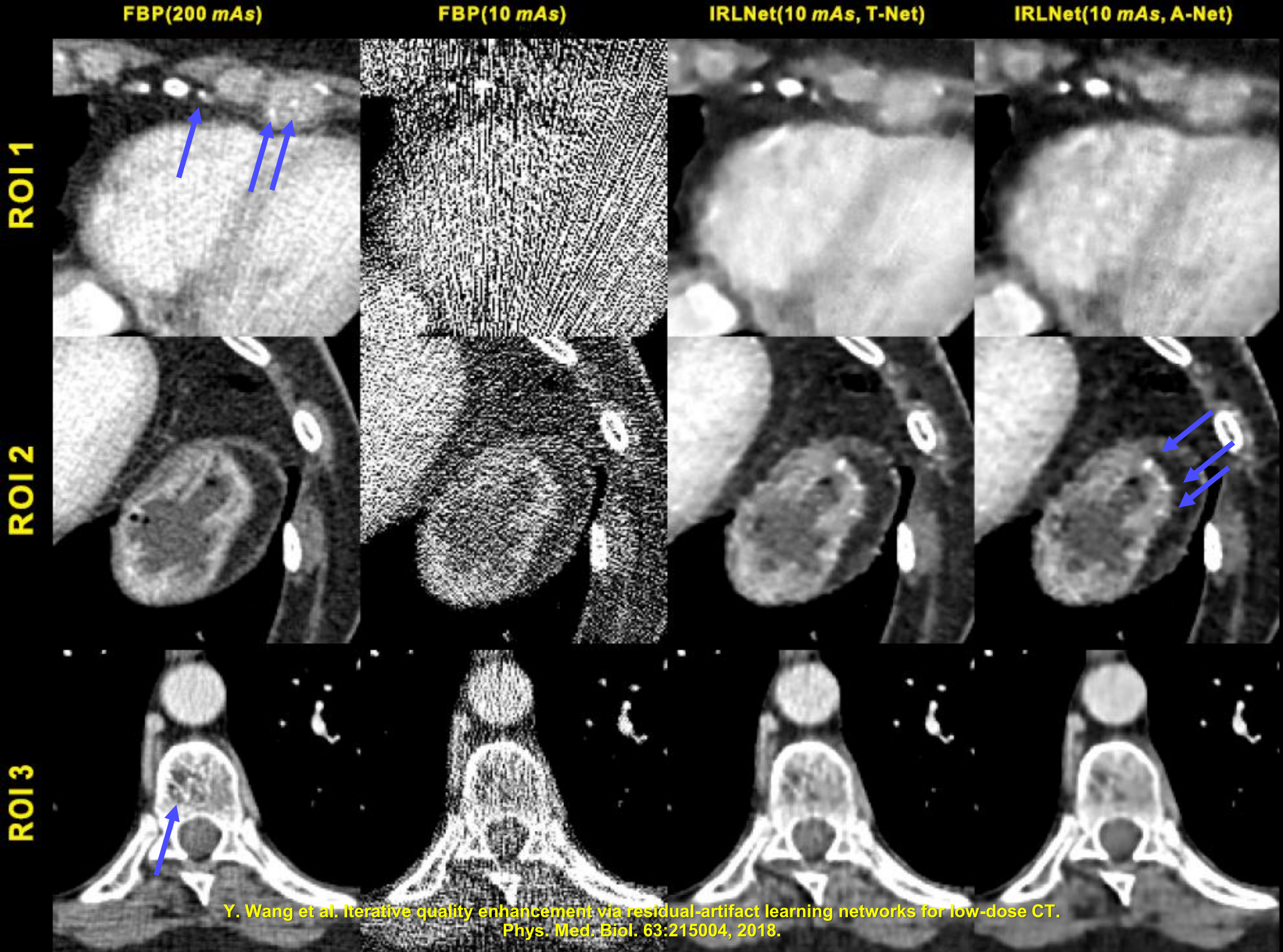
This is a no-go.
No functional information about the vessels is measured.
All information shown is fake.



From [4]

Minimize pitfalls: User action, when in doubt

COMPARE WITH NON-AI IMAGES



Y. Wang et al. Iterative quality enhancement via residual-artifact learning networks for low-dose CT. Phys. Med. Biol. 63:215004, 2018.

Minimize pitfalls: Only show images that are consistent with the measurement

ENFORCE RAWDATA AGREEMENT

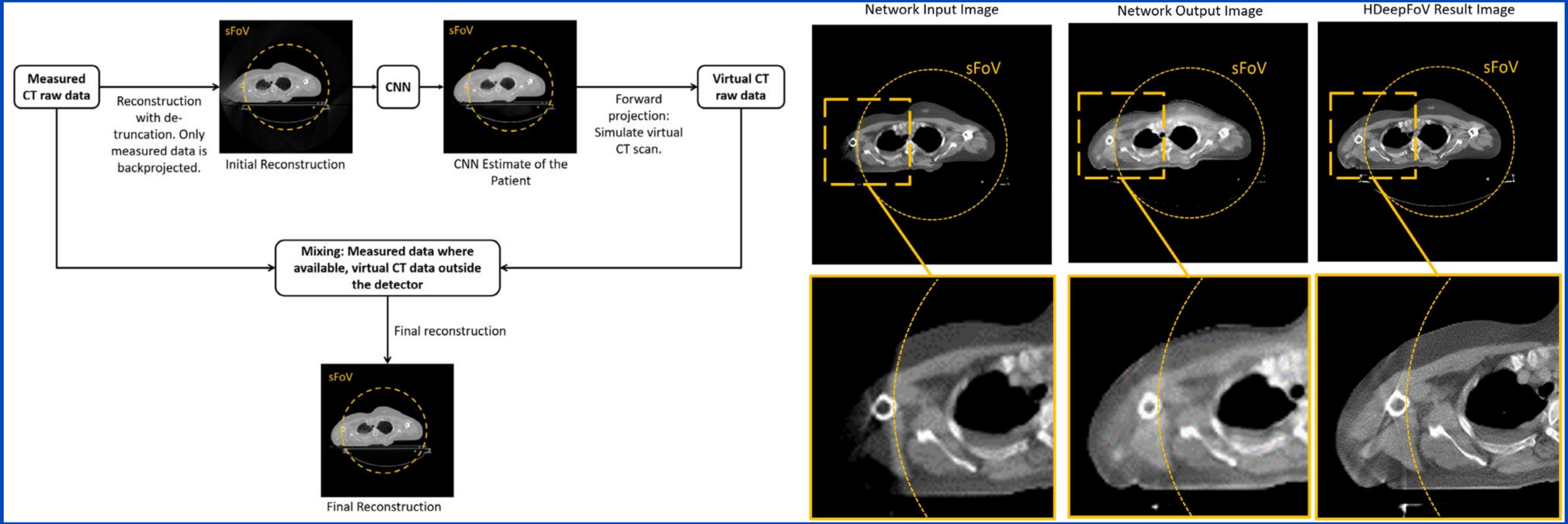
Evaluation of novel AI-based extended field-of-view CT reconstructions

Gabriel Paiva Fonseca[a)*]
Department of Radiation Oncology (MAASTRO), GROW School for Oncology and Developmental Biology, Maastricht University Medical Centre+, Maastricht 6229 ET, The Netherlands

Matthias Baer-Beck[*] Eric Fournie and Christian Hofmann
Siemens Healthcare GmbH, Forchheim, Germany

Ilaria Rinaldi, Michel C Ollers, Wouter J.C. van Elmpt and Frank Verhaegen
Department of Radiation Oncology (MAASTRO), GROW School for Oncology and Developmental Biology, Maastricht University Medical Centre+, Maastricht 6229 ET, The Netherlands

(Received 28 February 2021; revised 27 April 2021; accepted for publication 30 April 2021; published 31 May 2021)



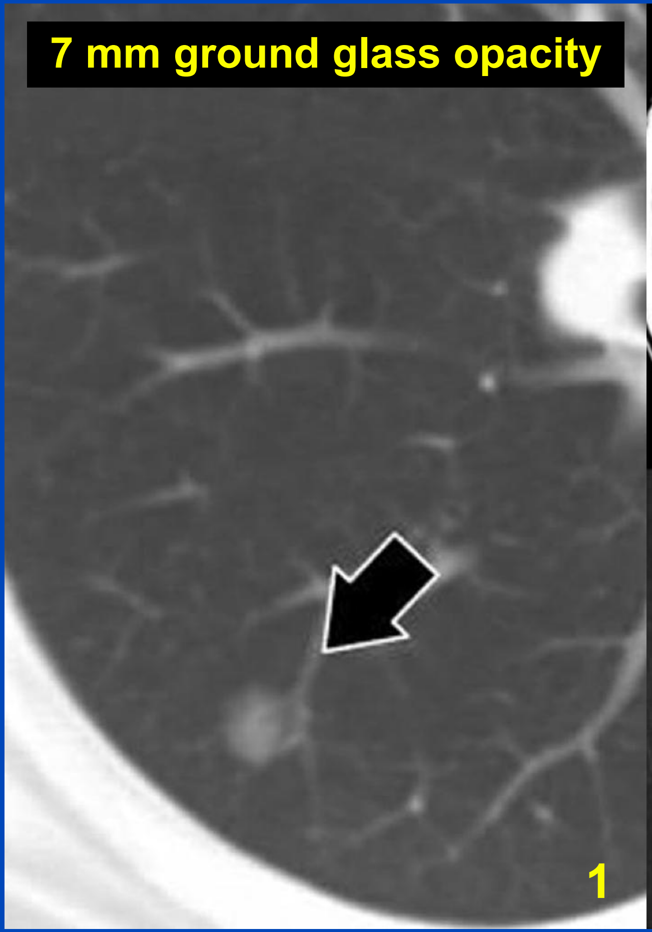
Fonseca, Gabriel Paiva, et al. "Evaluation of novel AI-based extended field-of-view CT reconstructions." *Medical Physics* (2021).

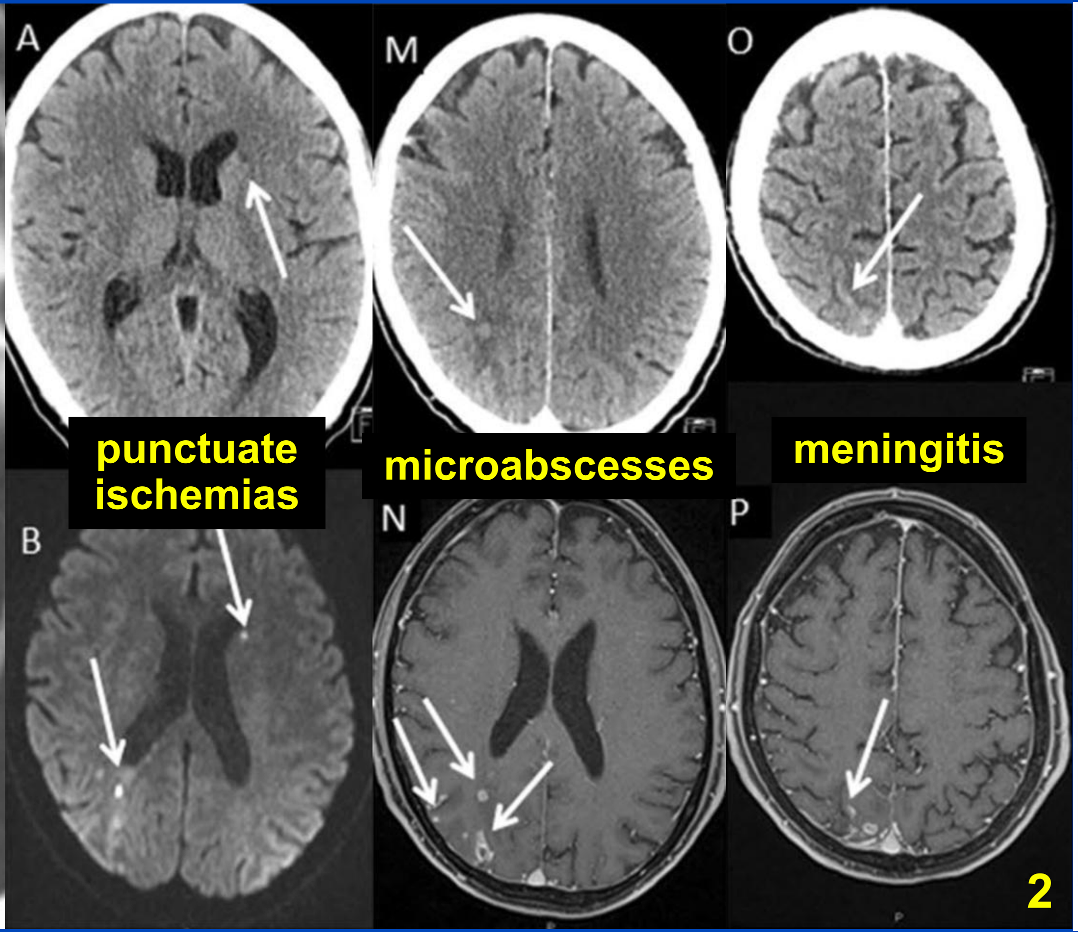
Minimize pitfall: Let small structures be just as important as large structures

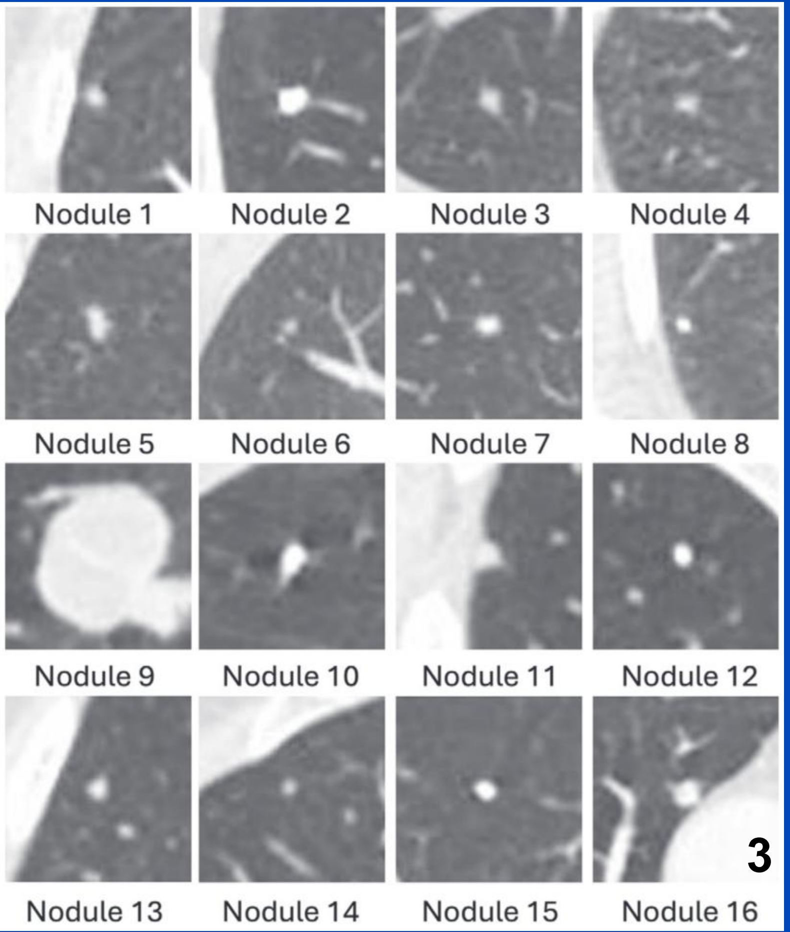
A NEW METRIC FOR SUBTLE DETAILS

Motivation

In medical imaging (CT, MRI) pathological features are often present as small, potentially low-contrast structures



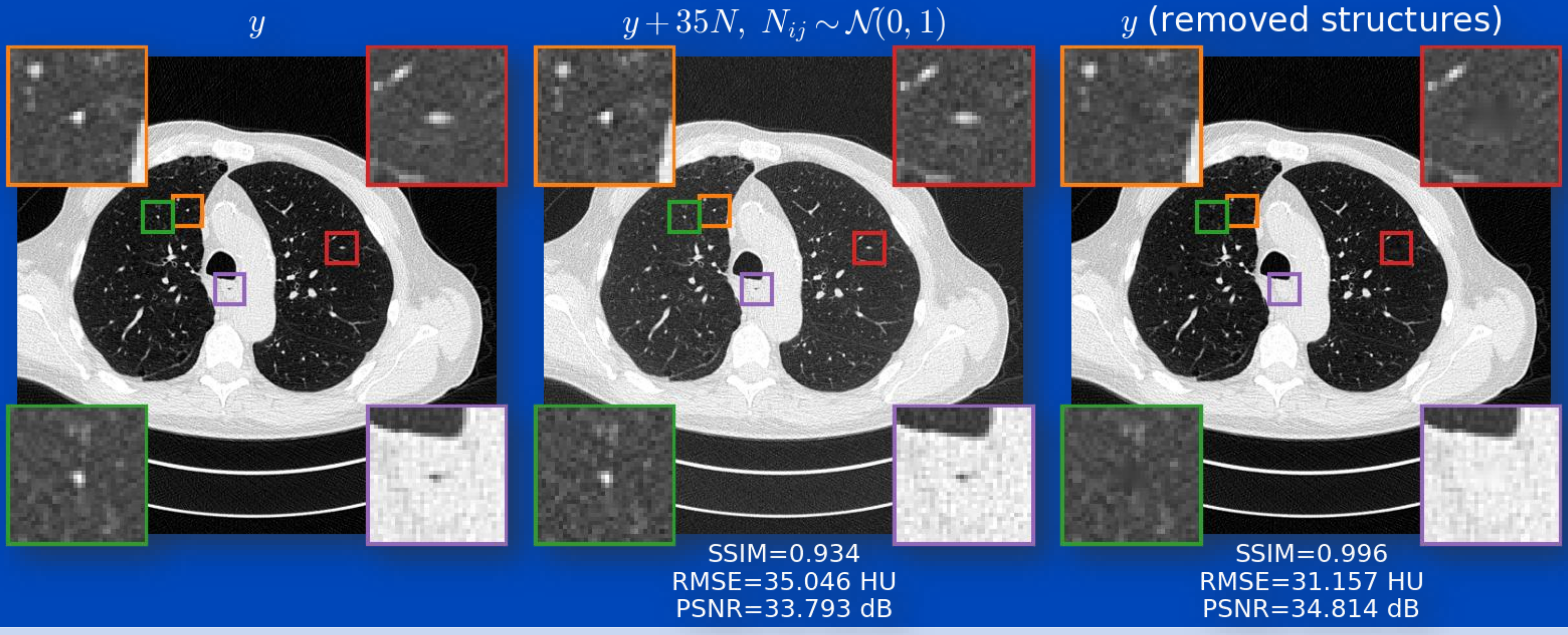




[1]H. K. Kim *et al.*, "Management of Multiple Pure Ground-Glass Opacity Lesions in Patients with Bronchioloalveolar Carcinoma," *Journal of Thoracic Oncology*, vol. 5, no. 2, 2010.
[2]P. Vitali *et al.*, "MRI versus CT in the detection of brain lesions in patients with infective endocarditis before or after cardiac surgery," *Neuroradiology*, vol. 64, no. 5, 2022.
[3]G. J. DiGirolamo *et al.*, "Non-conscious Detection of 'Missed' Lung Nodules by Radiologists: Expanding the Boundaries of Successful Processing during the Visual Assessment of Chest CT Scans," *Radiology*, vol. 314, no. 2, 2025.

Attention: Each Pixel May be Significant!

- MAE, PSNR, RMSE and SSIM* are often used to quantify image quality, e.g. in loss functions or to rank algorithms.
- Alteration of a few pixels may mislead diagnosis.



*SSIM also accounts in parts for the human visual system by using luminance, contrast and structure to estimate perceptual quality.

dkfz.

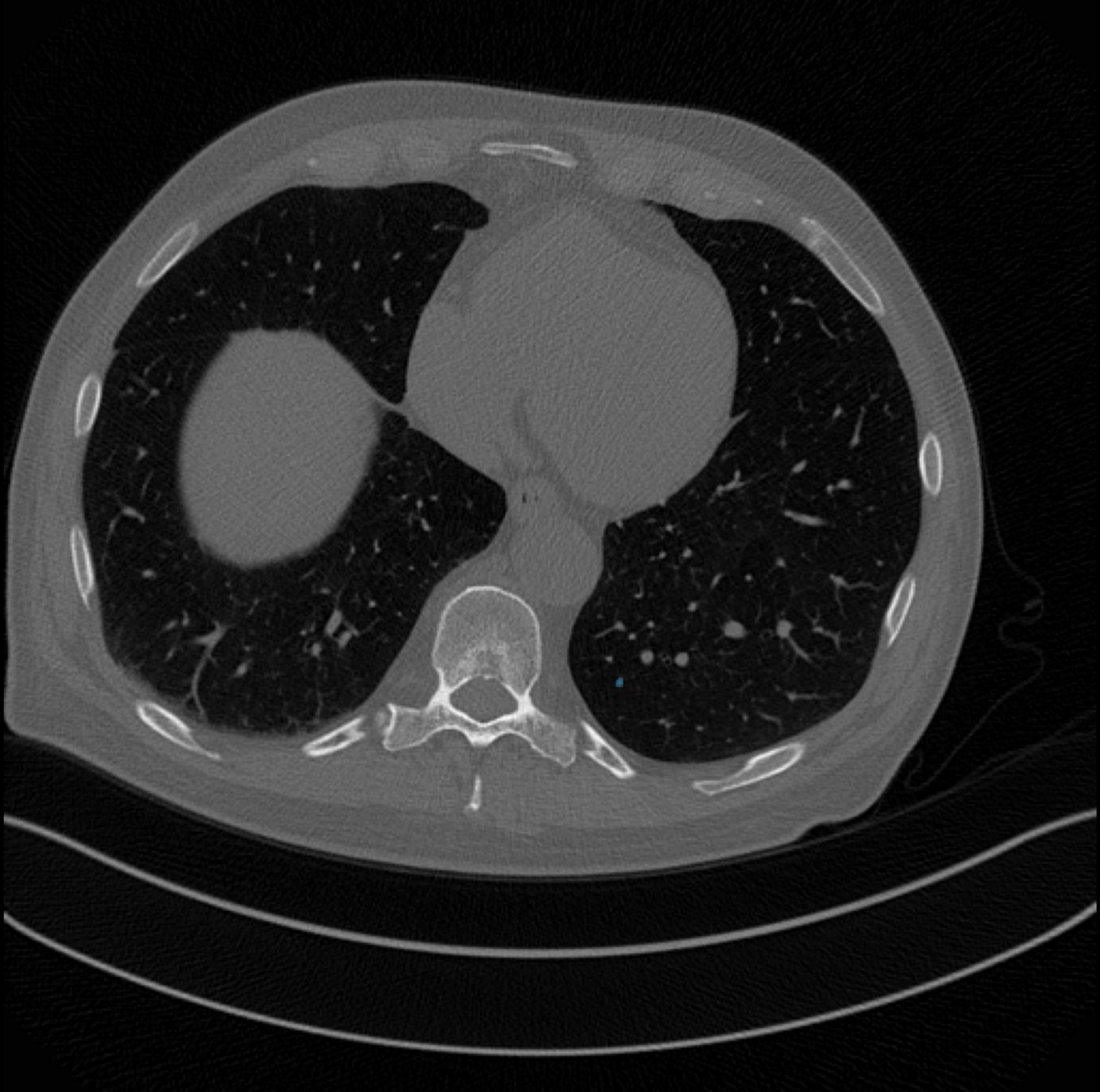
Methods

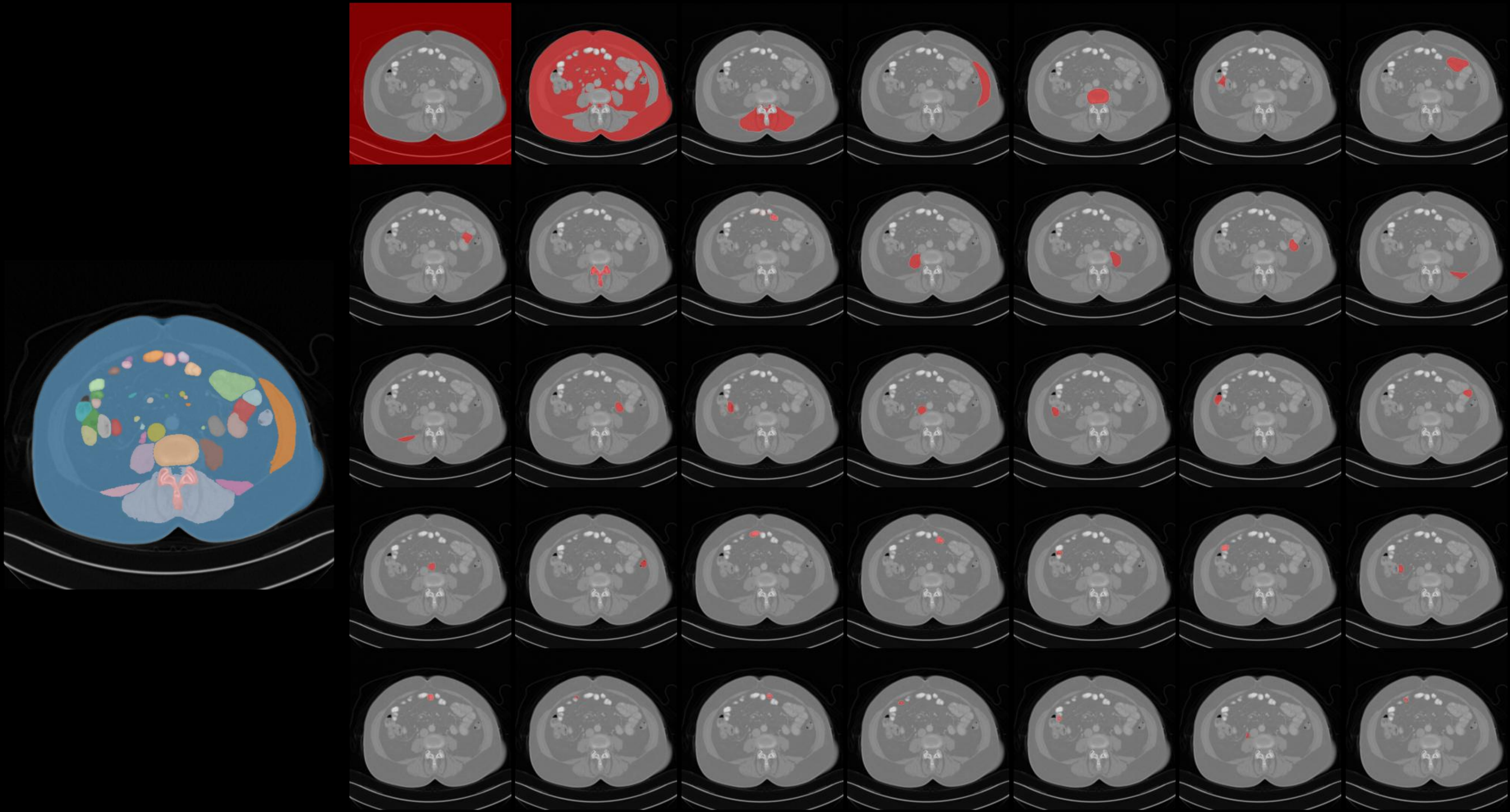
Segment Anything Model (SAM)[1]

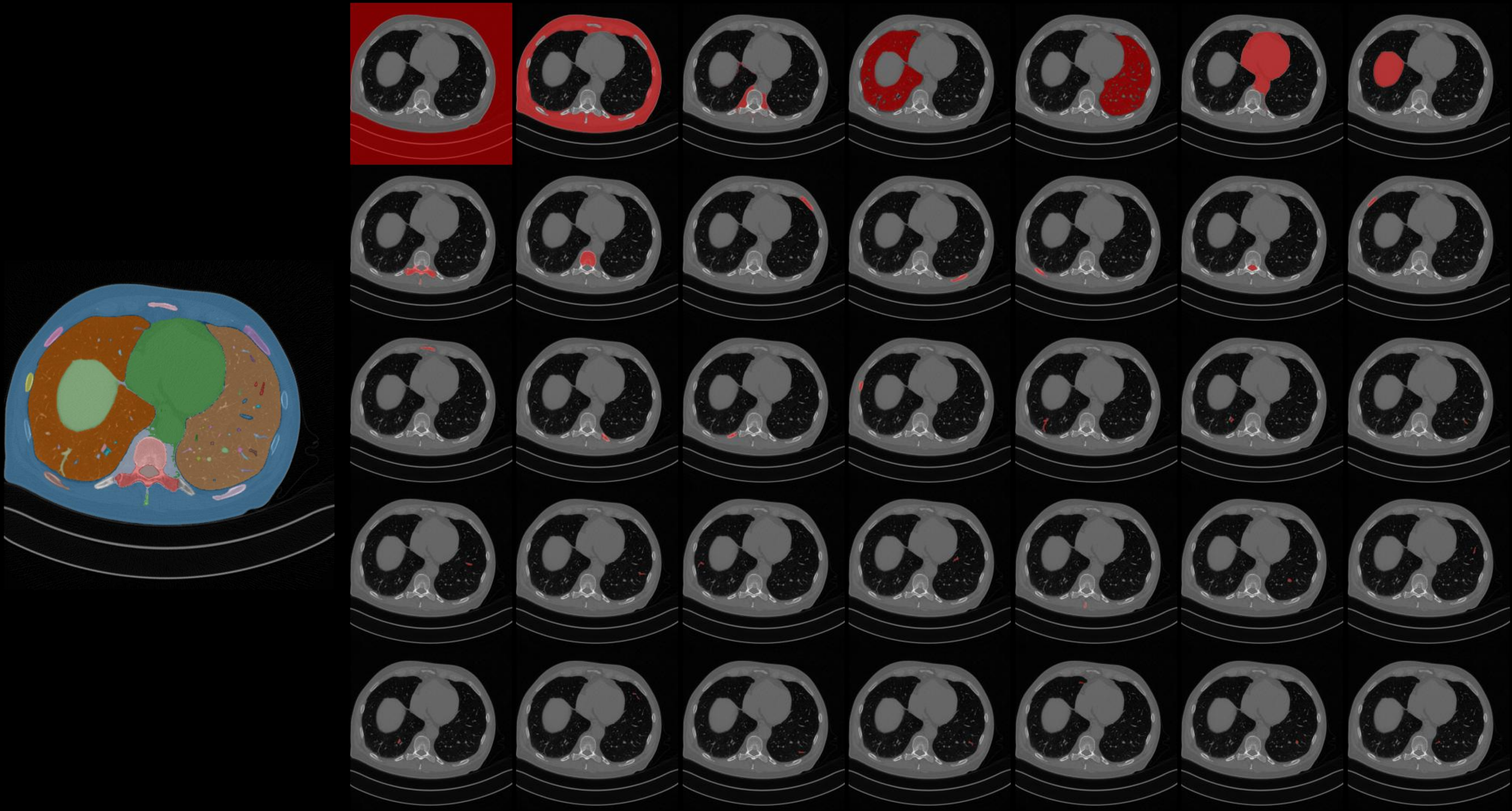
- Foundation model to segment arbitrary structures in natural images
- Based on three components
  - Image encoder (masked autoencoder (MAE)-pretrained vision transformer (ViT)
  - Prompt encoder (can process points, boxes, text, and masks)
  - Mask decoder (transformer decoder block + dynamic mask predictor)

- SAM can automatically segment all structures in an image by predicting masks for all points on a grid.
- Many works proposed tuned versions of SAM for medical image segmentation[2]

[1]Kirillov, Alexander, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, et al. 2023. “Segment Anything.” arXiv.
[2]Ma, Jun, Yuting He, Feifei Li, Lin Han, Chenyu You, and Bo Wang. 2024. “Segment Anything in Medical Images.” *Nature Communications* 15 (1): 654.







Segment RMSE (SRMSE)

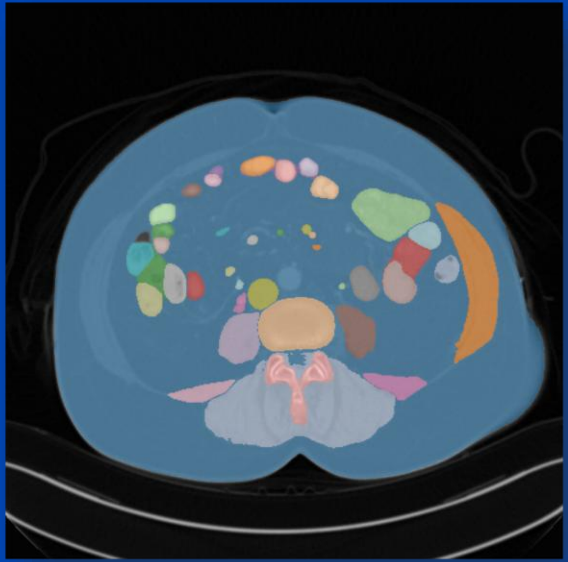
- **Assume *S* segments from SAM for a given patient volume.**
- **Represent each segment by a binary mask volume $m^{(s)} \in \{0, 1\}^N$ with N being the number of voxels.**
- **Define the segment-wise root mean square error between two images *x* and *y*, and segment *s*:**

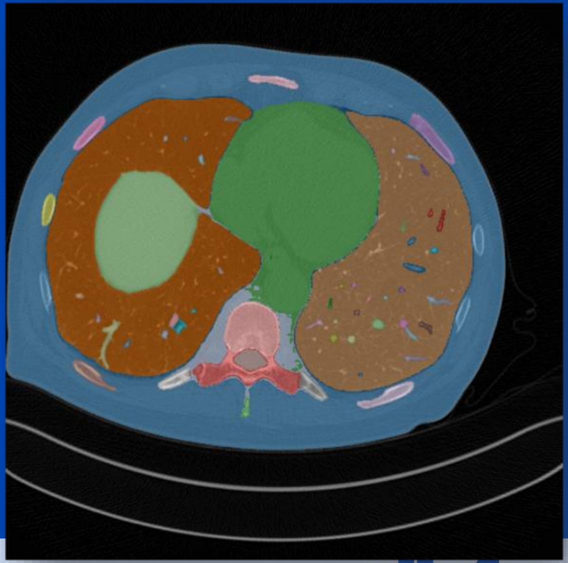
$$\mathrm{SRMSE}(x, y; s) = \sqrt{\frac{\sum_{n=1}^{N} m_n^{(s)} (x_n - y_n)^2}{\sum_{n=1}^{N} m_n^{(s)}}}$$

- **Using the set of all SRMSEs, define**

$$\mathrm{MeanSRMSE}(x, y) = \frac{1}{S} \sum_{s} \mathrm{SRMSE}(x, y; s)$$

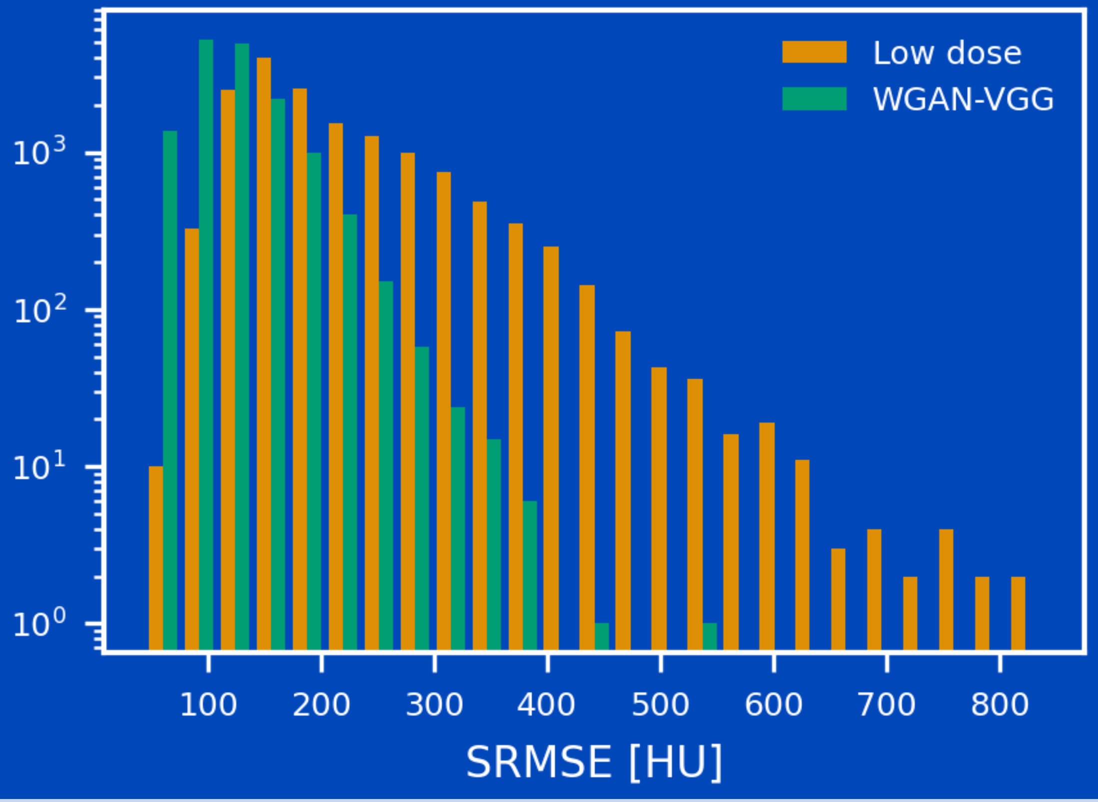
$$\mathrm{MaxSRMSE}(x, y) = \max_{s} \mathrm{SRMSE}(x, y; s)$$

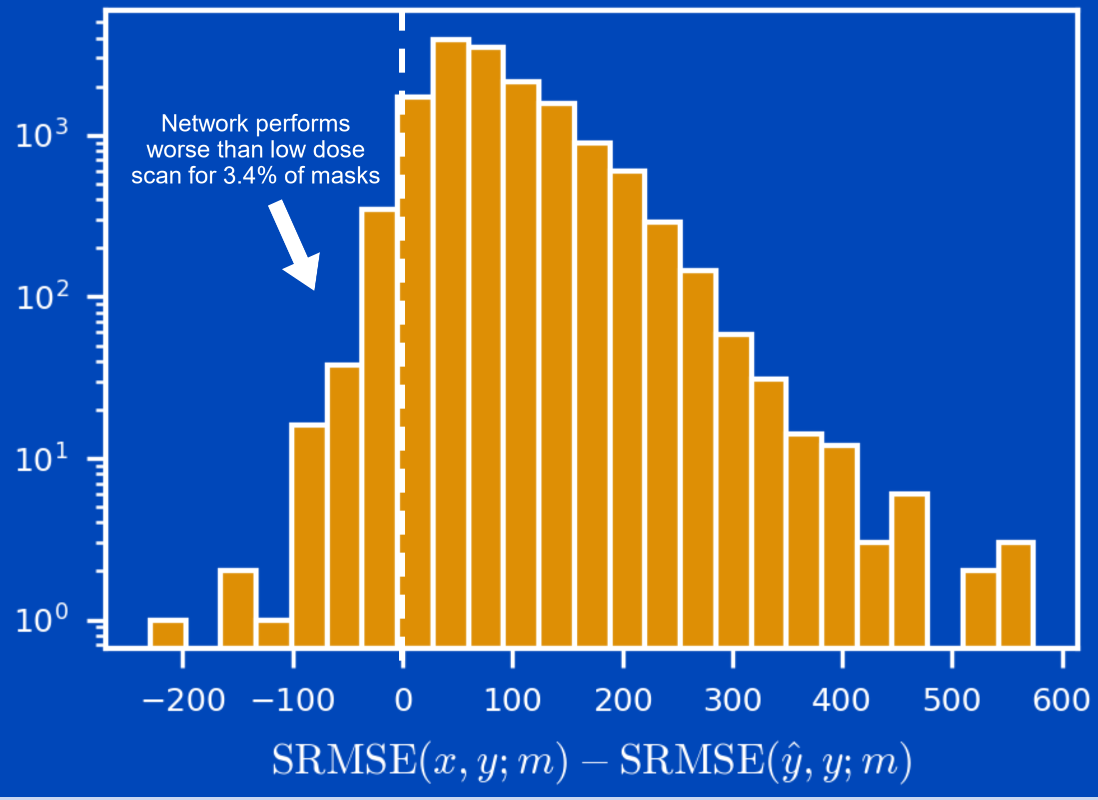




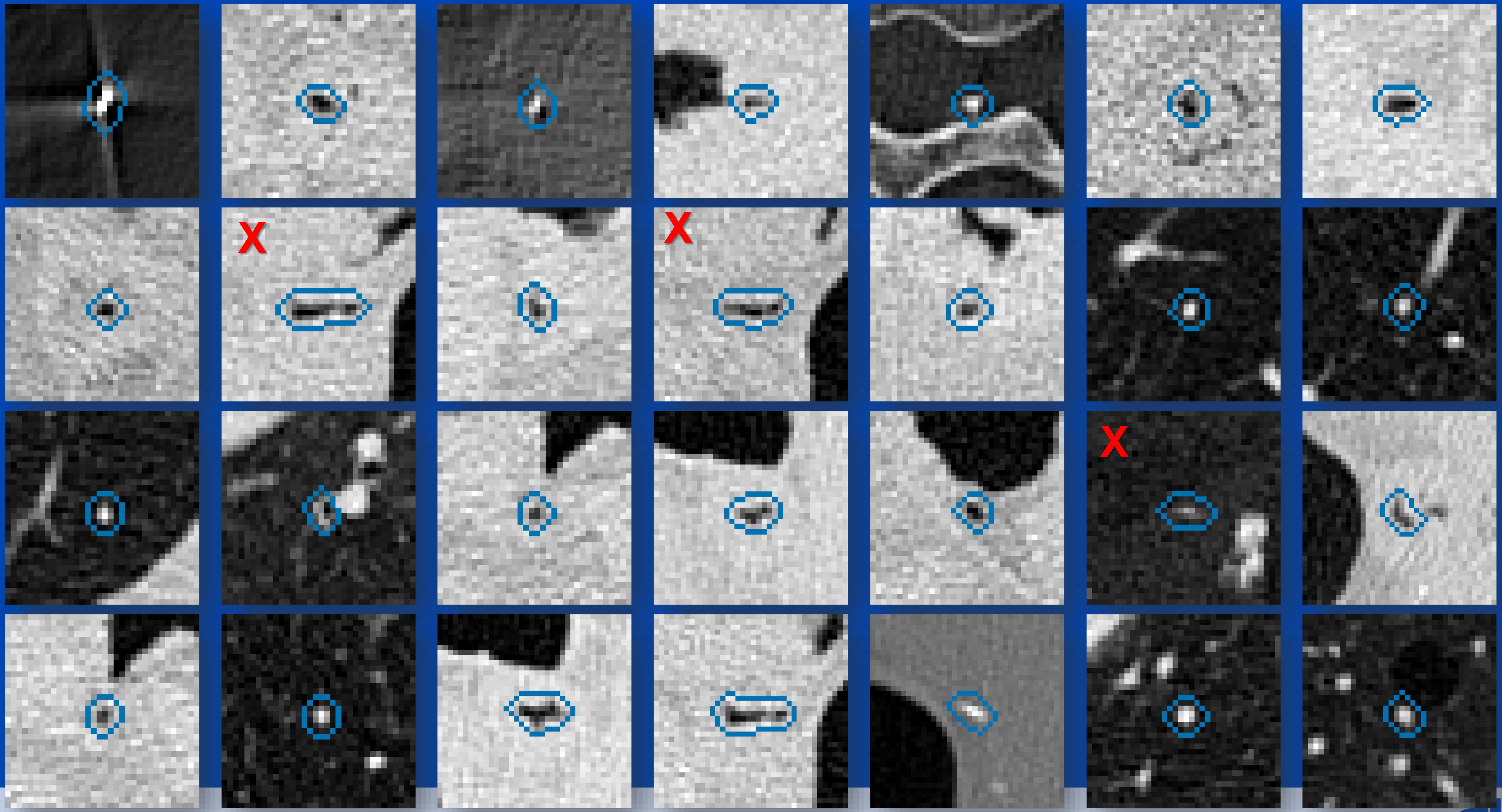
Detecting Hallucinations

- Compare SRMSE of low dose scan (x) with network prediction ($\hat{y}$)
- On a chest scan with 392 axial slices we have a total of 15,547 masks

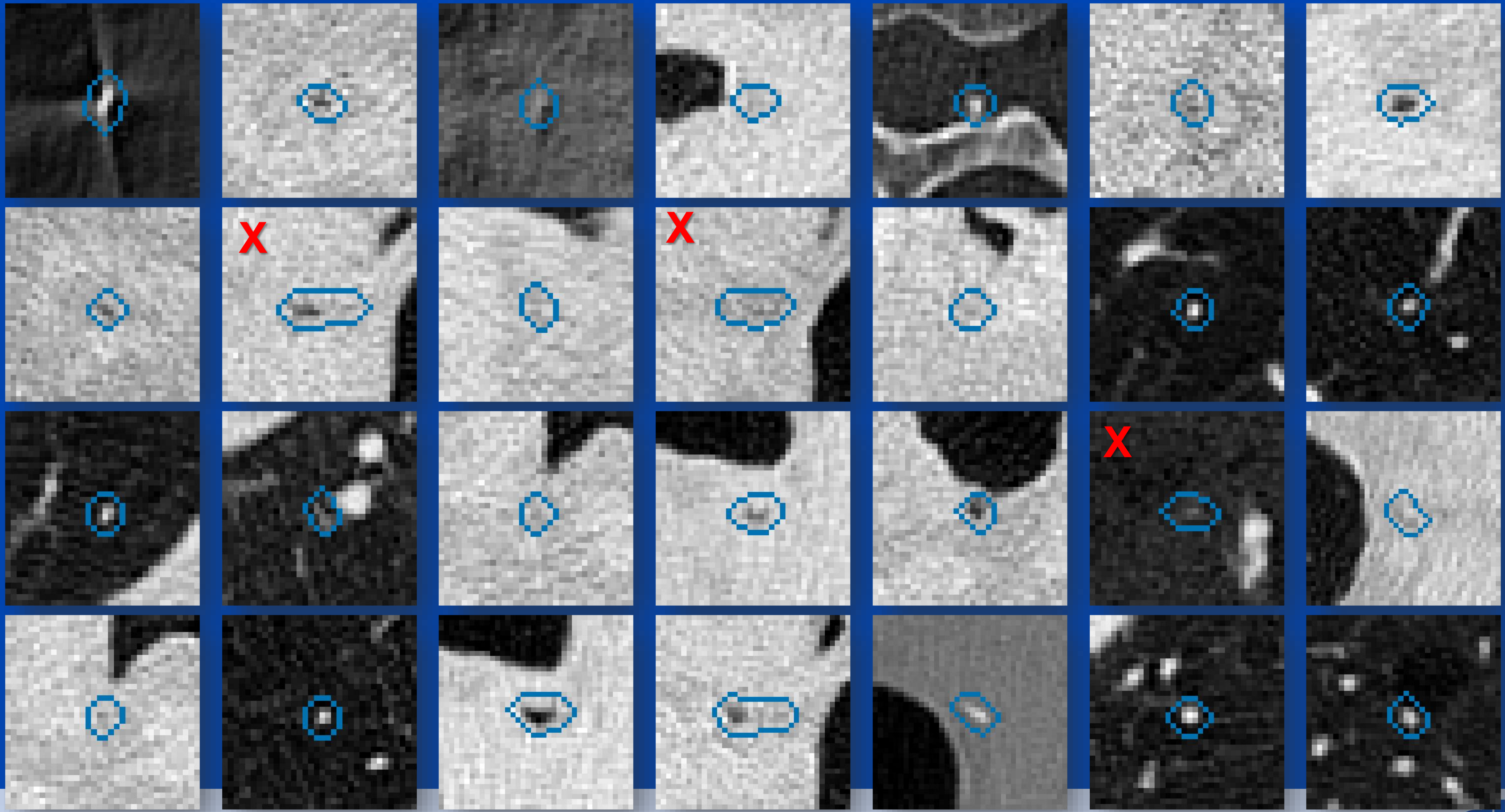




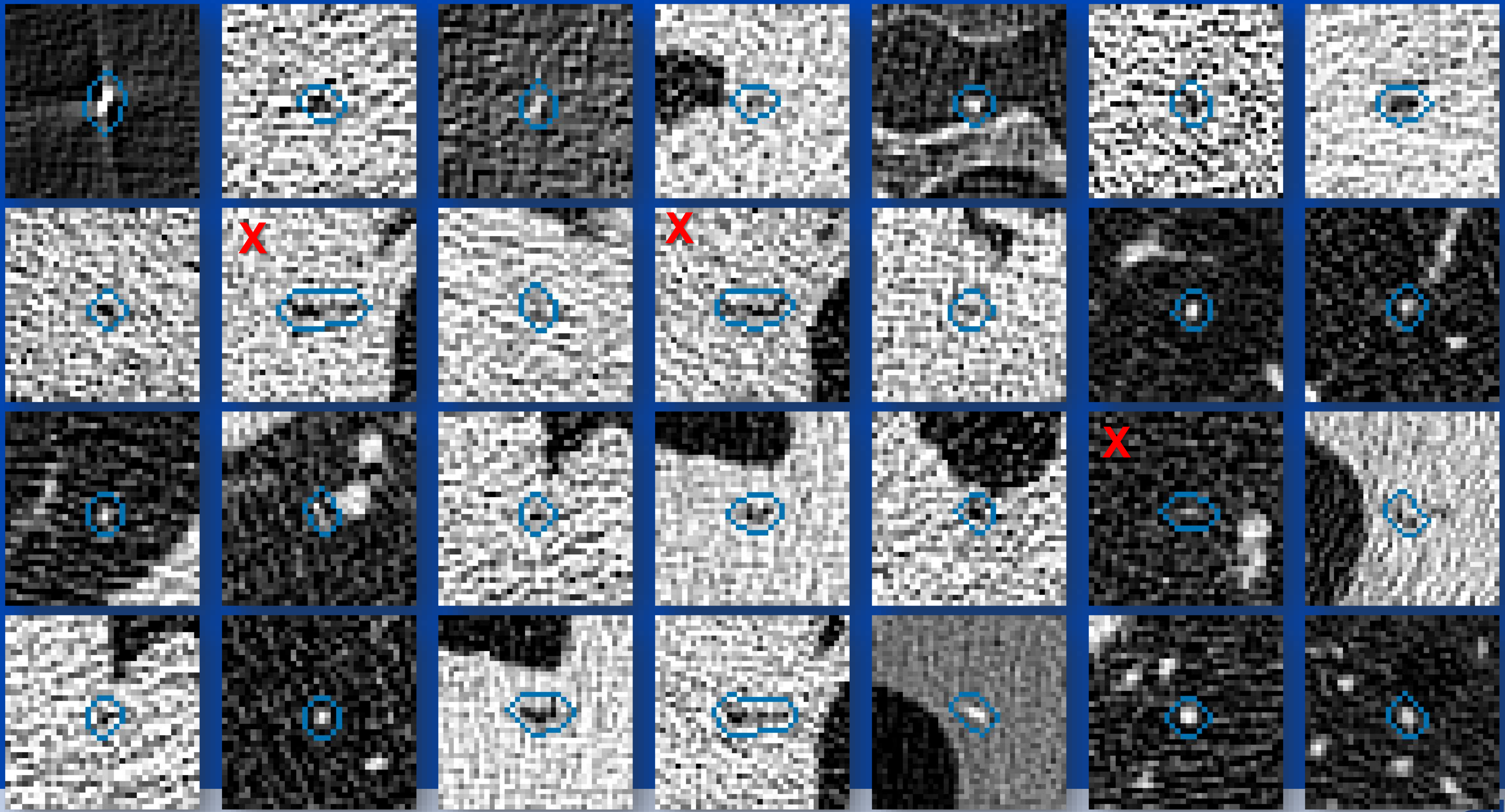
High Dose Images



Network Predictions (WGAN-VGG)

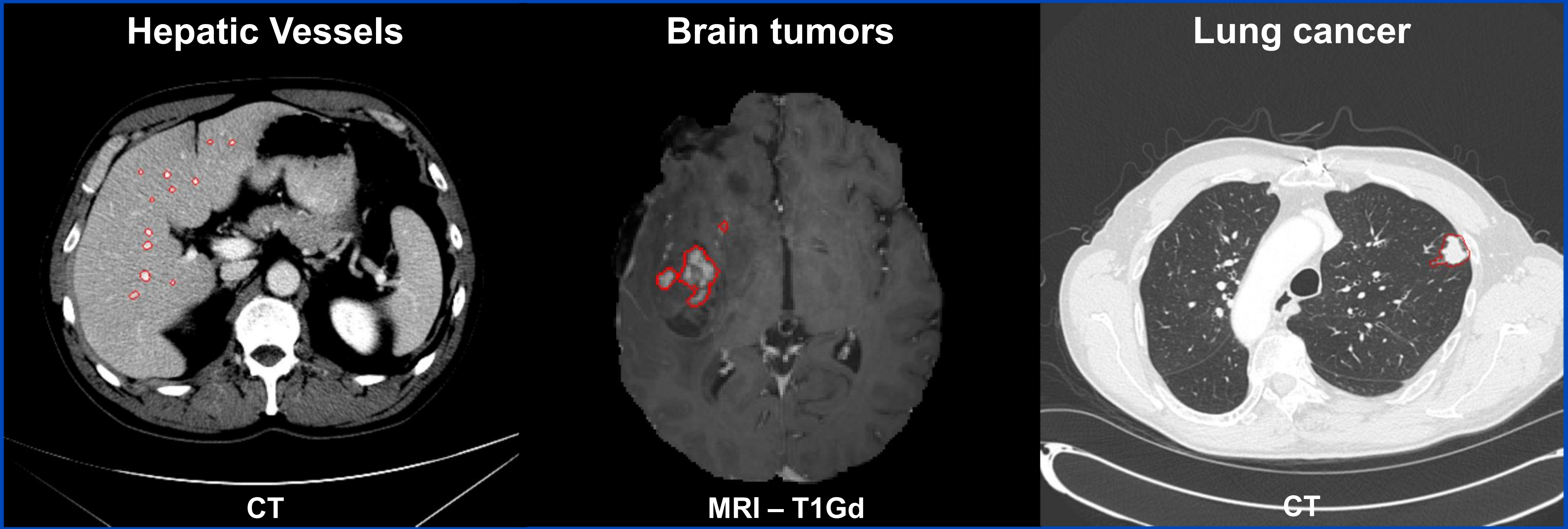


Low Dose Images



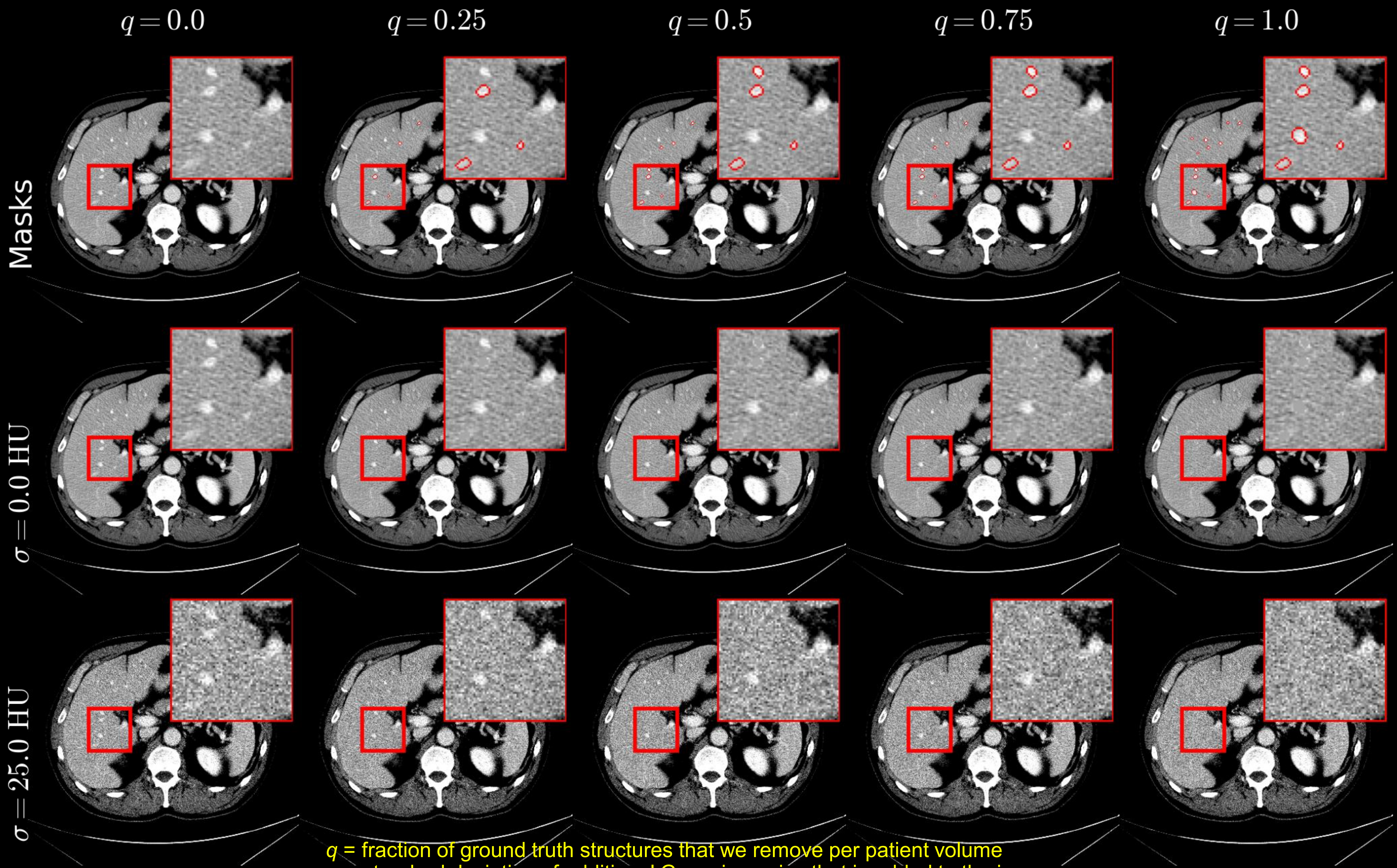
Experiments

- Evaluate the proposed metric on synthetic datasets where the amount of removed structures is known.
- Utilize three datasets from the Medical Decathlon[1], a collection of ten medical image segmentation tasks with ground truth annotations.



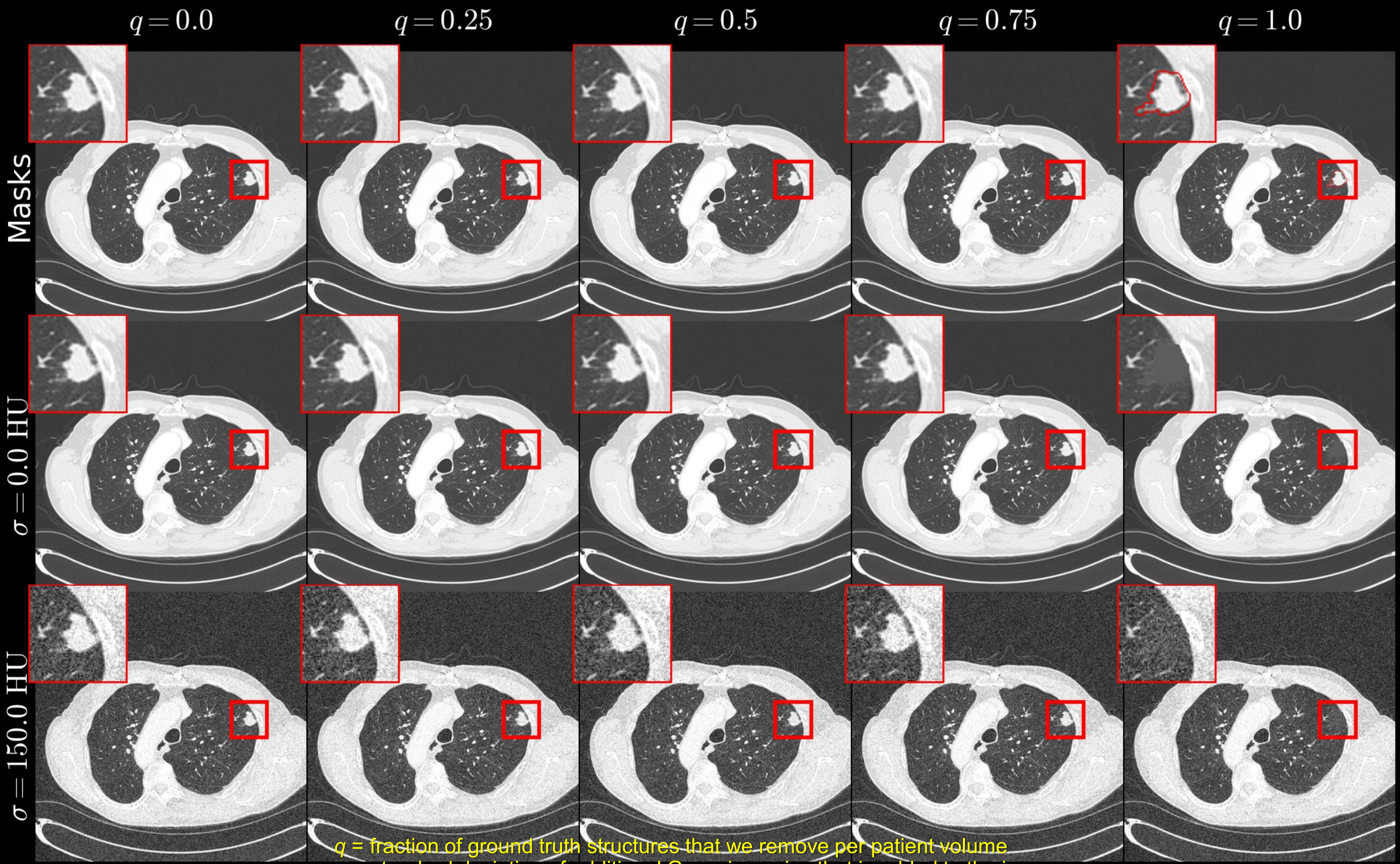
[1]Simpson, Amber L., Michela Antonelli, Spyridon Bakas, Michel Bilello, Keyvan Farahani, Bram van Ginneken, Annette Kopp-Schneider, et al. 2019. “A Large Annotated Medical Image Dataset for the Development and Evaluation of Segmentation Algorithms.” arXiv.

Hepatic Vessels



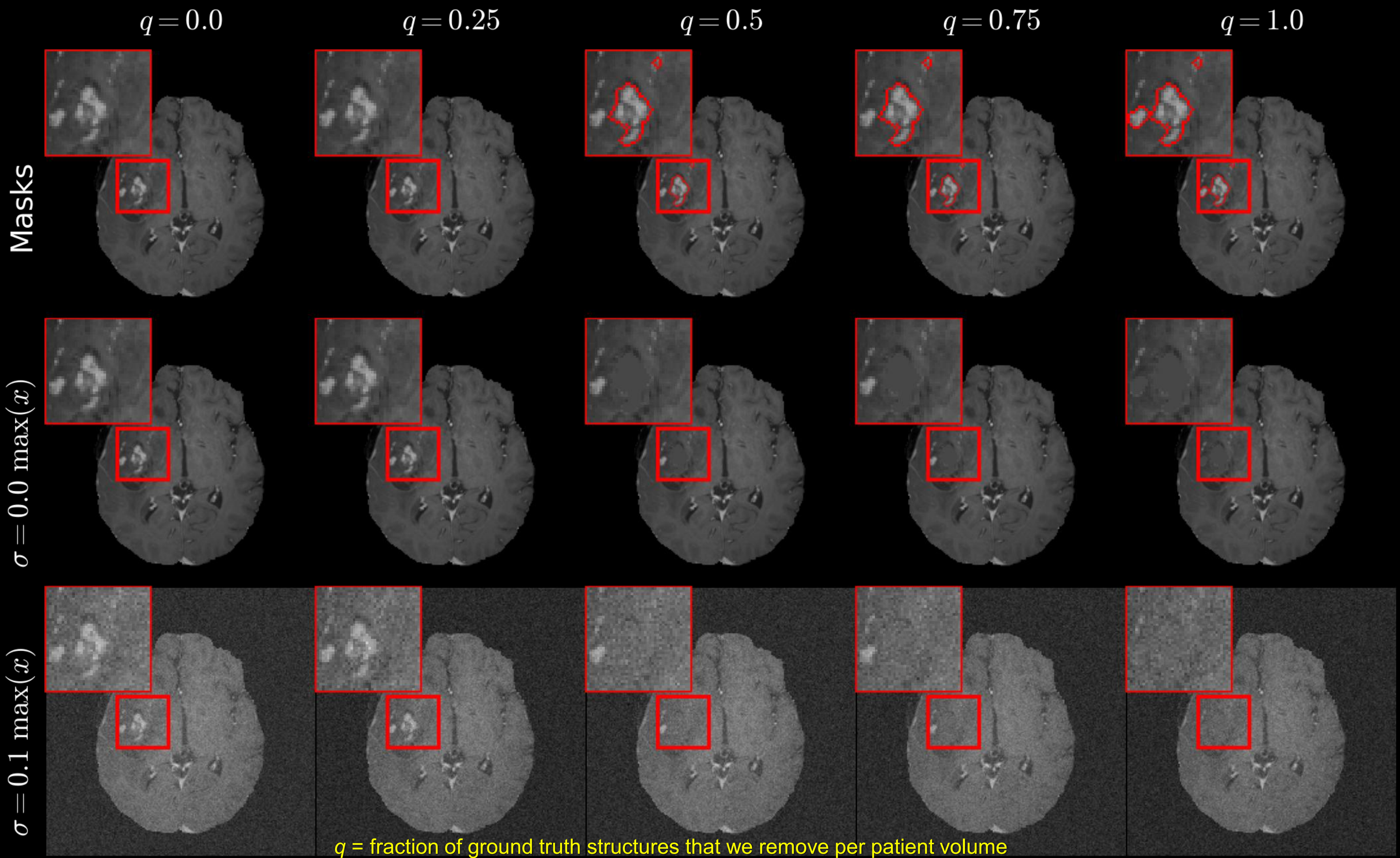
q = fraction of ground truth structures that we remove per patient volume
σ = standard deviation of additional Gaussian noise that is added to the images

Lung Cancer



q = fraction of ground truth structures that we remove per patient volume
σ = standard deviation of additional Gaussian noise that is added to the images

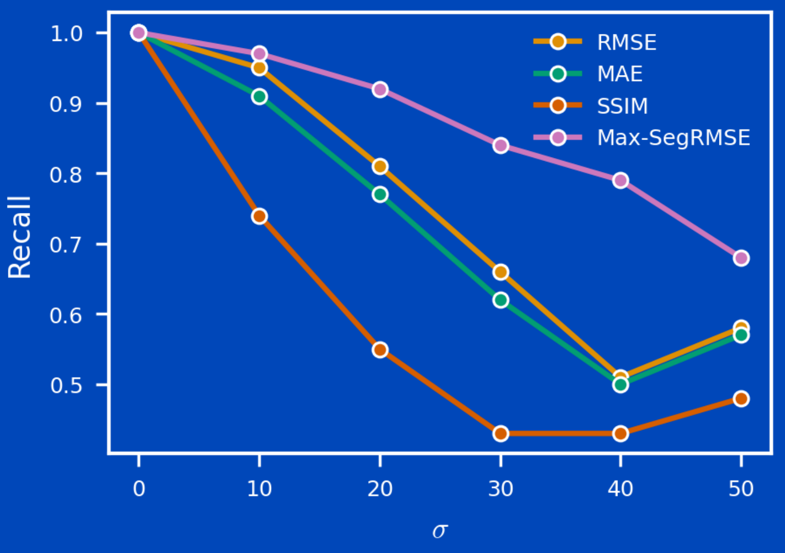
Brain Tumor



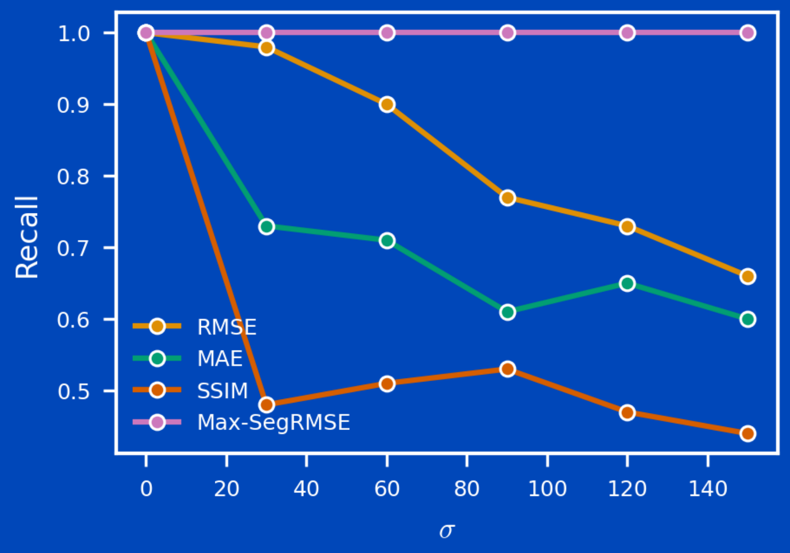
q = fraction of ground truth structures that we remove per patient volume
σ = standard deviation of additional Gaussian noise that is added to the images

Results: True Positive Fraction (Recall)

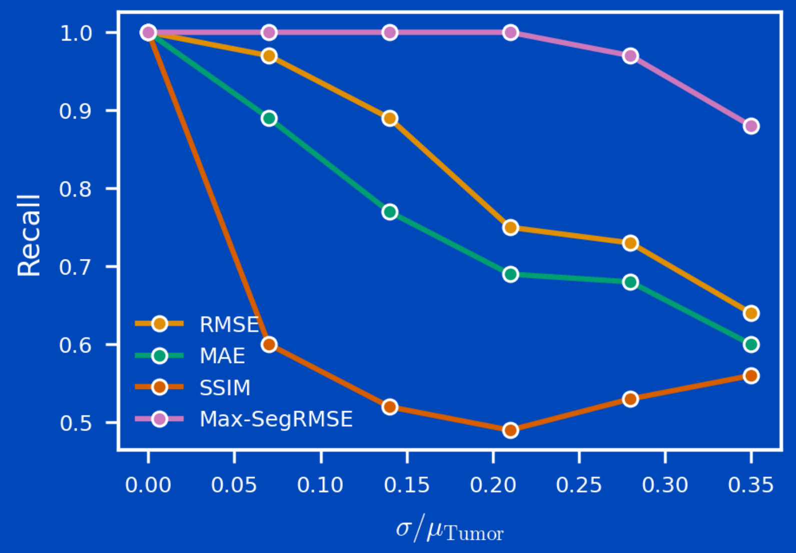
Hepatic Vessels



Lung Cancer



Brain Tumor



The plots show the likelihood that a metric correctly detects that images with q = 0.1 are worse (have more structures removed) than those with q = 0.

Note that q = 0.1 corresponds to about 0.007% to 0.03% modified voxels.

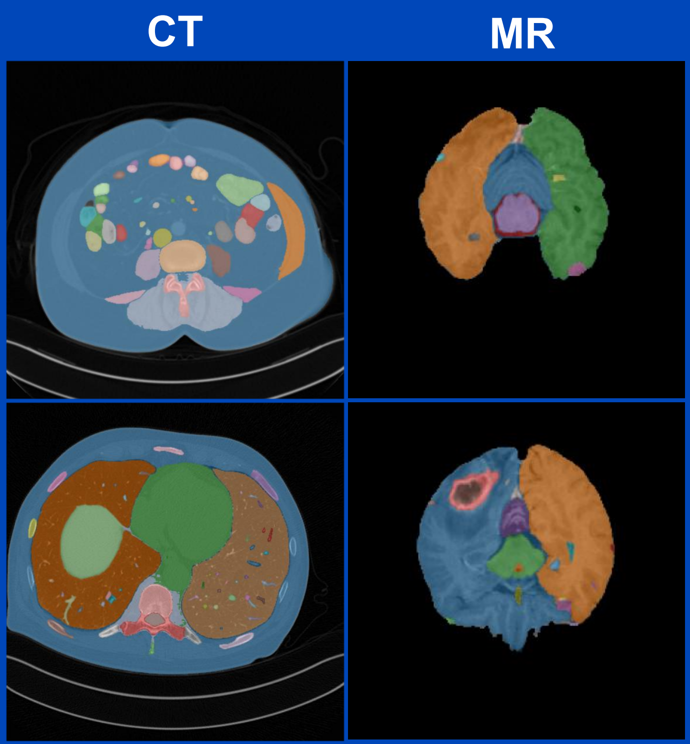
Structure-Aware Metrics for the Evaluation of Deep Learning-Based Image Reconstruction Algorithms

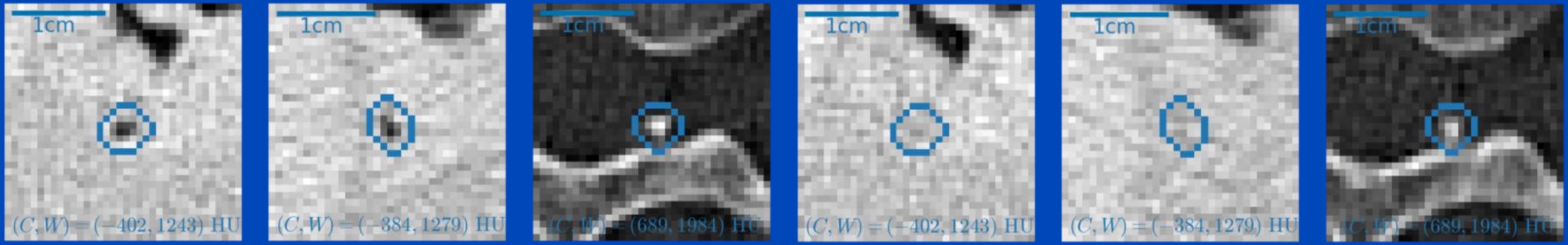
Elias Eulig, Joscha Maier, and Marc Kachelrieß

Metric available on GitHub

github.com/eeulig/structure-aware-metrics

- Evaluate whether your reconstruction algorithm **preserves small structures**
- Our metric **weights pixels inversely proportional to the structure size**
- Uses segment anything model (SAM) to segment structures
  → **works with both CT and MR images**
- Can also be used to **detect hallucinations**





Ground truth

Network prediction

dkfz.

Networks are black boxes. We probe such networks to learn about their internals.

RECONSTRUCTING HALLUCINATIONS

Finding Invariances x^{inv} of Denoising Networks $f_\theta(x)$

Invariances are a special case of hallucinations. Mathematically, they are related to the preimage of *f*.

Adversarial perturbations		Invariances
Small perturbations in the input that lead to large alterations in predictions	⟷	Large perturbations in the input that leave network predictions unaffected

Find invariances x^{inv} via

$$\arg\min_{x^{\text{inv}}} \left(\| f_\theta(x) - f_\theta(x^{\text{inv}}) \| - \alpha \| x - x^{\text{inv}} \| \right)$$

$$\alpha \in \mathbb{R}^+$$

$$x_0^{\text{inv}} = x + n$$

$$n \sim \mathcal{N}(0, 10^{-2})$$

For $\|\cdot\|$ we use:

- **Mean-squared-error (MSE), bounded by +1**
- **Structural dissimilarity: (1-SSIM)/2**
- **Perceptual loss using ImageNet-pretrained VGG16**

Optimize x^{inv} using Adam optimizer for 3000 iterations


May Denoising Remove Structures? How to Reconstruct Invariances of CT Denoising Algorithms

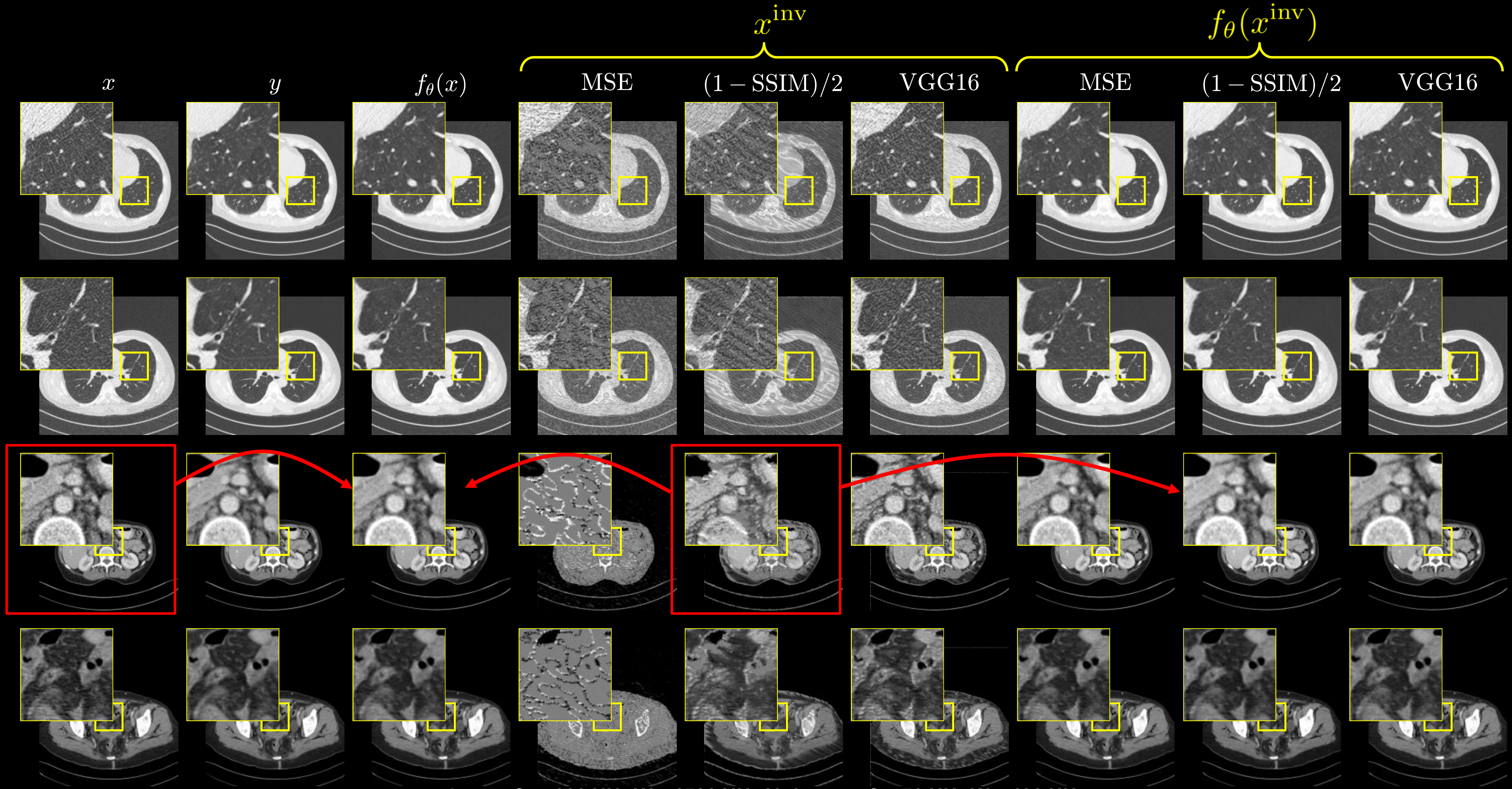
Elias Eulig[a,b,*], Joscha Maier[a], Björn Ommer[c], and Marc Kachelrieß[a,d]

[a]German Cancer Research Center (DKFZ), Heidelberg, Germany
[b]Faculty of Physics and Astronomy, Heidelberg University, Germany
[c]LMU Munich, Germany
[d]Medical Faculty, Heidelberg University, Germany

ABSTRACT

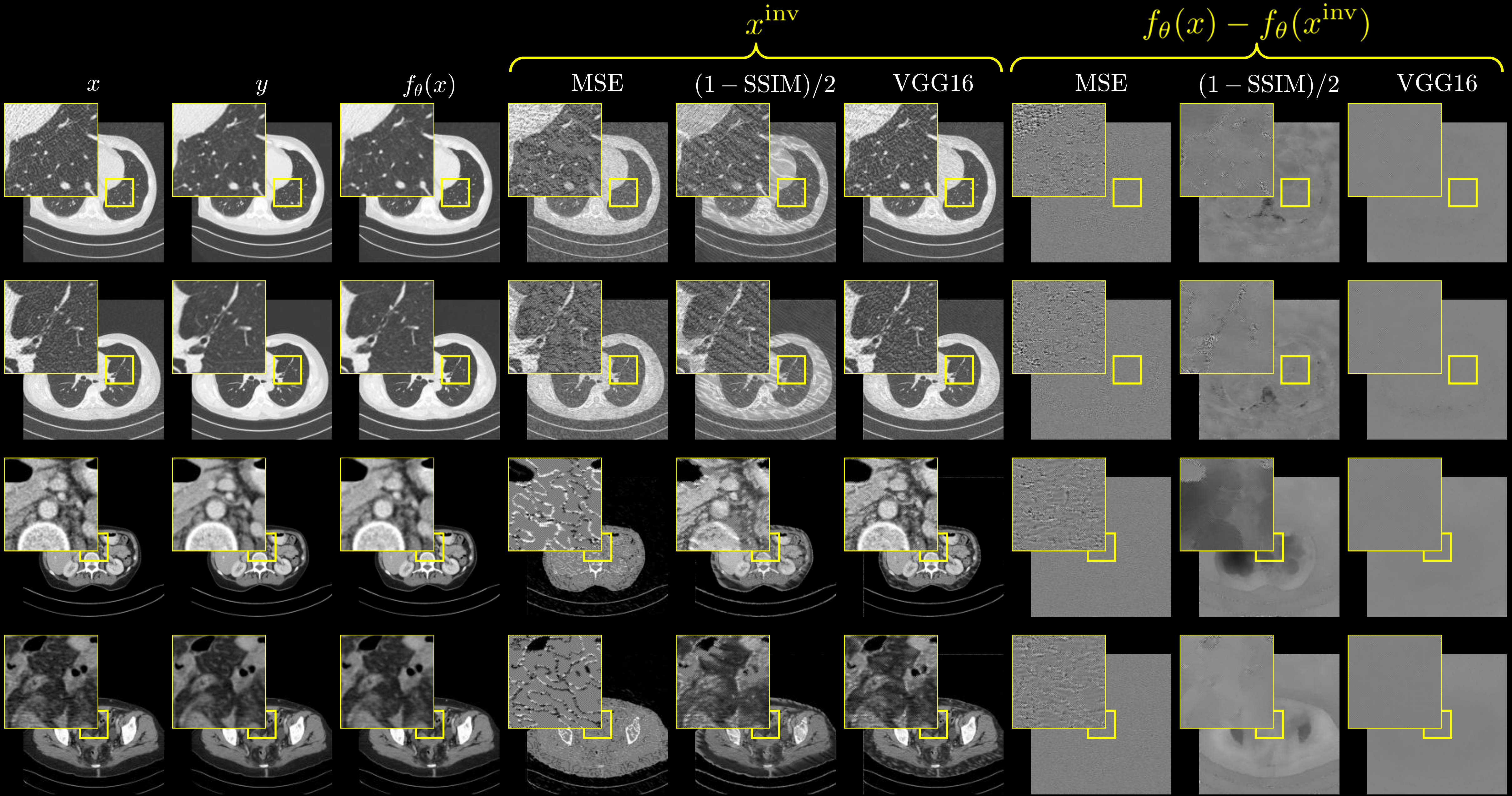
E. Eulig, J. Maier, B. Ommer, and M. Kachelrieß. May denoising remove structures? How to reconstruct invariances of CT denoising algorithms. SPIE Medical Imaging Conference Record, 1292506:1-6, April 2024.

Results for *f*(*x*) = WGAN-VGG



Lung: *C* = -600 HU, *W* = 1500 HU. Abdomen: *C* = 50 HU, *W* = 400 HU.

Results for *f*(*x*) = WGAN-VGG

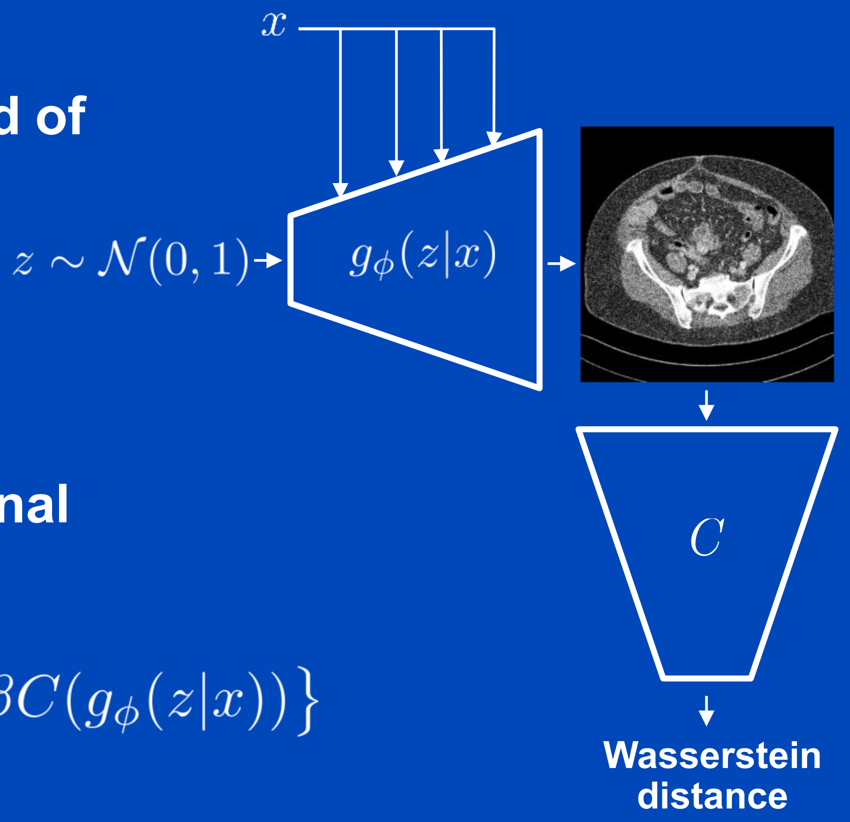


Lung: *C* = -600 HU, *W* = 1500 HU. Abdomen: *C* = 50 HU, *W* = 400 HU.

Future Step: Reconstruct Clinically Relevant Invariances

Drawbacks of previous approach

- Generated x^{inv} generally do not lie on the data manifold of low dose images
- Sampling new invariances requires new x_0^{inv}



Natural invariances

Generate natural (on-manifold) x^{inv} by training a conditional generator $g_\phi(z \sim \mathcal{N}(0,1)|x)$ together with a critic *C*

$$\arg\min_{\phi}\left\{\|f_\theta(x) - f_\theta(g_\phi(z|x))\| - \alpha\|x - g_\phi(z|x)\| + \beta C(g_\phi(z|x))\right\}$$

E. Eulig, J. Maier, B. Ommer, and M. Kachelrieß. May denoising remove structures? How to reconstruct invariances of CT denoising algorithms. SPIE Medical Imaging Conference Record, 1292506:1-6, April 2024.

Generating Invariances

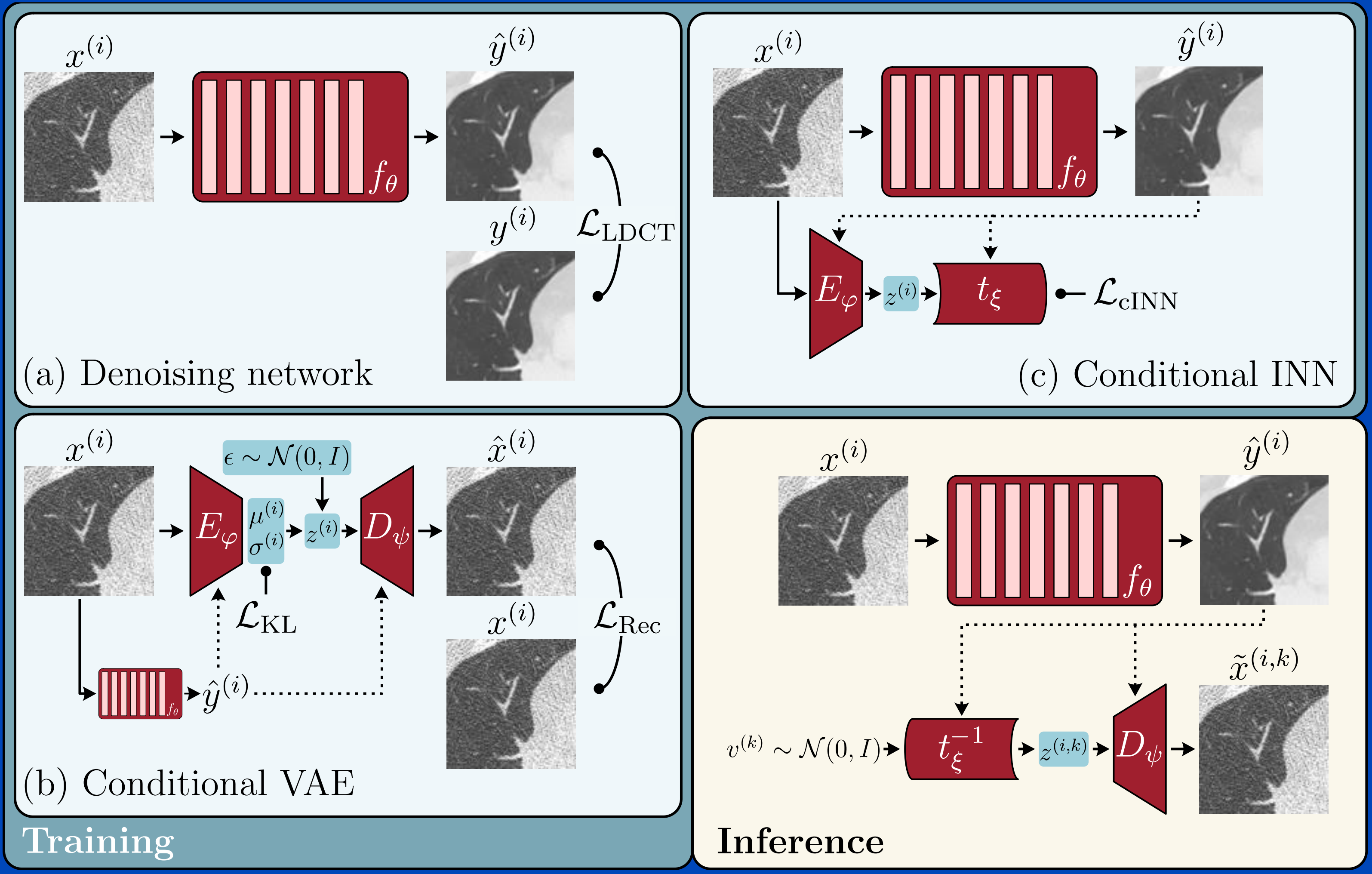
Pipeline to reconstruct invariances:

a) LDCT denoising network, that predicts high dose images $\hat{y}^{(i)} = f(x^{(i)})$ from low dose images

b) Conditional VAE that is trained to reconstruct low dose images $x^{(i)}$ and is conditioned on denoised images $\hat{y}^{(i)}$

c) cINN that disentangles the information in z that the denoising network is invariant to from the one it is not invariant to

To reconstruct invariances we can sample from the Gaussian distribution of invariances, apply the inverse cINN and decode samples using (fixed) conditional decoder



E. Eulig, F. Jäger, J. Maier, B. Ommer, and M. Kachelrieß. Reconstructing and analyzing the invariances of low-dose CT image denoising networks. Med. Phys. 52(1):188-200, January 2025.

Conclusions

- New metrics are needed to quantify changes in subtle details.
  - Needed to evaluate the quality of AI-based algorithms.
  - Could become part of the loss function to train networks.
  - May help to determine the amount of dose reduction possible for a given algorithm.
- Methods to reconstruct hallucinations are required
  - Probe existing algorithms
  - Improve the development of new algorithms

Thank you!

- This presentation will soon be available at www.dkfz.de/ct.
- Job opportunities through DKFZ's international PhD or Postdoctoral Fellowship programs (marc.kachelriess@dkfz.de).
- Parts of the reconstruction software were provided by RayConStruct® GmbH, Nürnberg, Germany.