Problems in Assessing AI Image Quality

Marc Kachelrieß

marc.kachelriess@dkfz.de

German Cancer Research Center (DKFZ)

Heidelberg, Germany

www.dkfz.de/ct

dkfz. DEUTSCHES KREBSFORSCHUNGSZENTRUM IN DER HELMHOLTZ-GEMEINSCHAFT

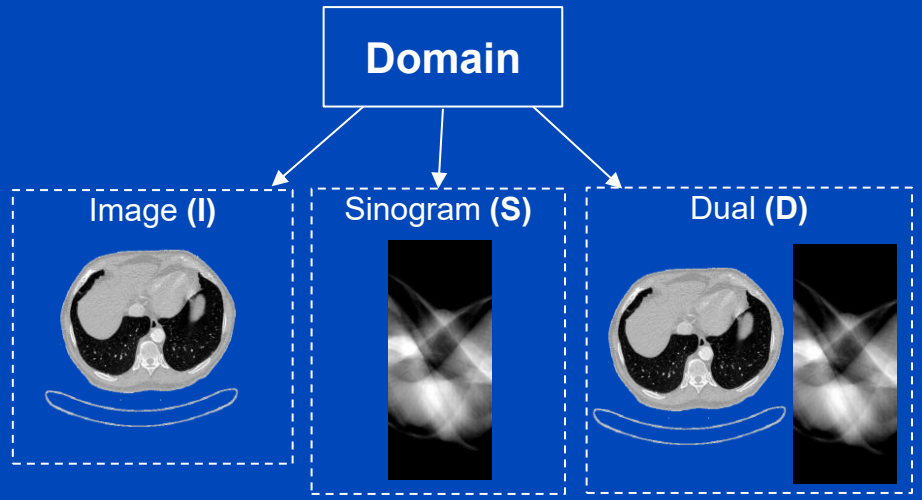
Potential Pitfalls and their Clinical Risk

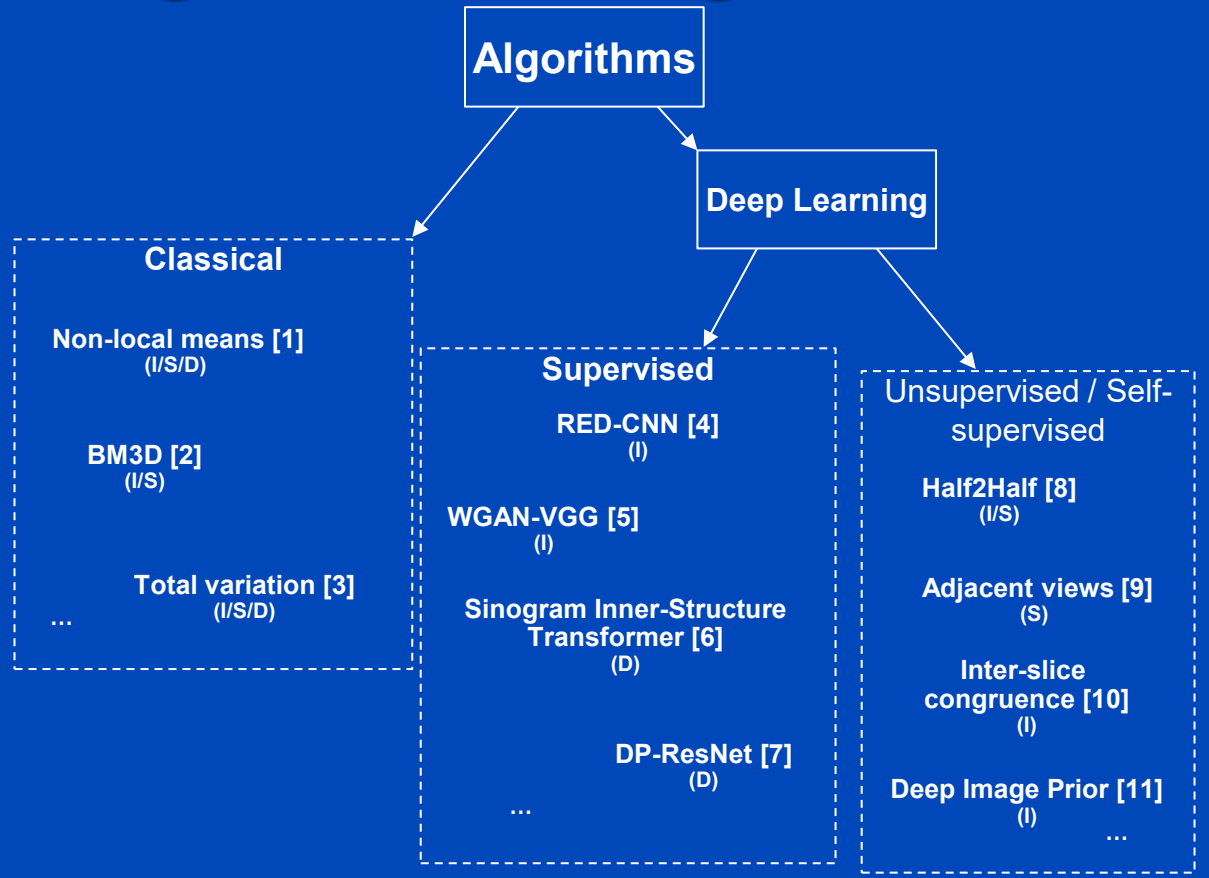
Pitfall	Likelihood	Clinical Risk	Comments
Lesions are removed or introduced: hallucinations	unkown	high	• May completely impair the diagnosis. • The user cannot detect the error (looks like a real lesion). • May easily happen with image-based AI. Very unlikely in projection-based AI.
Apparent spatial 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.

Denoising benchmark with surprising results

IS NEWER ALWAYS BETTER?

Low Dose CT Image Denoising

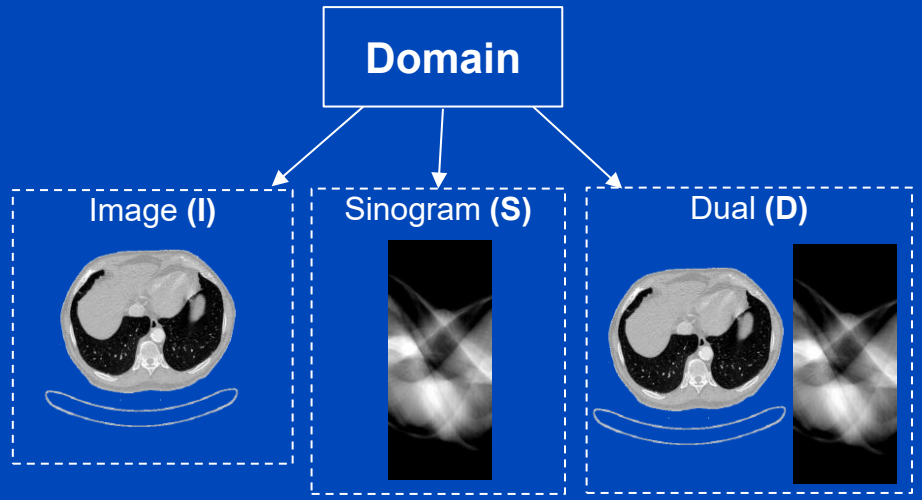


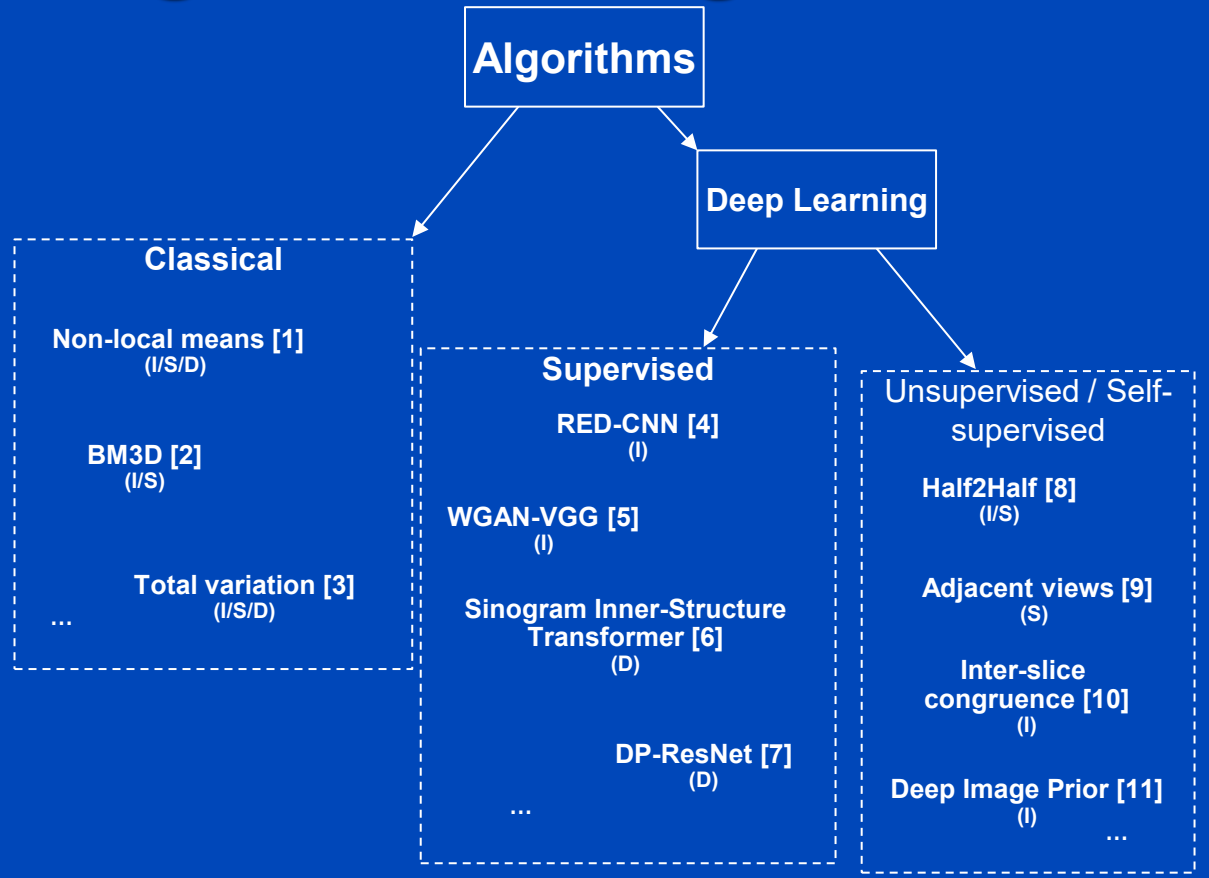


[1] Yuan Y, Zhang YB, Yu HY (2018) “Adaptive nonlocal means method for denoising basis material images […]”. J Comput Assist Tomogr 42:972–981.
[2] Feruglio, P Fumene, C Vinegoni, J Gros, A Sbarbati, and R Weissleder (2010) “Block Matching 3D Random Noise Filtering for Absorption Optical Projection Tomography.” *Physics in Medicine and Biology* 55 (18): 5401–15.
[3] Jia L, Zhang Q, Shang Y, et al. (2018) “Denoising for low-dose CT image by discriminative weighted nuclear norm minimization”. IEEE Access
[4] Chen, Hu, Yi Zhang, Mannudeep K. Kalra, Feng Lin, Yang Chen, Peixi Liao, Jiliu Zhou, and Ge Wang. 2017. “Low-Dose CT with a Residual Encoder-Decoder Convolutional Neural Network.” *IEEE Transactions on Medical Imaging* 36 (12): 2524–35.
[5] Yang, Qingsong, Pingkun Yan, Yanbo Zhang, et al.. 2018. “Low-Dose CT Image Denoising Using a Generative Adversarial Network with Wasserstein Distance and Perceptual Loss.” *IEEE Transactions on Medical Imaging* 37 (6): 1348–57.
[6] L. Yang, Z. Li, R. Ge, J. Zhao, H. Si and D. Zhang, "Low-Dose CT Denoising via Sinogram Inner-Structure Transformer," in IEEE Transactions on Medical Imaging, vol. 42, no. 4, pp. 910-921, 2023.
[7] Yin, Xiangrui, Qianlong Zhao, Jin Liu, Wei Yang, Jian Yang, Guotao Quan, Yang Chen, Huazhong Shu, Limin Luo, and Jean-Louis Coatrieux. 2019. “Domain Progressive 3D Residual Convolution Network to Improve Low-Dose CT Imaging.” *IEEE TMI* 38 (12).
[8] Yuan, N., Zhou, J., & Qi, J. (2020). Half2Half: deep neural network based CT image denoising without independent reference data. *Physics in Medicine & Biology*, *65*(21), 215020.
[9] Hong, Zixuan, Dong Zeng, Xi Tao, and Jianhua Ma. 2023. “Learning CT Projection Denoising from Adjacent Views.” *Medical Physics* 50 (3): 1367–77.
[10] Bera, Sutanu, and Prabir Kumar Biswas. 2023. “Self Supervised Low Dose Computed Tomography Image Denoising Using Invertible Network Exploiting Inter Slice Congruence.” In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 5603–12.
[11] Baguer, Daniel Otero, Johannes Leuschner, and Maximilian Schmidt. 2020. “Computed Tomography Reconstruction Using Deep Image Prior and Learned Reconstruction Methods.” *Inverse Problems* 36 (9.

dkfz.

Low Dose CT Image Denoising

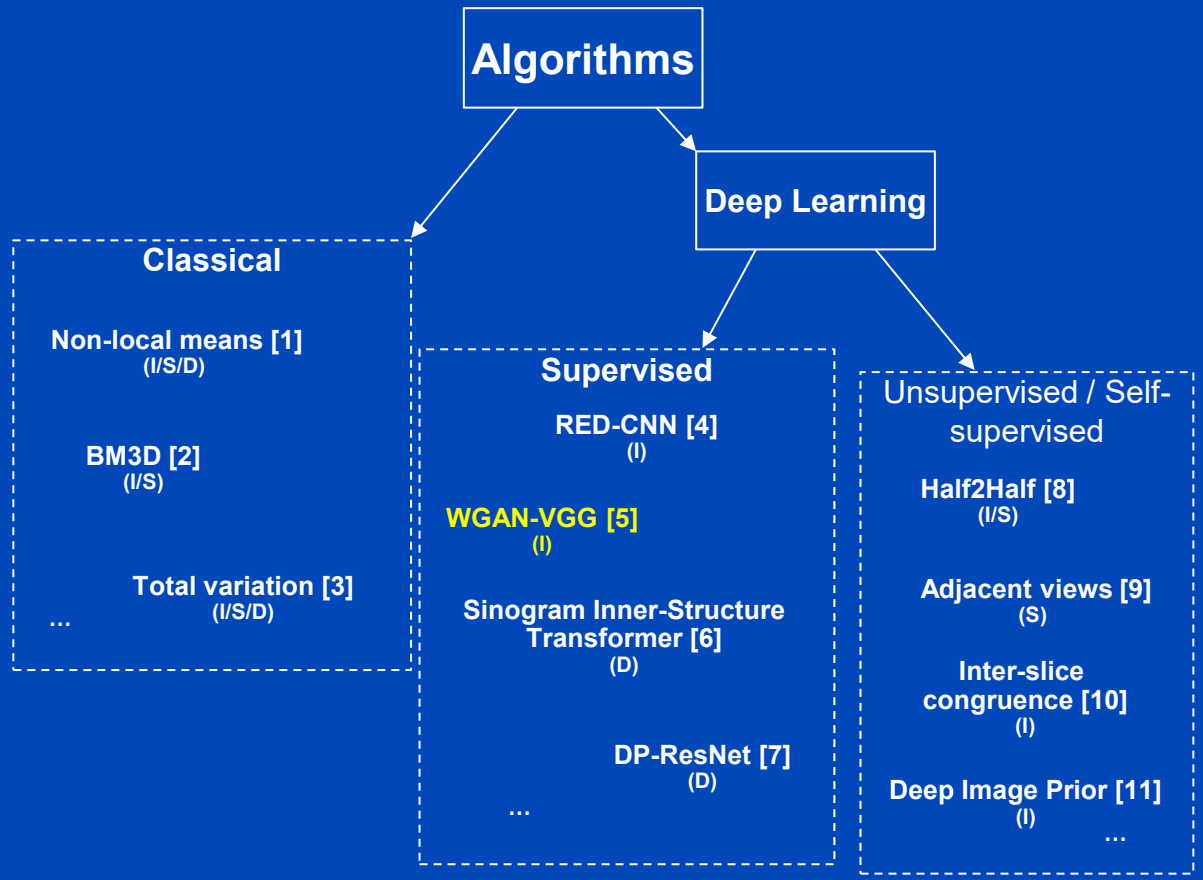


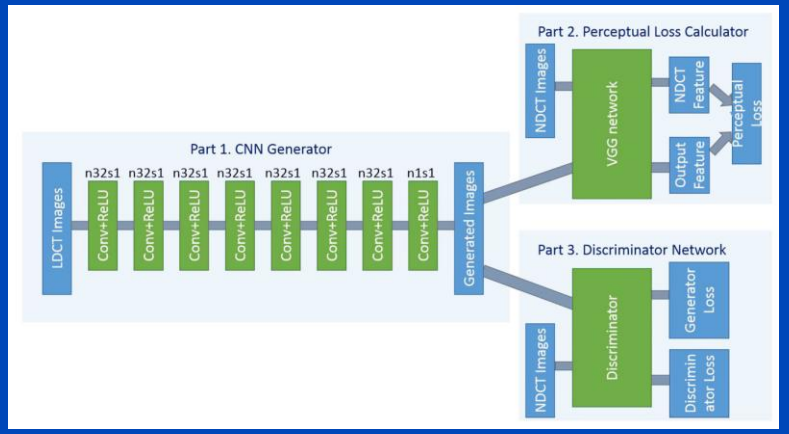


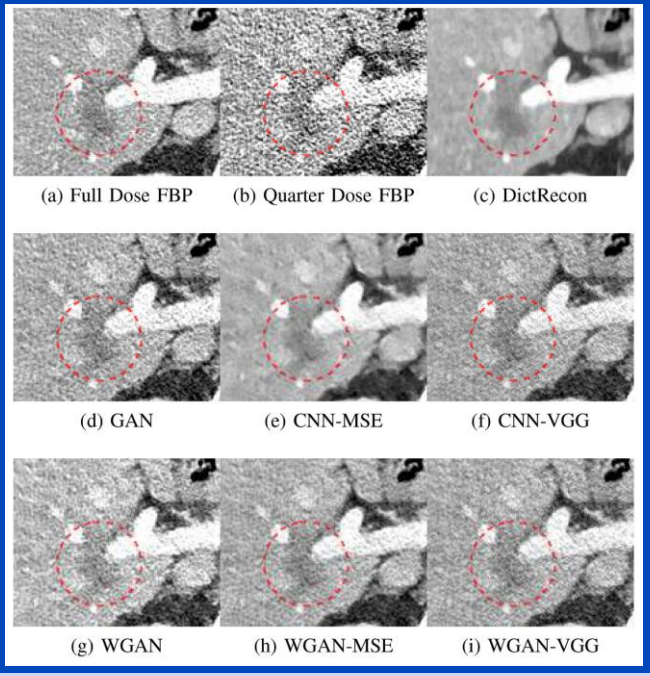
[1] Yuan Y, Zhang YB, Yu HY (2018) “Adaptive nonlocal means method for denoising basis material images […]”. J Comput Assist Tomogr 42:972–981.
[2] Feruglio, P Fumene, C Vinegoni, J Gros, A Sbarbati, and R Weissleder (2010) “Block Matching 3D Random Noise Filtering for Absorption Optical Projection Tomography.” *Physics in Medicine and Biology* 55 (18): 5401–15.
[3] Jia L, Zhang Q, Shang Y, et al. (2018) “Denoising for low-dose CT image by discriminative weighted nuclear norm minimization”. IEEE Access
[4] Chen, Hu, Yi Zhang, Mannudeep K. Kalra, Feng Lin, Yang Chen, Peixi Liao, Jiliu Zhou, and Ge Wang. 2017. “Low-Dose CT with a Residual Encoder-Decoder Convolutional Neural Network.” *IEEE Transactions on Medical Imaging* 36 (12): 2524–35.
[5] Yang, Qingsong, Pingkun Yan, Yanbo Zhang, et al.. 2018. “Low-Dose CT Image Denoising Using a Generative Adversarial Network with Wasserstein Distance and Perceptual Loss.” *IEEE Transactions on Medical Imaging* 37 (6): 1348–57.
[6] L. Yang, Z. Li, R. Ge, J. Zhao, H. Si and D. Zhang, "Low-Dose CT Denoising via Sinogram Inner-Structure Transformer," in IEEE Transactions on Medical Imaging, vol. 42, no. 4, pp. 910-921, 2023.
[7] Yin, Xiangrui, Qianlong Zhao, Jin Liu, Wei Yang, Jian Yang, Guotao Quan, Yang Chen, Huazhong Shu, Limin Luo, and Jean-Louis Coatrieux. 2019. “Domain Progressive 3D Residual Convolution Network to Improve Low-Dose CT Imaging.” *IEEE TMI* 38 (12).
[8] Yuan, N., Zhou, J., & Qi, J. (2020). Half2Half: deep neural network based CT image denoising without independent reference data. *Physics in Medicine & Biology*, *65*(21), 215020.
[9] Hong, Zixuan, Dong Zeng, Xi Tao, and Jianhua Ma. 2023. “Learning CT Projection Denoising from Adjacent Views.” *Medical Physics* 50 (3): 1367–77.
[10] Bera, Sutanu, and Prabir Kumar Biswas. 2023. “Self Supervised Low Dose Computed Tomography Image Denoising Using Invertible Network Exploiting Inter Slice Congruence.” In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 5603–12.
[11] Baguer, Daniel Otero, Johannes Leuschner, and Maximilian Schmidt. 2020. “Computed Tomography Reconstruction Using Deep Image Prior and Learned Reconstruction Methods.” *Inverse Problems* 36 (9.

dkfz.

Low Dose CT Image Denoising







Low Dose CT Image Denoising

Algorithms

Deep Learning

Classical

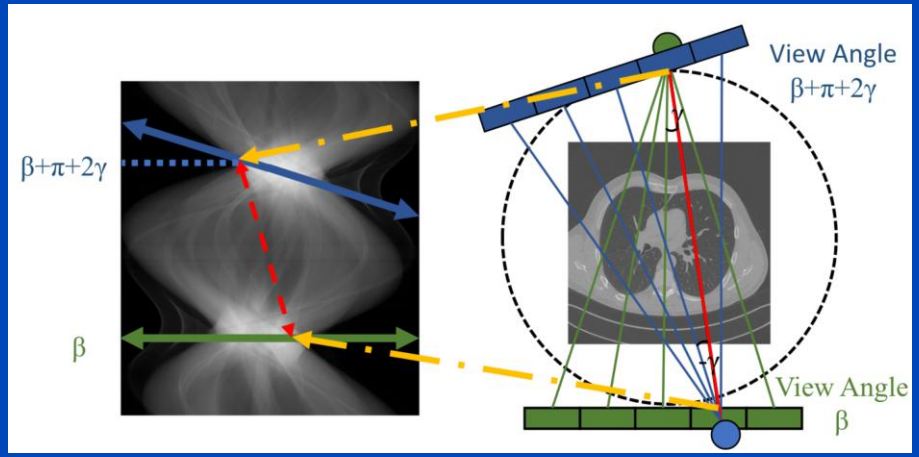
- Non-local means [1] (I/S/D)
- BM3D [2] (I/S)
- Total variation [3] (I/S/D)
- ...

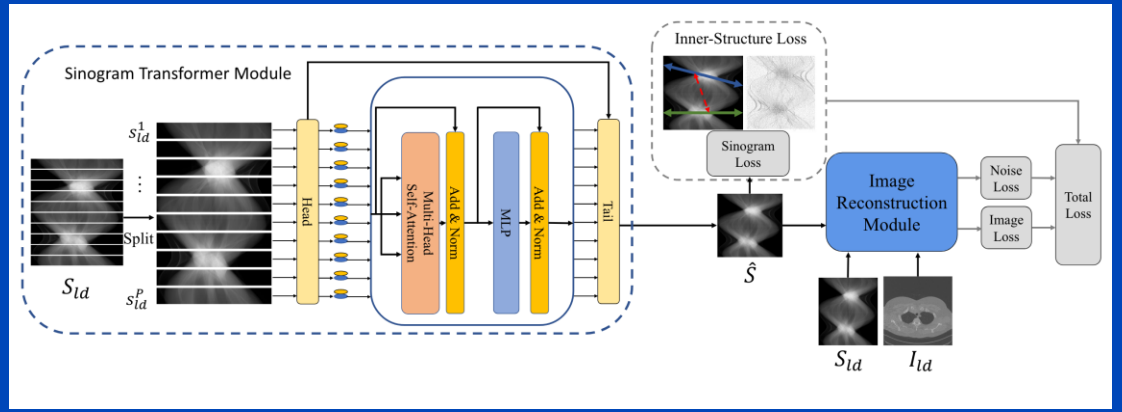
Supervised

- RED-CNN [4] (I)
- WGAN-VGG [5] (I)
- Sinogram Inner-Structure Transformer [6] (D)
- DP-ResNet [7] (D)
- ...

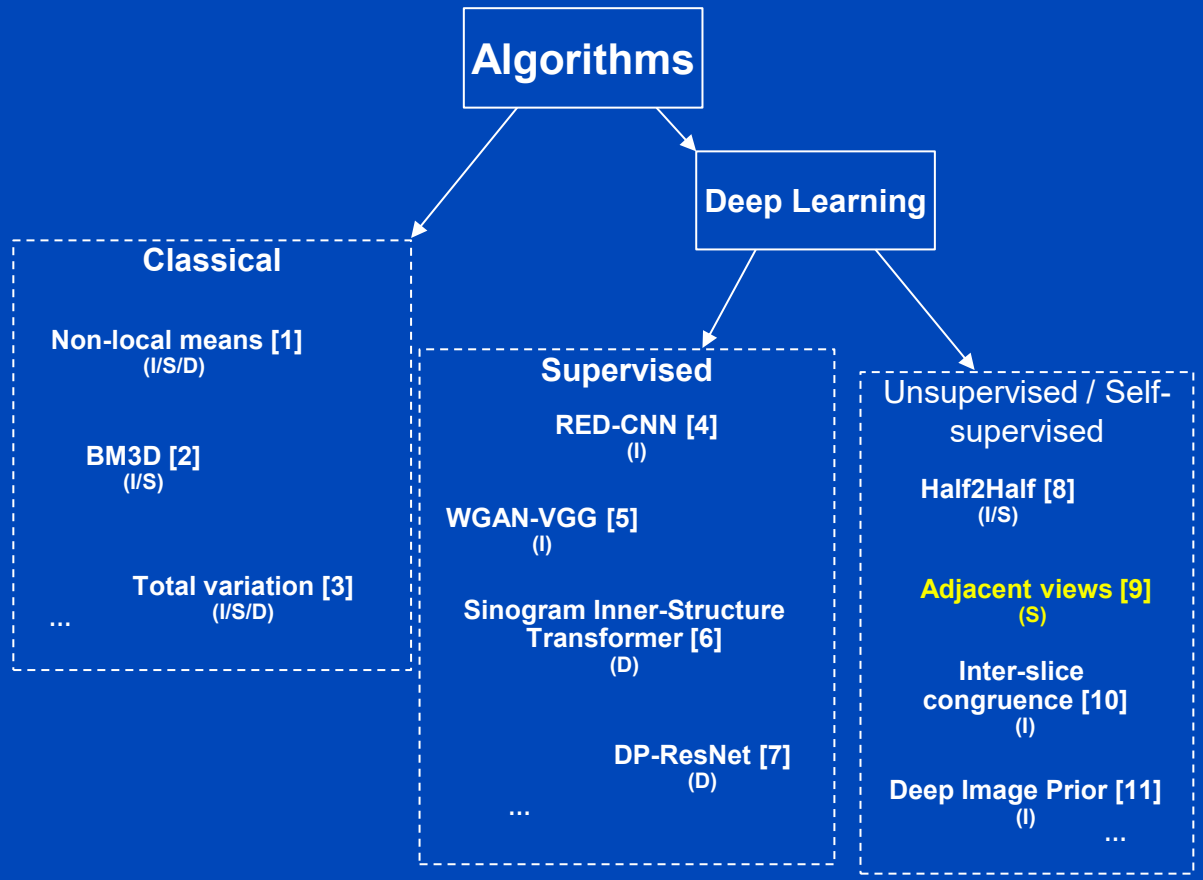
Unsupervised / Self-supervised

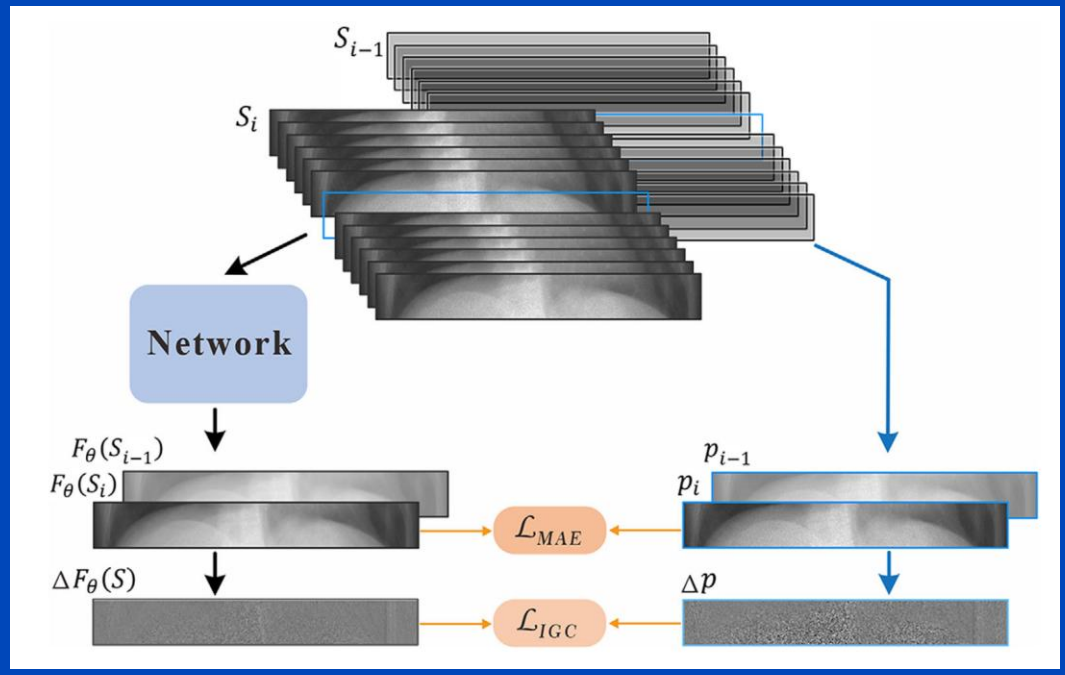
- Half2Half [8] (I/S)
- Adjacent views [9] (S)
- Inter-slice congruence [10] (I)
- Deep Image Prior [11] (I)
- ...





Low Dose CT Image Denoising

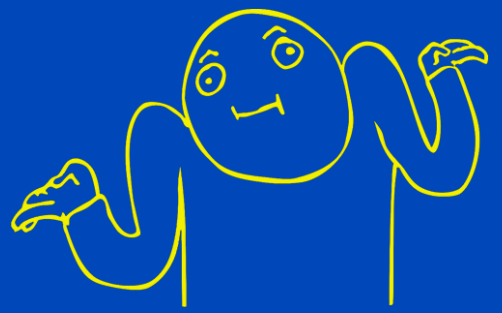




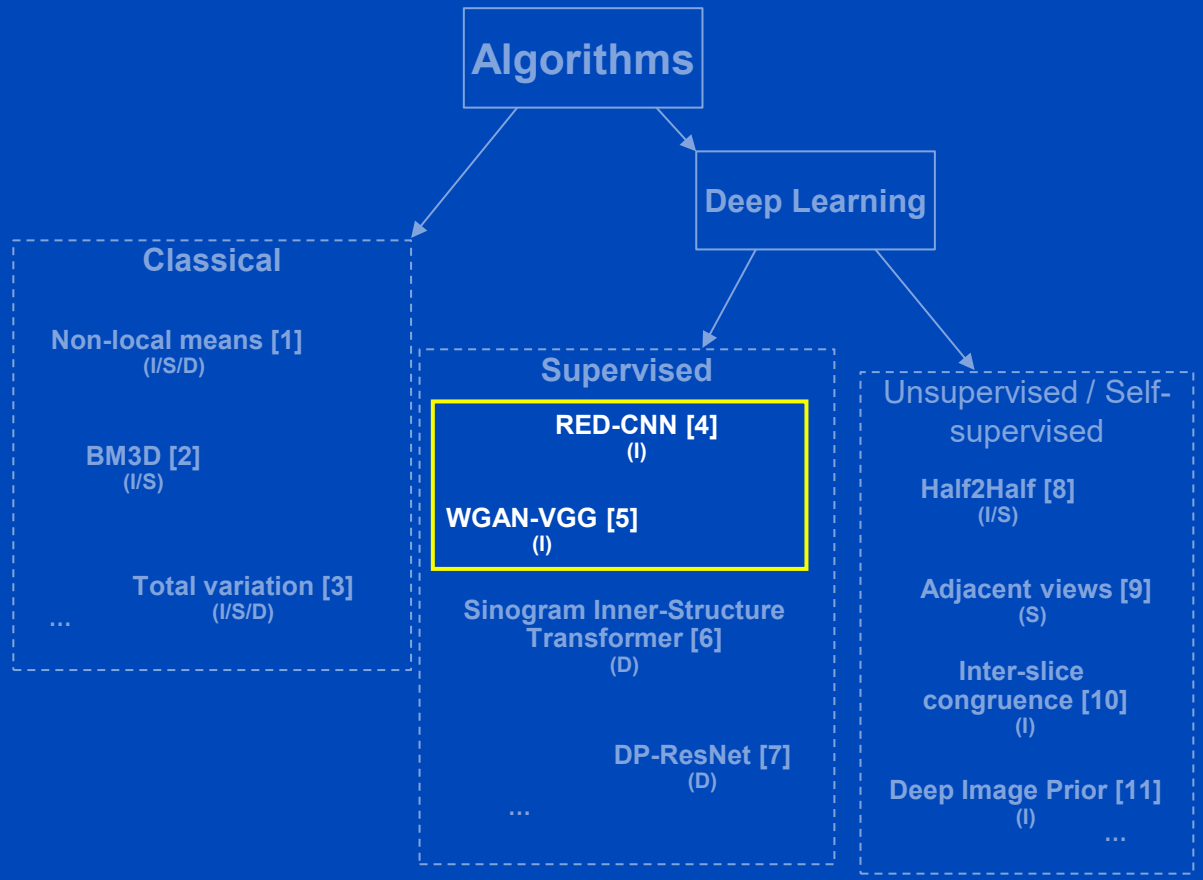
Low Dose CT Image Denoising

Q: Which algorithm performs best?

A:



Low Dose CT Image Denoising



$$\theta^* = \arg\min_{\theta} \mathbb{E}_{x,y\sim\mathcal{D}^{\text{train}}} \|f_\theta(x) - y\|$$

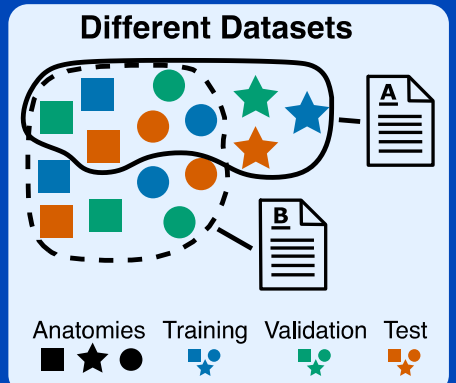
Flaws of Current Evaluation Protocols

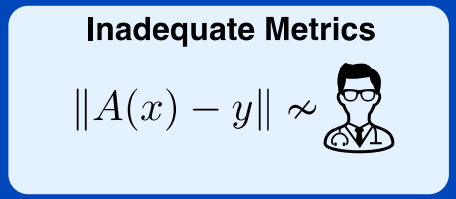
No open-source implementations: Increases chance of (un)intentional errors

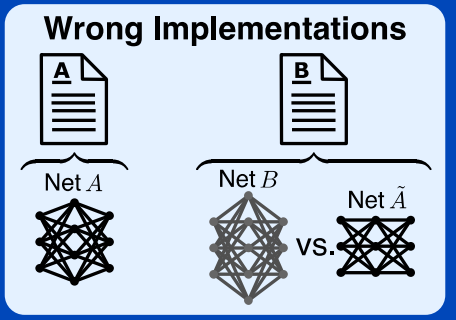
> *Meanwhile, to keep the reasonable model complexity, we reduced 96 filters to 32 filters in each layer.* [1]

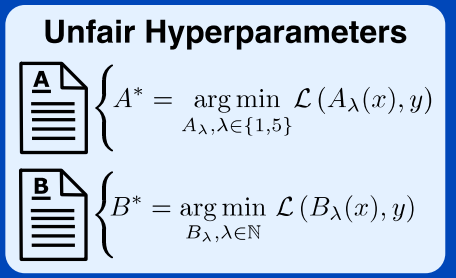
Inadequate metrics: Standard IQA metrics (MSE, SSIM, …) do not correlate well with human reader ratings [2]

Unfair hyperparameters: Either no HP optimization or limited to subset of parameters / methods. Often authors use HPs reported in reference publications
→ Problematic since no consensus dataset exists









[1] Fan, Fenglei, Hongming Shan, Mannudeep K. Kalra, Ramandeep Singh, Guhan Qian, Matthew Getzin, Yueyang Teng, Juergen Hahn, and Ge Wang. 2020. "Quadratic Autoencoder (Q-AE) for Low-Dose CT Denoising." *IEEE Transactions on Medical Imaging.*
[2] K. Ohashi, Y. Nagatani, M. Yoshigoe, K. Iwai, K. Tsuchiya, A. Hino, Y. Kida, A. Yamazaki, and T. Ishida, "Applicability evaluation of fullreference image quality assessment methods for computed tomography images," Journal of Imaging Informatics in Medicine, vol. 36, no. 6, pp. 2623–2634, Dec. 2023.

dkfz.

Received: 9 March 2024 | Revised: 5 August 2024 | Accepted: 17 August 2024

DOI: 10.1002/mp.17379

RESEARCH ARTICLE

MEDICAL PHYSICS

Benchmarking deep learning-based low-dose CT image denoising algorithms

Elias Eulig[1,2] | **Björn Ommer**[3] | **Marc Kachelrieß**[1,4]

[1]Division of X-Ray Imaging and Computed Tomography, German Cancer Research Center (DKFZ), Heidelberg, Germany

[2]Faculty of Physics and Astronomy, Heidelberg University, Heidelberg, Germany

[3]CompVis @ LMU Munich, MCML, Munich, Germany

[4]Medical Faculty Heidelberg, Heidelberg University, Heidelberg, Germany

Abstract

Background: Long-lasting efforts have been made to reduce radiation dose and thus the potential radiation risk to the patient for computed tomography (CT) acquisitions without severe deterioration of image quality. To this end, various techniques have been employed over the years including iterative reconstruction methods and noise reduction algorithms.

Purpose: Recently, deep learning-based methods for noise reduction became increasingly popular and a multitude of papers claim ever improving performance both quantitatively and qualitatively However the lack of a standardized

LDCT Benchmark

- **Algorithms used for our benchmark:**
  - CNN-10 (2017)
  - RED-CNN (2017)
  - ResNet (2018)
  - WGAN-VGG (2017)
  - QAE (2019)
  - DU-GAN (2021)
  - TransCT (2021)
  - Bilateral (2022)

Standard CNNs trained with pixelwise losses

CNNs trained with adversarial losses

Specialized architectures trained with pixelwise losses

- **All tested methods**
  - do the same hyperparameter optimization
  - use the same train/validation set
  - were evaluated on the same test set

github.com/eeulig/ldct-benchmark

E. Eulig, B. Ommer, and M. Kachelrieß. Benchmarking deep learning-based low-dose CT image denoising algorithms. Med. Phys. 51(12):8776-8788, December 2024.

dkfz.

Benchmark Setup

- **Hyperparameter optimization**
  - Rigorous HP optimization including the weighting factors of loss function terms
  - 50 iterations of sequential model-based optimization (SMBO) using Gaussian processes and expected improvement as acquisition function
  - As metric to optimize we use SSIM on validation dataset
- **Retrain each method 10 times with**
  - optimal HPs and
  - different random seeds

	Parameter	Prior
All algorithms	Learning rate	$\log\mathcal{U}(1\times10^{-5}, 0.01)$
	Maximum iterations	$\mathcal{U}(1\times10^{3}, 1\times10^{5})$
	Mini-batch size	$\mathcal{U}(2, 128)$
CNN-10 (2017)	Patchsize	$\mathcal{U}(32, 128)$
RED-CNN (2017)	Patchsize	$\mathcal{U}(32, 128)$
WGAN-VGG (2017)	β_1 of Adam	$\mathcal{U}(0.3, 0.9)$
	Loss weight: $\lambda_{\text{perceptual}}$	$\mathcal{U}(0, 1)$
	Critic updates	$\mathcal{U}(1, 5)$
	Patchsize	$\mathcal{U}(32, 128)$
ResNet (2018)	Patchsize	$\mathcal{U}(32, 128)$
QAE (2019)	Patchsize	$\mathcal{U}(32, 128)$
DU-GAN (2021)	β_1 of Adam	$\mathcal{U}(0.3, 0.9)$
	Cutmix warmup	$\mathcal{U}(0, 1\times10^{4})$
	Loss weight: λ_{adv}	$\mathcal{U}(0, 1)$
	Loss weight: λ_{CM}	$\mathcal{U}(0, 10)$
	Loss weight: $\lambda_{\text{px,grad}}$	$\mathcal{U}(0, 40)$
	Critic updates	$\mathcal{U}(1, 5)$
	Patchsize	$\mathcal{U}(32, 128)$
TransCT (2021)	-	-
Bilateral (2022)	Learning rate for σ_r	$\log\mathcal{U}(1\times10^{-5}, 0.01)$
	Patchsize	$\mathcal{U}(32, 128)$
	Initalization for σ_r	$\mathcal{U}(0, 1)$
	Initalization for $\sigma_{x,y}$	$\mathcal{U}(0, 1)$

$\mathcal{U}$: Uniform distribution; $\log\mathcal{U}$: Log-uniform distribution

E. Eulig, B. Ommer, and M. Kachelrieß. Benchmarking deep learning-based low-dose CT image denoising algorithms. Med. Phys. 51(12):8776-8788, December 2024.

dkfz.

Benchmark Setup

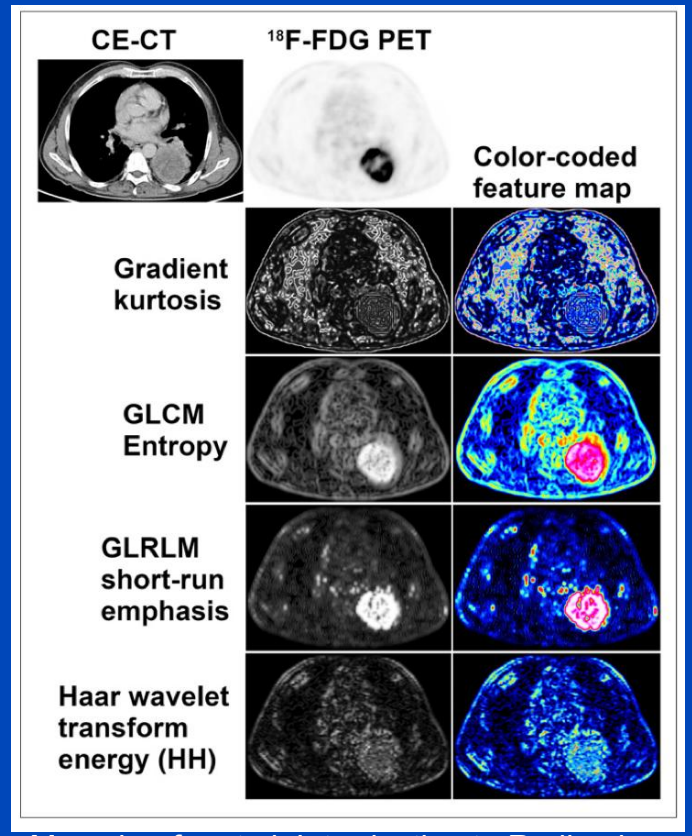
Metrics

- Standard metrics: SSIM, PSNR, Visual information fidelity (VIF)
- Clinically relevant image properties: Radiomic feature similarity (RFS)
  1. Automatically segment organs in high-dose scan (s)
  2. Compare features on high-dose scan with those on denoised scans

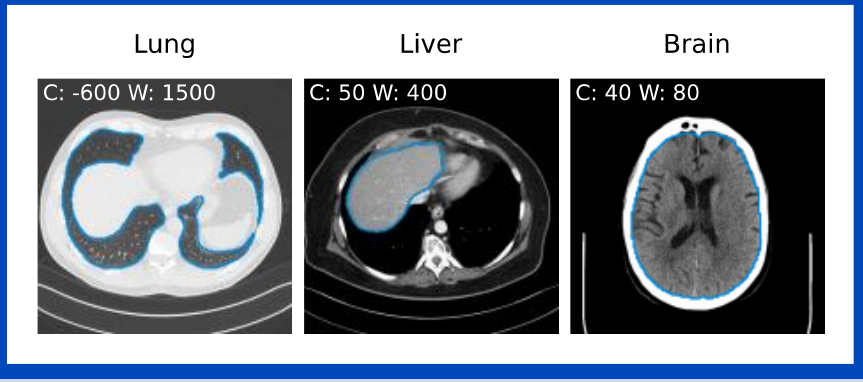
$$\mathrm{RFS}_i^{(s)} = \cos\left(r_i^{(s)}, r_{\mathrm{GT}}^{(s)}\right), \quad r_i^{(s)} = \underbrace{\left(R_{i,1}^{(s)}, \ldots, R_{i,J}^{(s)}\right)}_{\text{Normalized radiomic features for some algorithm } i \text{ on scan } s}$$

$j \in \{1, 2, \ldots, J\}$: Radiomic features

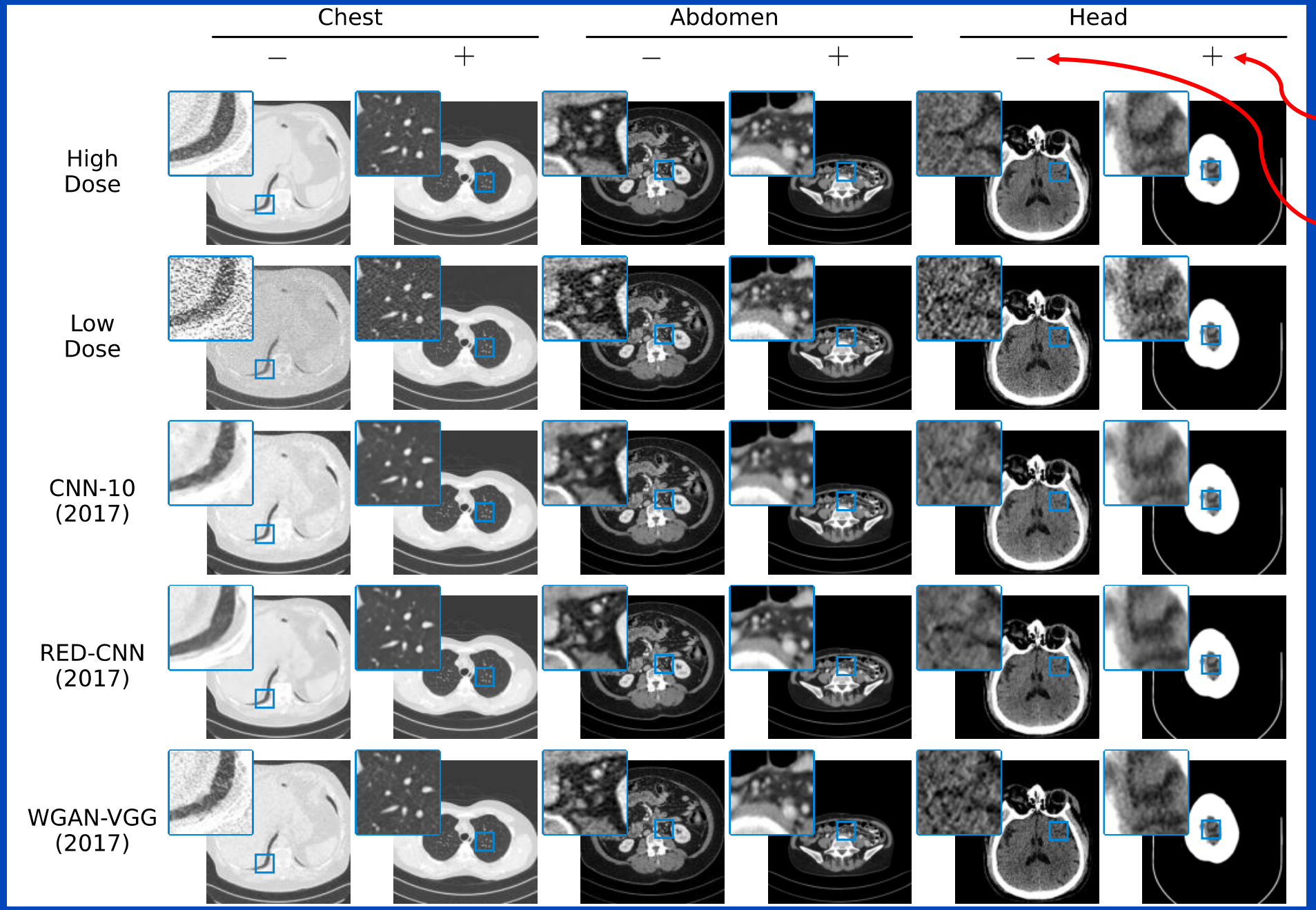
$i \in \{1, 2, \ldots, N\}$: Algorithms



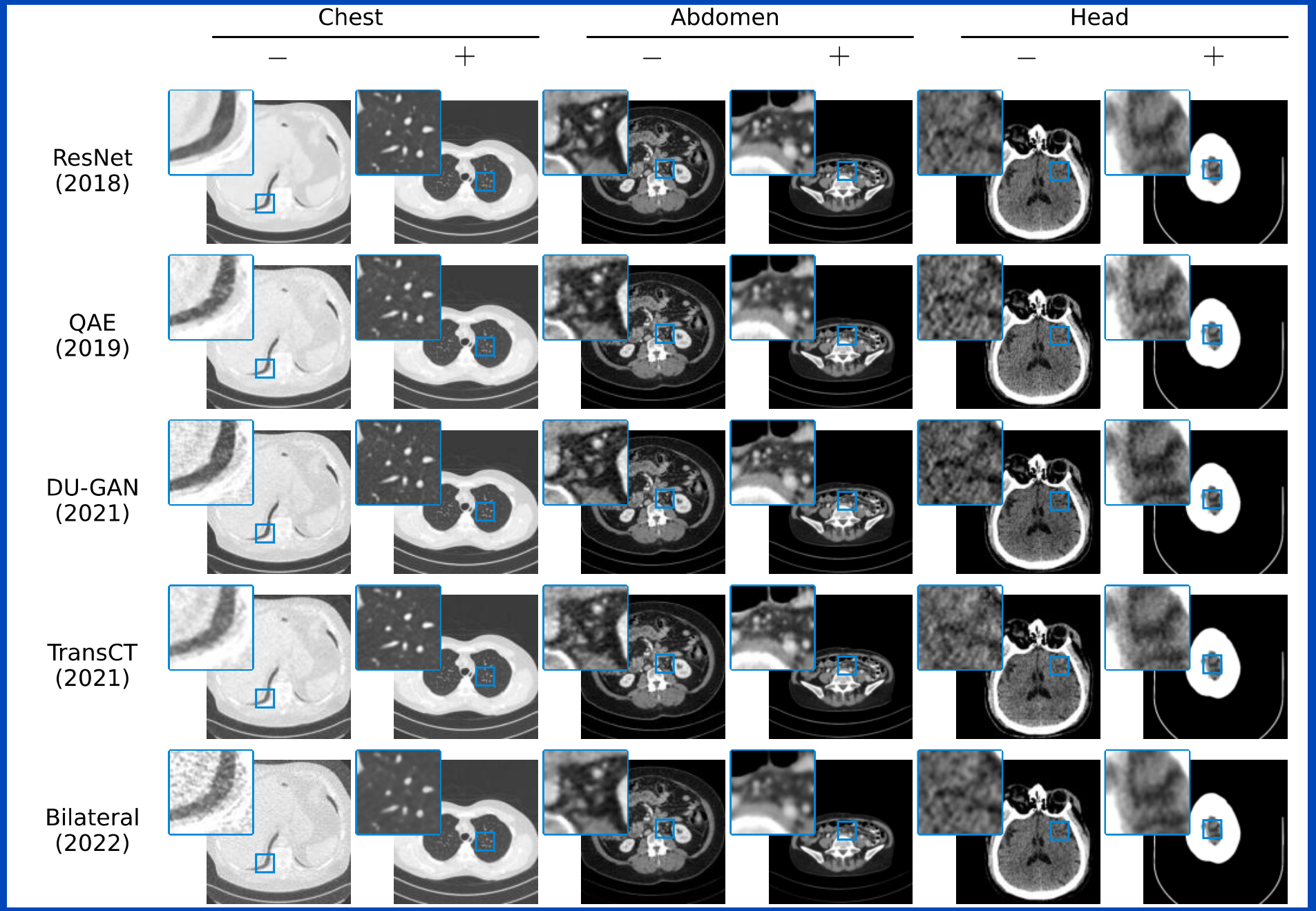
Mayerhoefer et al. Introduction to Radiomics. J. Nucl. Med. 61(4):488–495, 2020.



dkfz.



E. Eulig, B. Ommer, and M. Kachelrieß. Benchmarking deep learning-based low-dose CT image denoising algorithms. Med. Phys. 51(12):8776-8788, December 2024.



E. Eulig, B. Ommer, and M. Kachelrieß. Benchmarking deep learning-based low-dose CT image denoising algorithms. Med. Phys. 51(12):8776-8788, December 2024.

Image Quality Metrics

Given reference image *x* and test image *y*, *N* pixels each.

Peak signal-to-noise ratio (PSNR)

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

$$\mathrm{PSNR}(x,y) = 20\log\left(\frac{I_{\max}}{\mathrm{RMSE}(x,y)}\right)$$

Visual information fidelity (VIF)

Compare information *I* extracted by a human visual system (HVS) model of the test image *y* with that of the reference image *x*.

$$\mathrm{VIF}(x,y) = \frac{\sum_k I(G_k * x)}{\sum_k I(G_k * y)}$$

G_k are Gaussians of different scale

Structural similarity index measure (SSIM)

$$\mathrm{SSIM}(u,v) = \frac{(2\mu_u\mu_v + C_1)(2\sigma_{uv} + C_2)}{(\mu_u^2 + \mu_v^2 + C_1)(\sigma_u^2 + \sigma_v^2 + C_2)}$$

$$\mathrm{SSIM}(x,y) = \underset{u\in x, v\in y}{\mathrm{mean}}\ \mathrm{SSIM}(u,v)$$

μ_u : Mean of u $\quad$ σ_u^2 : Variance of u

μ_v : Mean of v $\quad$ σ_v^2 : Variance of v

σ_{uv} : Covariance of u and v

} in a sliding 7×7 window

Radiomic feature similarity (RFS)

1. Extract radiomic features R_x and R_y from segmentations in *x* and *y*
2. Compute cosine similarity between R_x and R_y

$$\mathrm{RFS}(x,y) = \frac{R_x \cdot R_y}{\|R_x\|\,\|R_y\|}$$

dkfz.

Quantitative Results

PSNR units are decibel (dB)	Head (25% dose)				Chest (10% dose)				Abdomen (25% dose)			
	SSIM	PSNR	VIF	RFS	SSIM	PSNR	VIF	RFS	SSIM	PSNR	VIF	RFS
Low dose scan	26.40	0.55	0.71	0.34	18.77	0.09	0.70	0.84	28.67	0.34	0.75	0.88
CNN-10 (2017)	28.86	0.62	0.94	0.59	27.71	0.19	0.80	0.90	32.39	0.45	0.88	0.90
RED-CNN (2017)	30.41	0.69	0.95	0.61	28.36	0.22	0.76	0.90	33.22	0.49	0.80	0.90
WGAN-VGG (2017)	25.36	0.53	0.86	0.51	25.54	0.15	0.98	0.88	30.51	0.38	0.92	0.88
ResNet (2018)	29.64	0.67	0.91	0.61	28.42	0.22	0.75	0.90	33.15	0.49	0.79	0.90
QAE (2019)	28.51	0.59	0.95	0.58	27.62	0.19	0.83	0.89	32.02	0.42	0.96	0.90
DU-GAN (2021)	28.76	0.62	0.94	0.57	26.68	0.17	0.96	0.89	32.13	0.43	0.97	0.90
TransCT (2021)	24.65	0.44	0.88	0.56	26.99	0.17	0.83	0.88	30.53	0.37	0.92	0.85
Bilateral (2022)	26.60	0.50	0.87	0.55	25.59	0.16	0.64	0.86	27.13	0.36	0.87	0.87

Green numbers indicate that a method is significantly better than the previously published best method.
Orange numbers indicate that it is significantly worse.

E. Eulig, B. Ommer, and M. Kachelrieß. Benchmarking deep learning-based low-dose CT image denoising algorithms. Med. Phys. 51(12):8776-8788, December 2024.

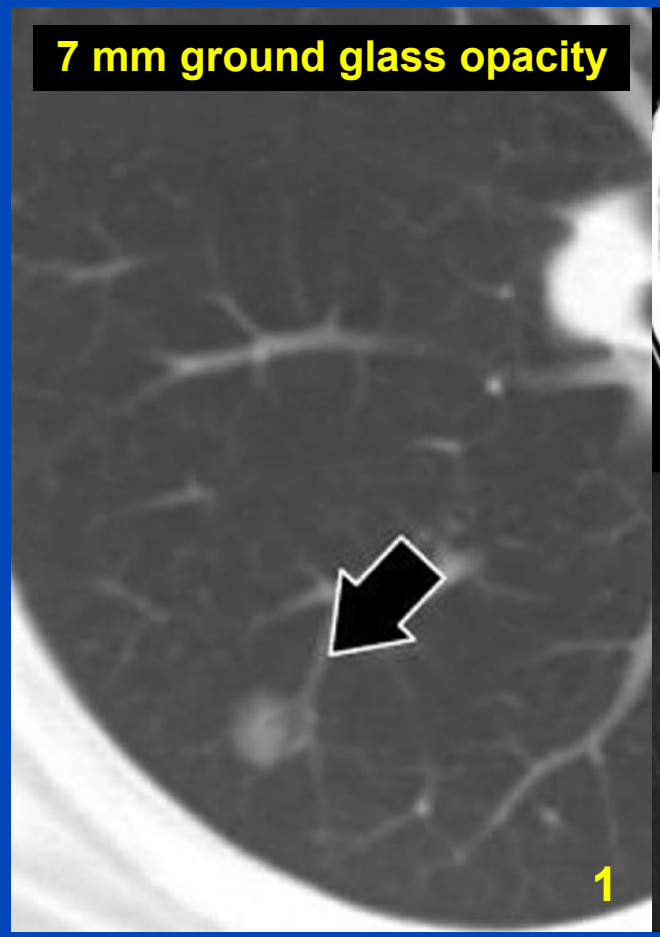
dkfz.

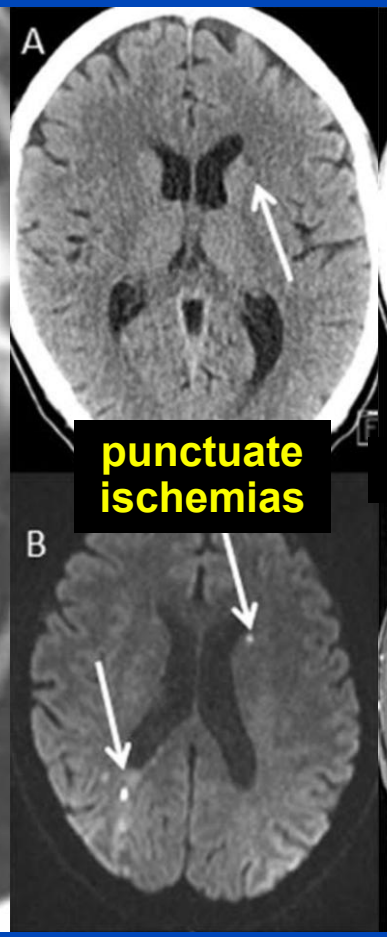
Minimize pitfall: Let small structures be just as important as large structures

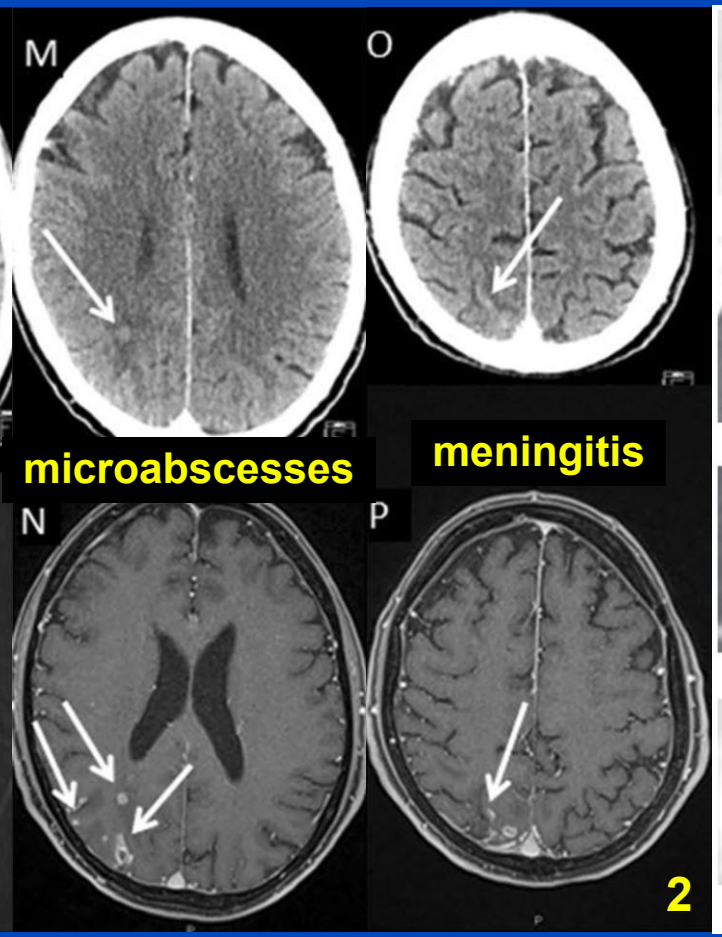
A NEW METRIC FOR SUBTLE DETAILS

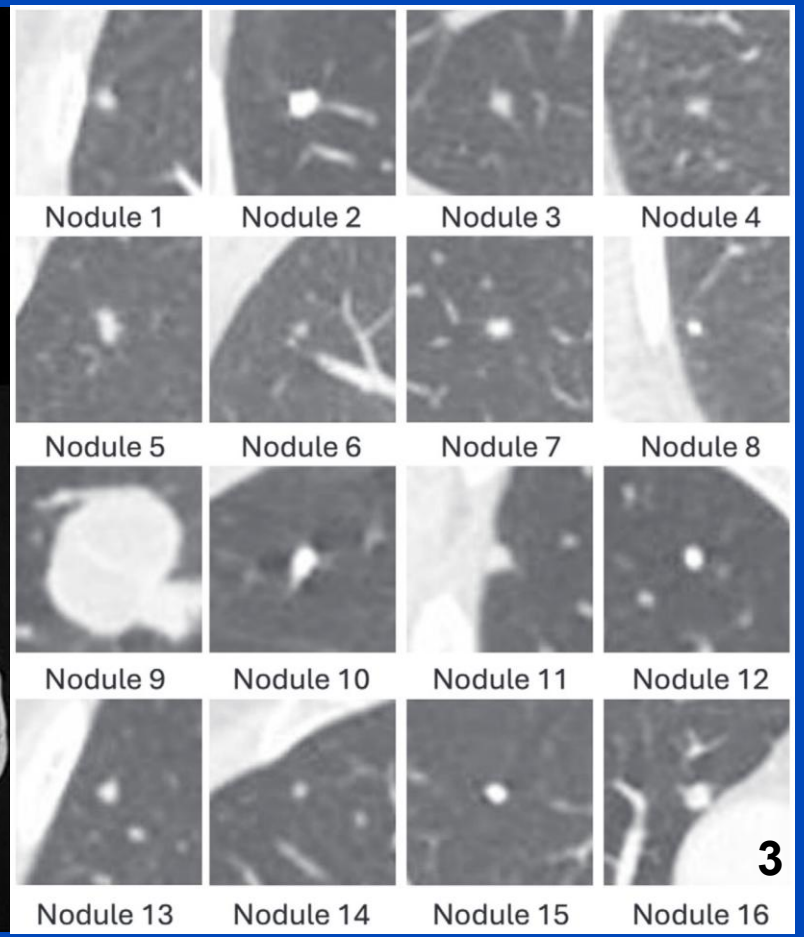
Motivation

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







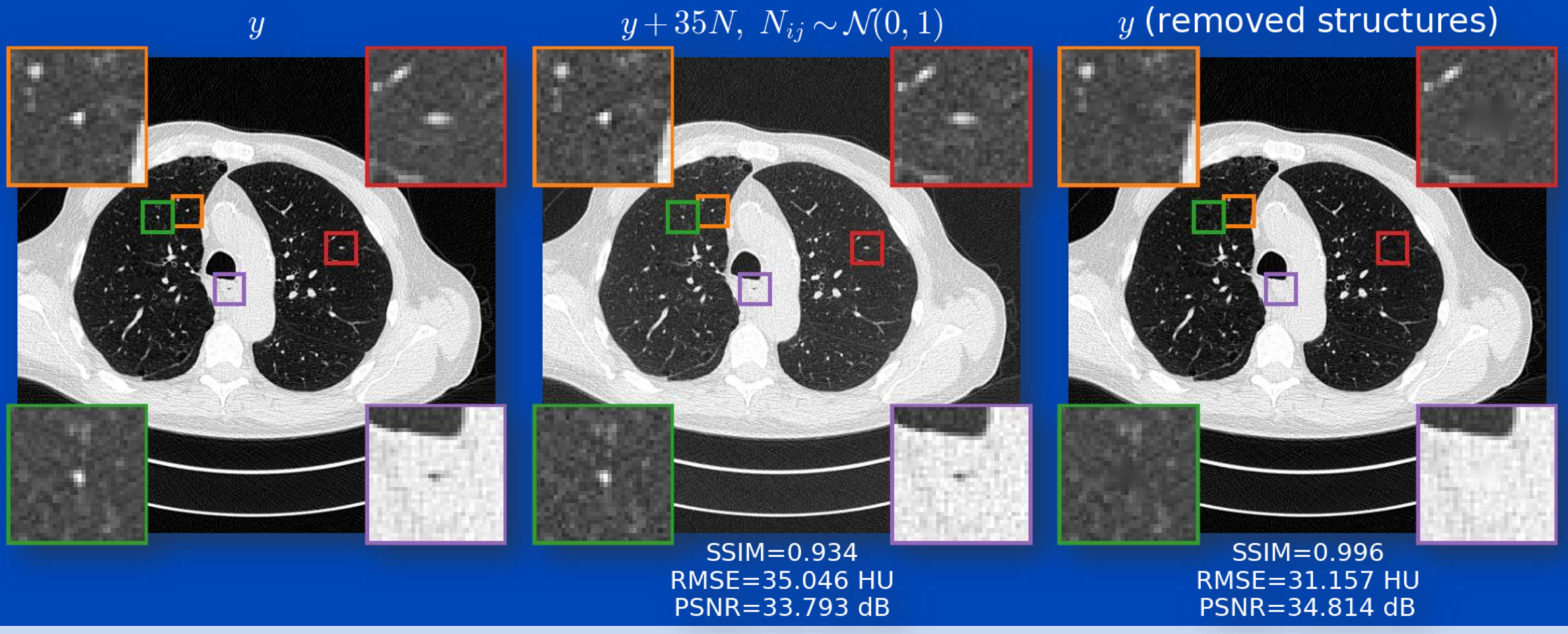


[1]H. K. Kim *et al.*, "Management of Multiple Pure Ground-Glass Opacity Lesions in Patients with Bronchioloalveolar Carcinoma," *Journal of Thoracic Oncology*, vol. 5, no. 2, 2010.
[2]P. Vitali *et al.*, "MRI versus CT in the detection of brain lesions in patients with infective endocarditis before 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.

dkfz.

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.

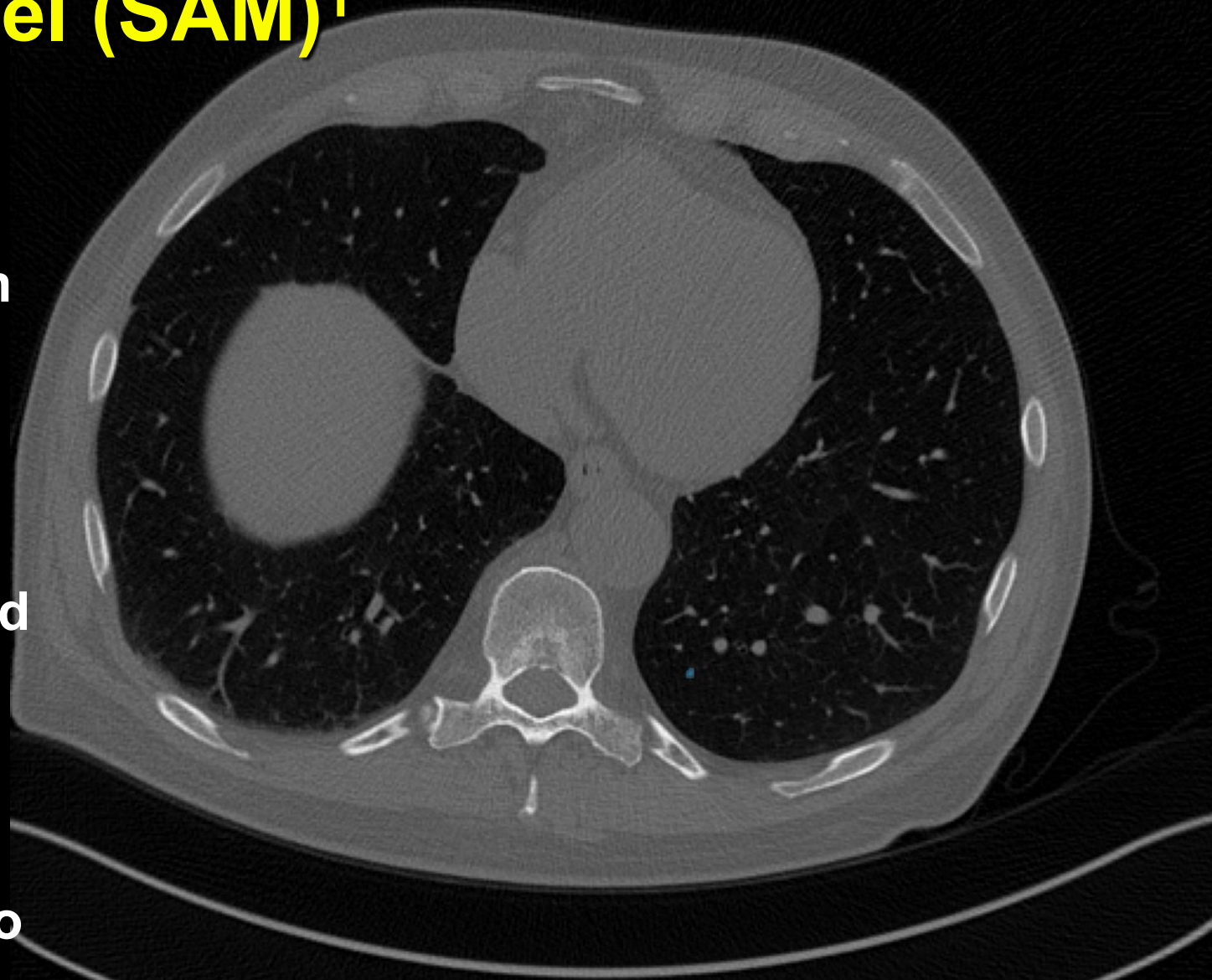


*SSIM also accounts in parts for the human visual system by using luminance, contrast and structure to estimate perceptual quality.

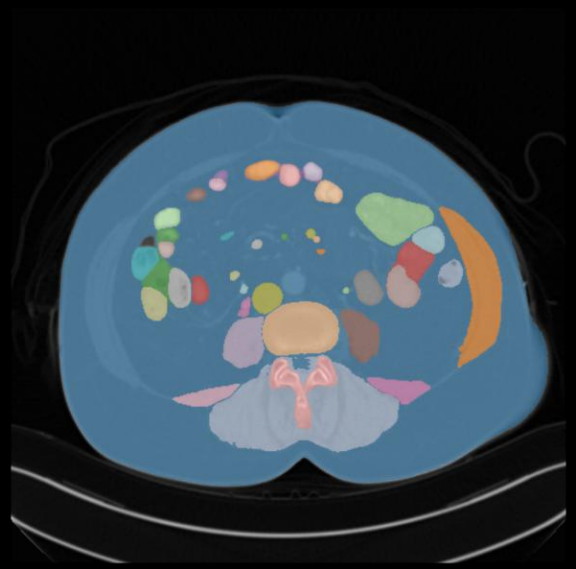
dkfz.

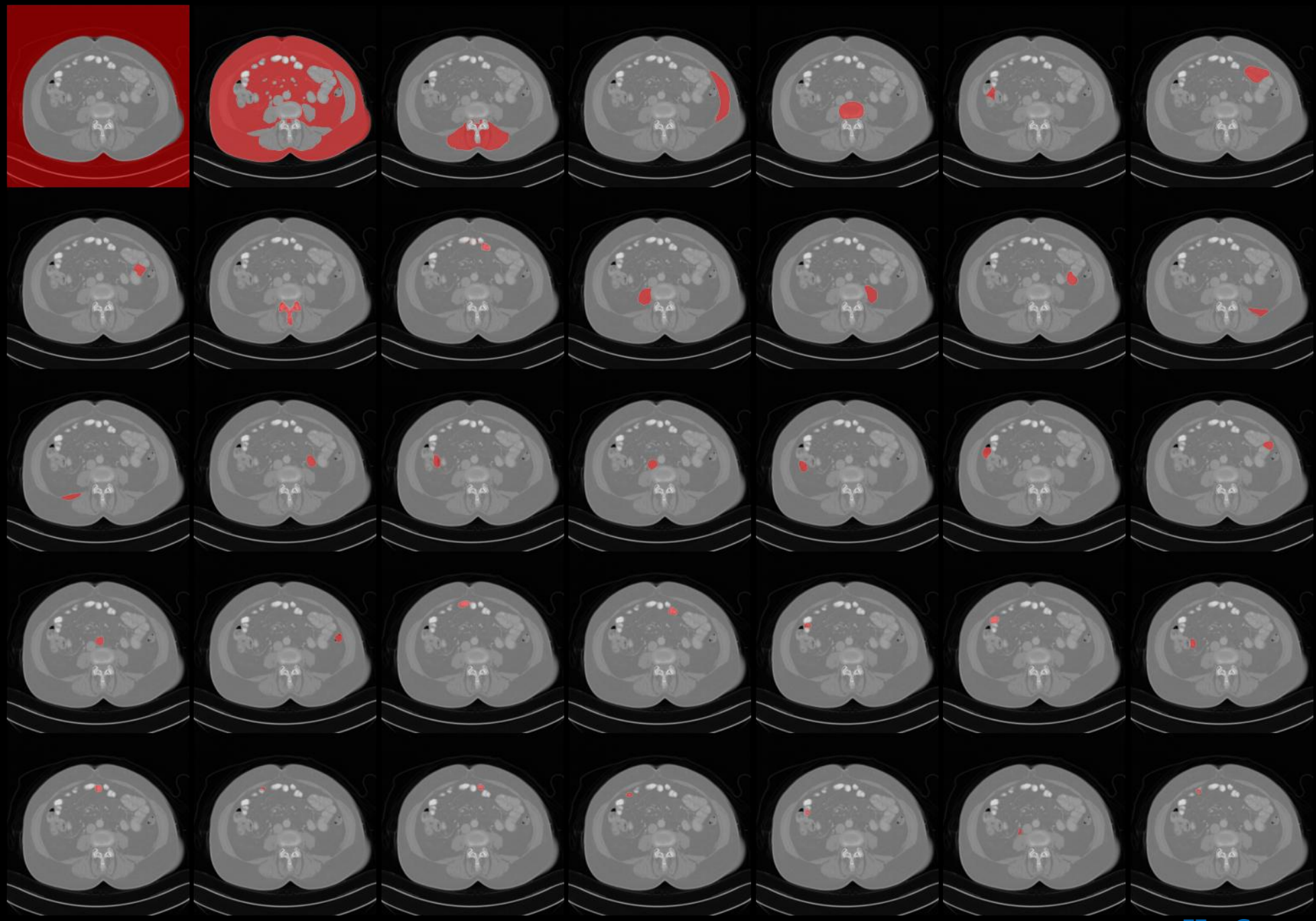
Segment Anything Model (SAM)[1]

- 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]
- To make our metric work, we need to apply SAM
  - to the ground truth image
  - or to the denoised image
  - or to the noisy image.
- In the following, SAM is applied to the GT image.

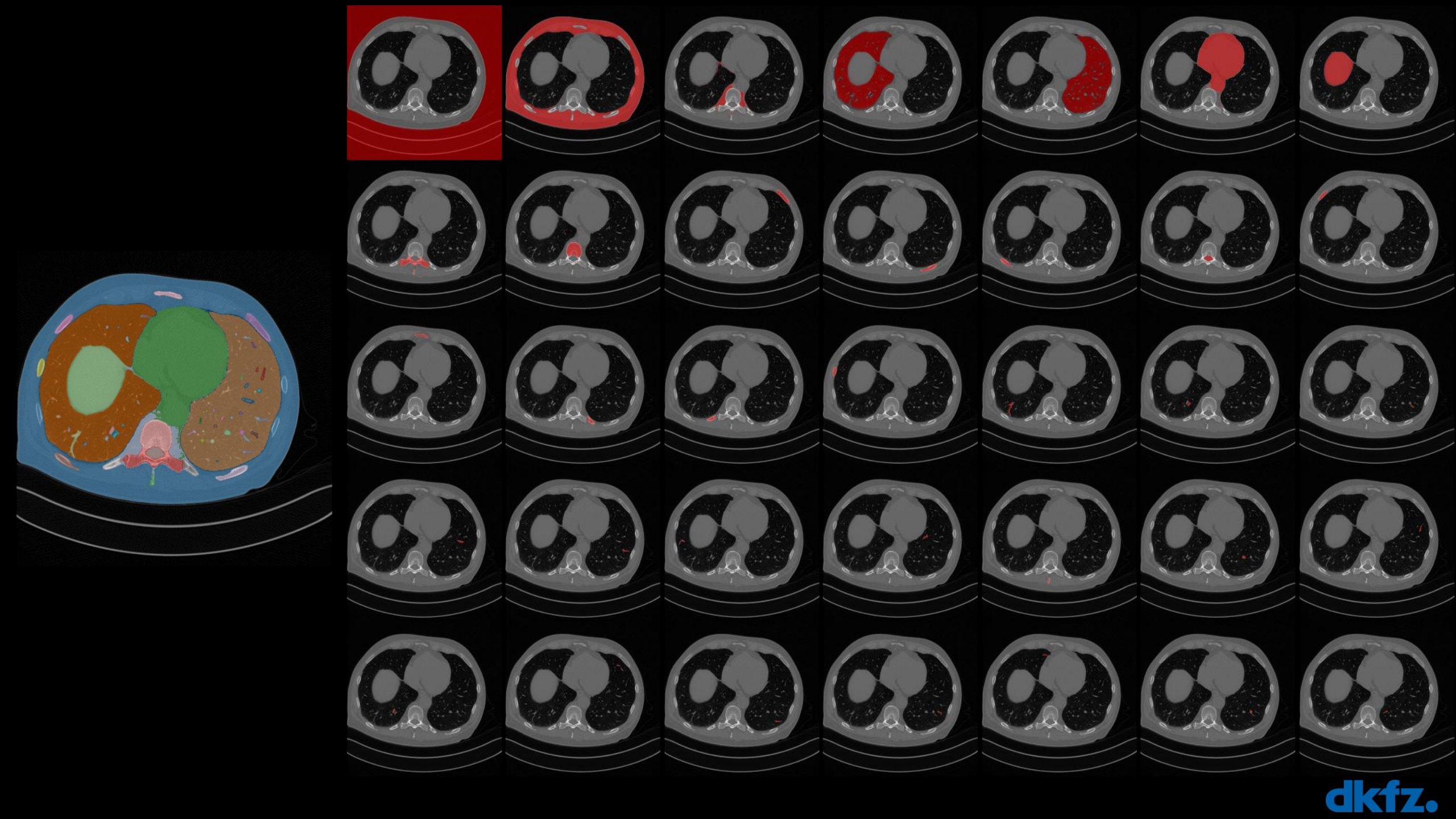


[1]Kirillov, Alexander, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, et al. 2023. "Segment Anything." arXiv.
[2]Ma, Jun, Yuting He, Feifei Li, Lin Han, Chenyu You, and Bo Wang. 2024. "Segment Anything in Medical Images." *Nature Communications* 15 (1): 654.





dkfz.


dkfz.

Segment RMSE (SRMSE)

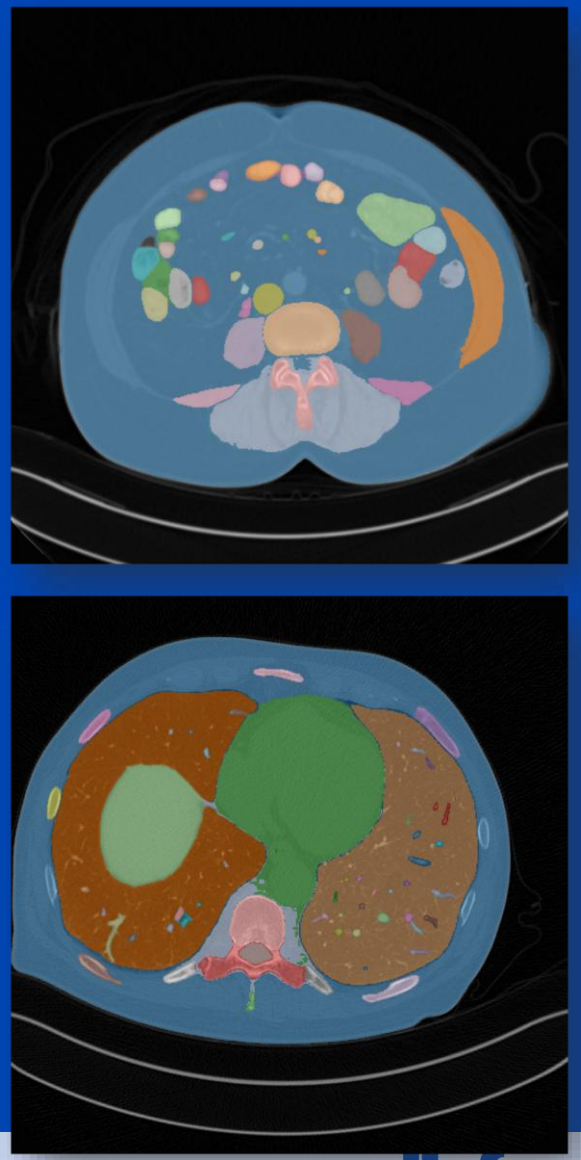
- Assume *S* segments from SAM for a given patient volume.
- Represent each segment by a binary mask volume $m^{(s)} \in \{0, 1\}^N$ with *N* being the number of voxels.
- Define the segment-wise root mean square error between two images *x* and *y*, and segment *s*:

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

- Using the set of all SRMSEs, define

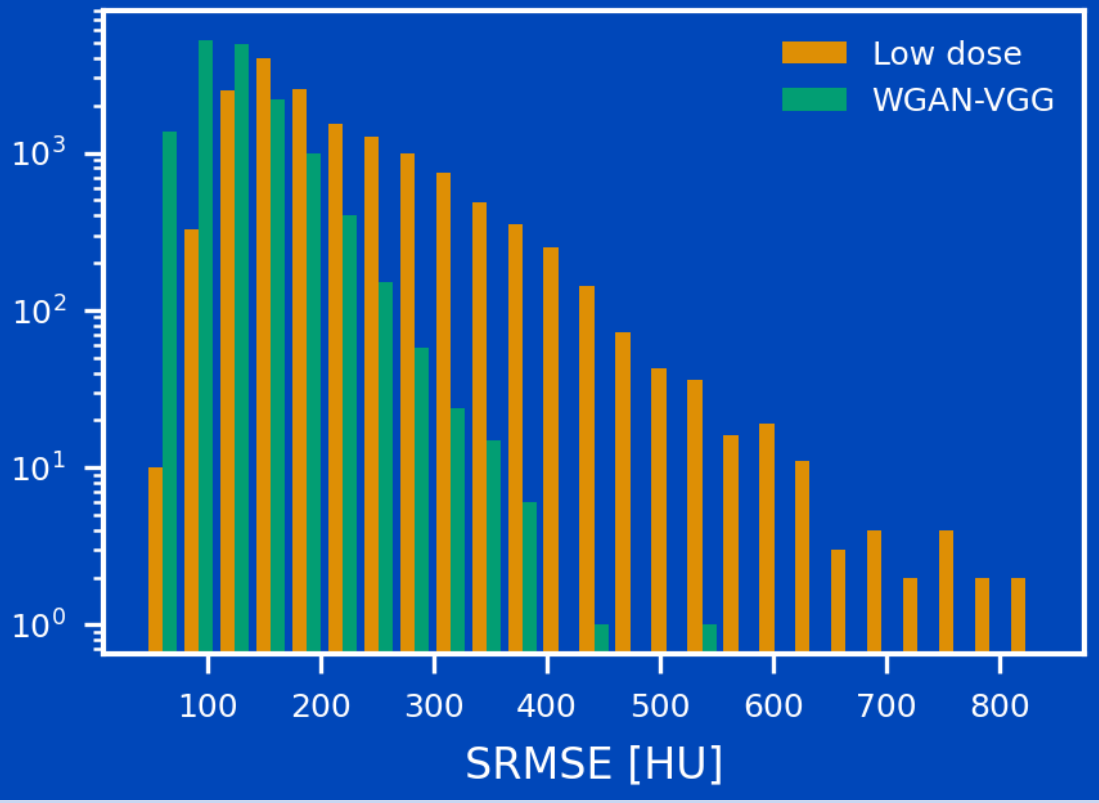
$$\mathrm{MeanSRMSE}(x, y) = \frac{1}{S} \sum_{s} \mathrm{SRMSE}(x, y; s)$$

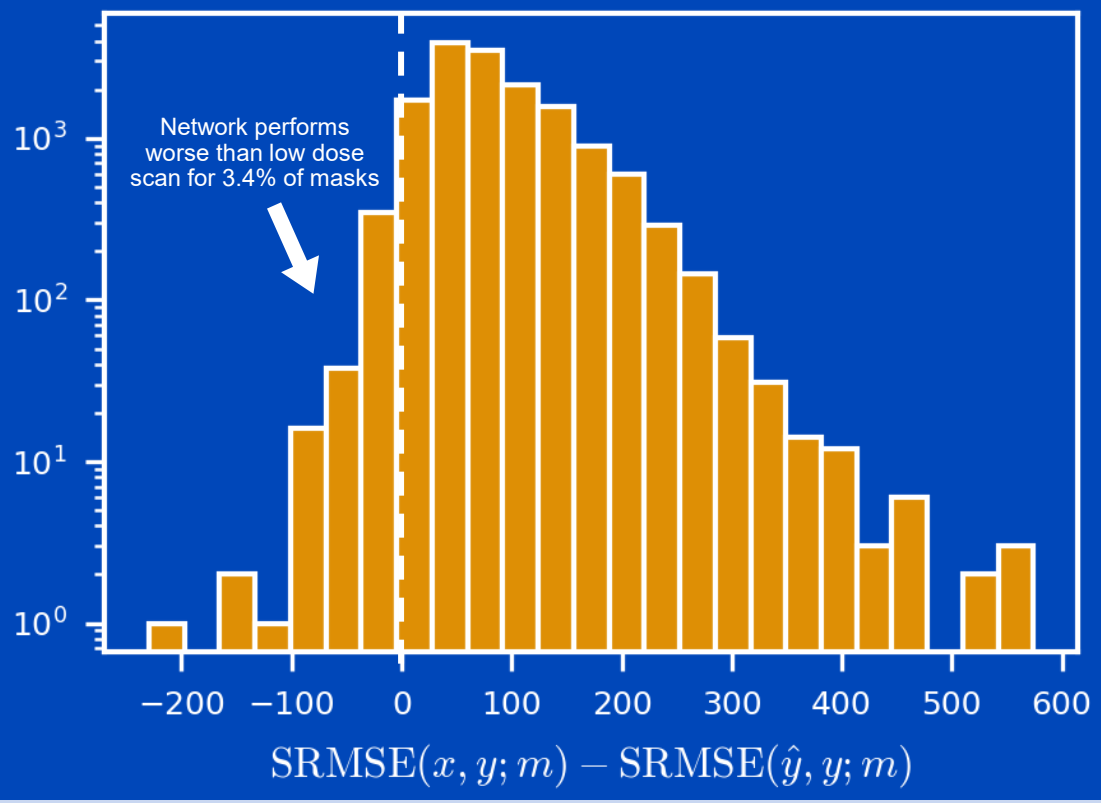
$$\mathrm{MaxSRMSE}(x, y) = \max_{s} \mathrm{SRMSE}(x, y; s)$$



Detecting Hallucinations

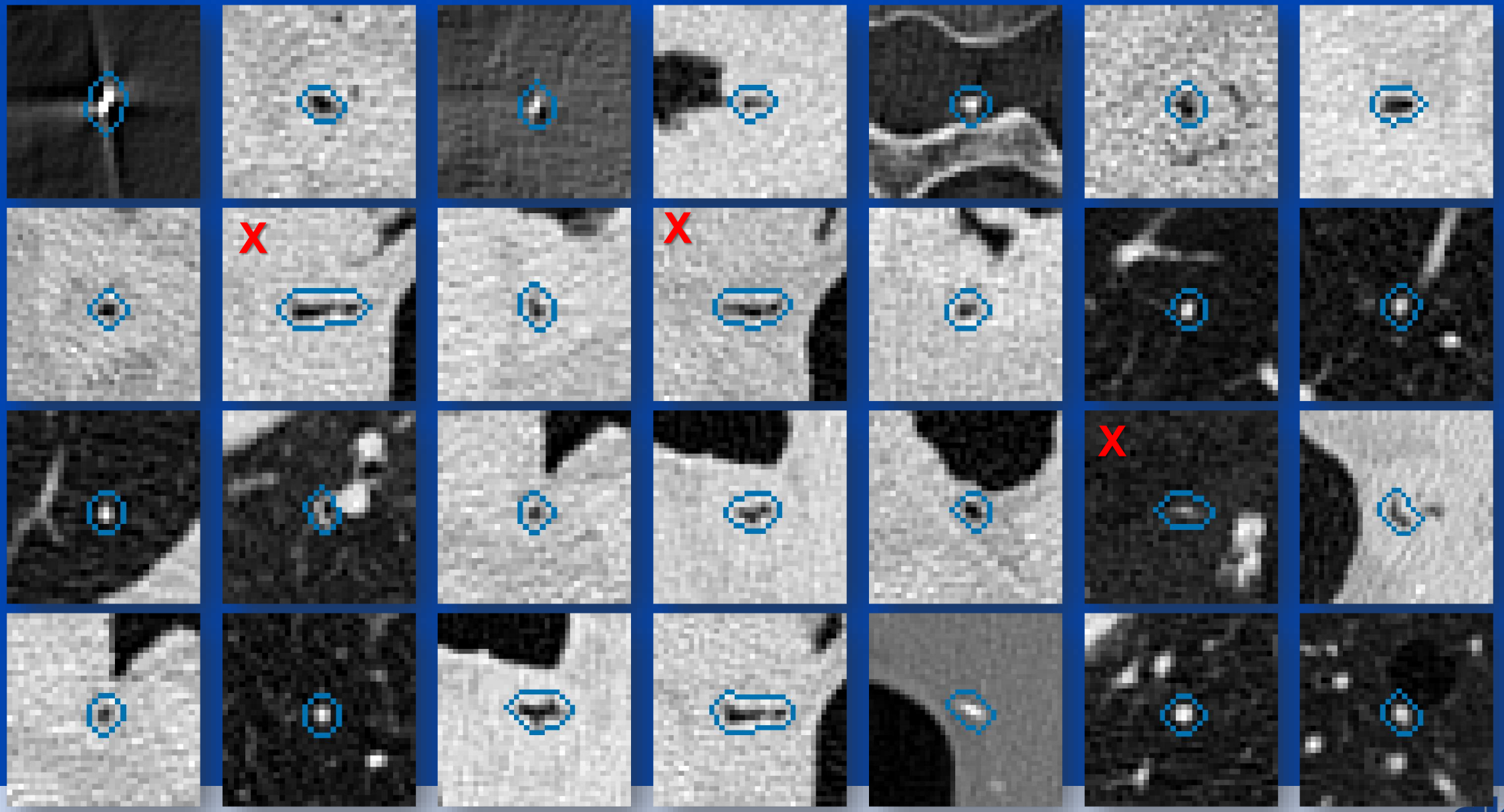
- Compare SRMSE of low dose scan (x) with network prediction ($\hat{y}$)
- On a chest scan with 392 axial slices we have a total of 15,547 masks





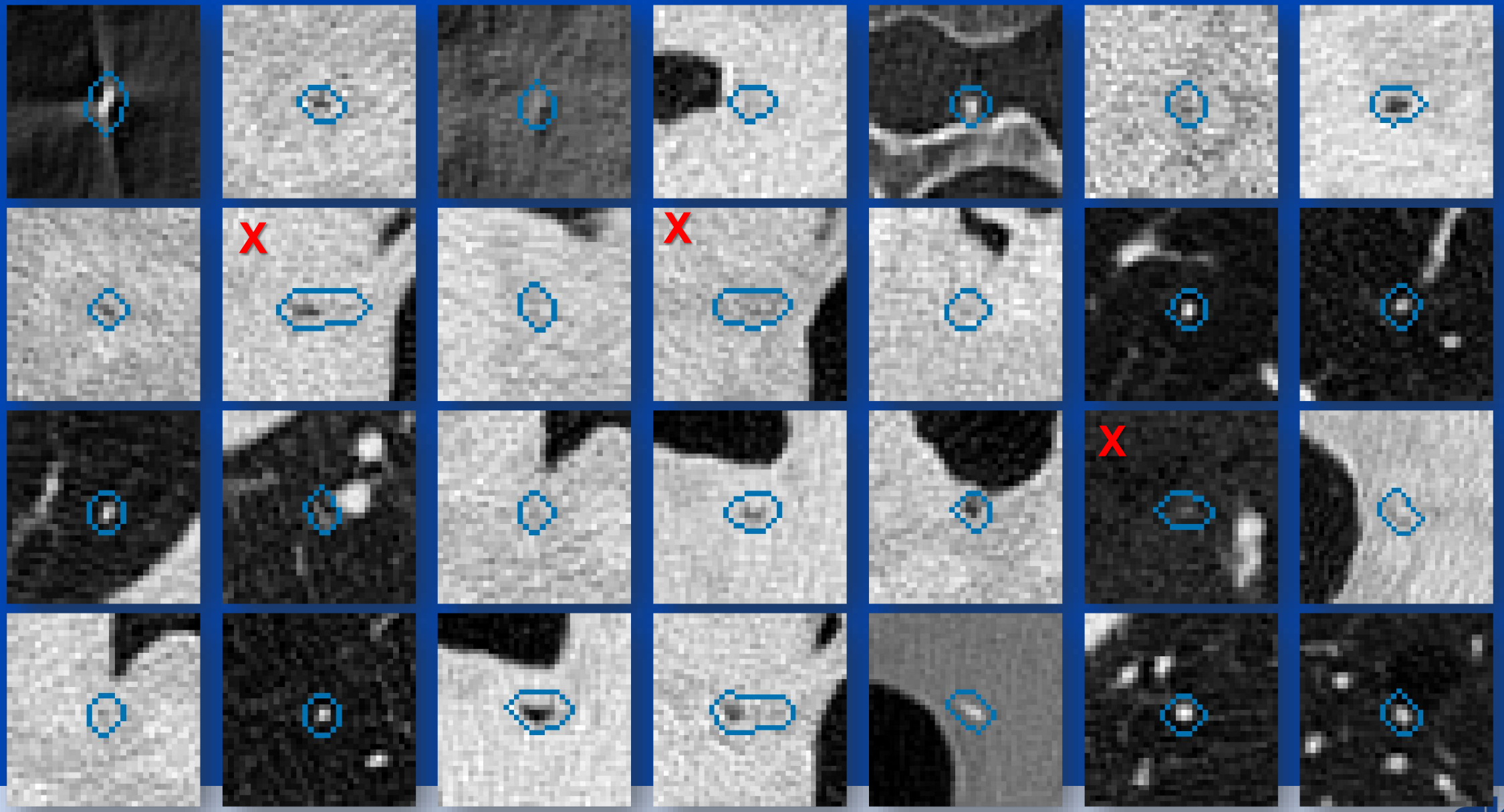
dkfz.

High Dose Images



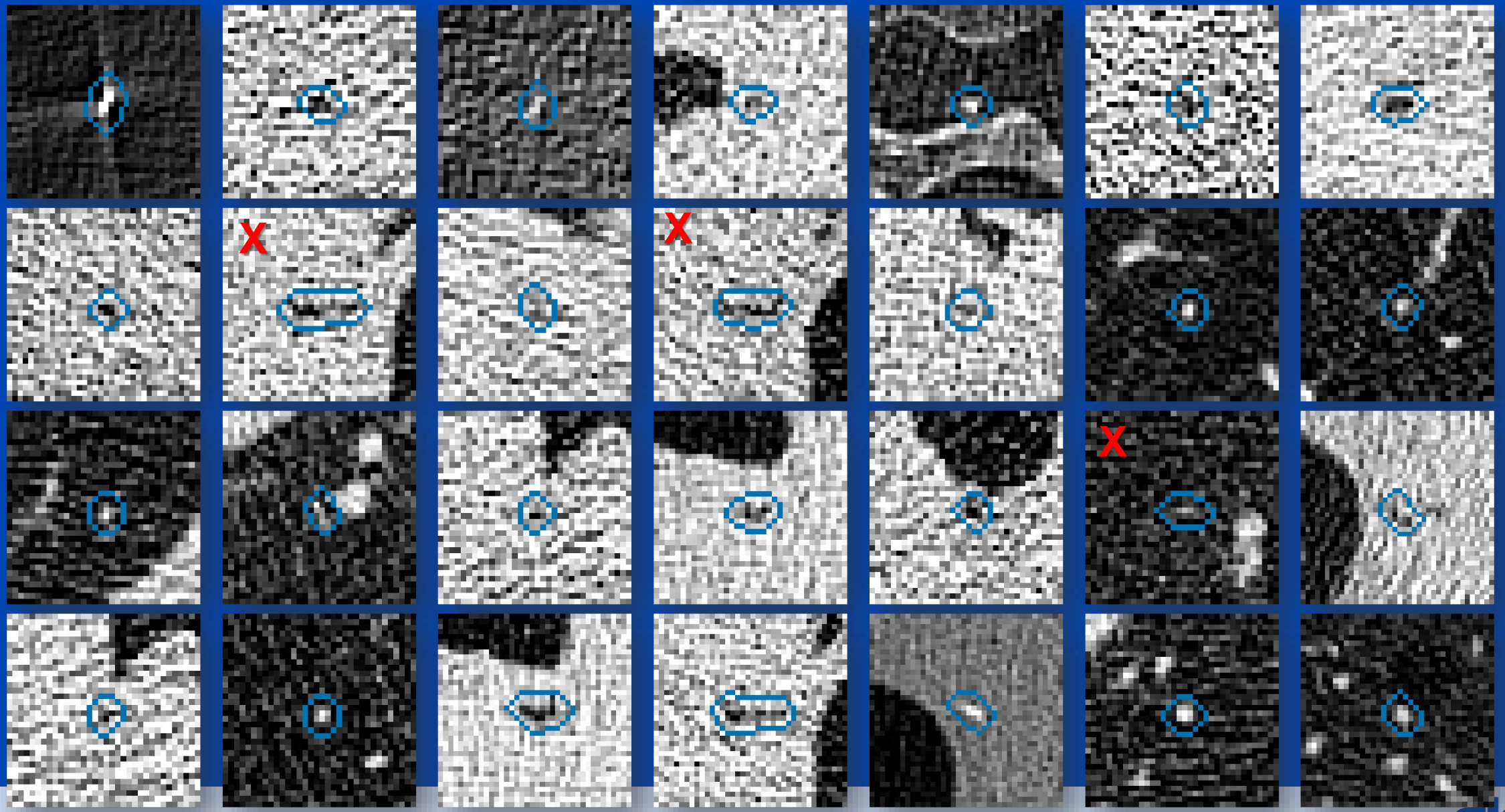
dkfz.

Network Predictions (WGAN-VGG)



dkfz.

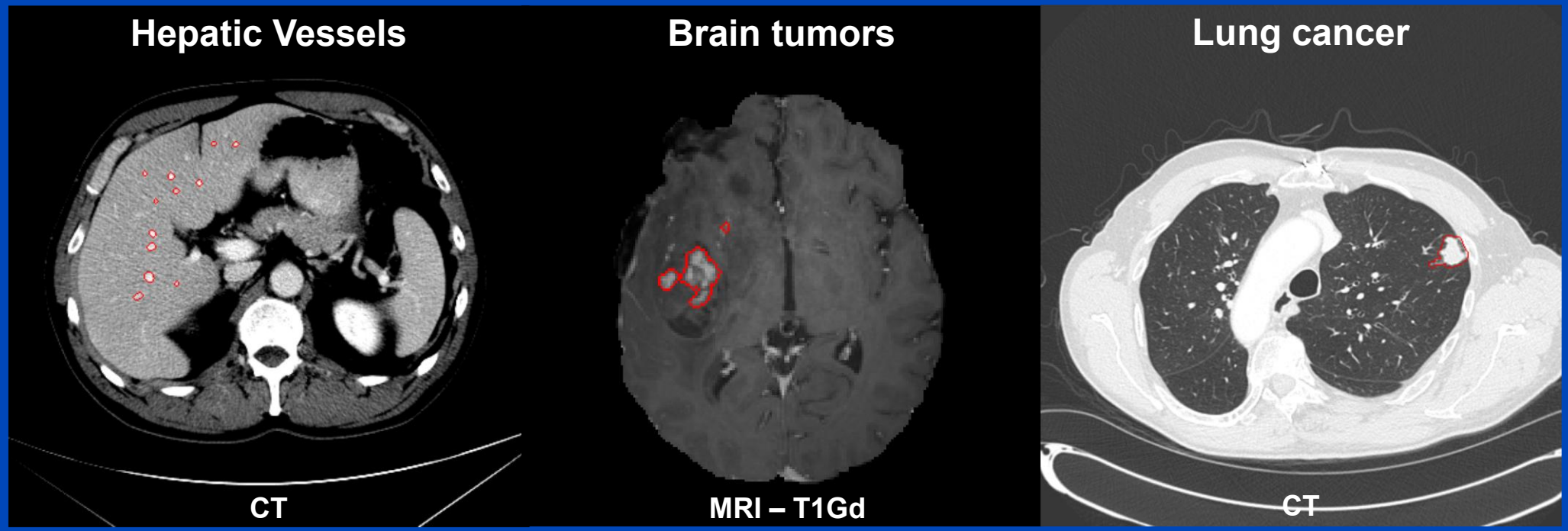
Low Dose Images



dkfz.

Experiments

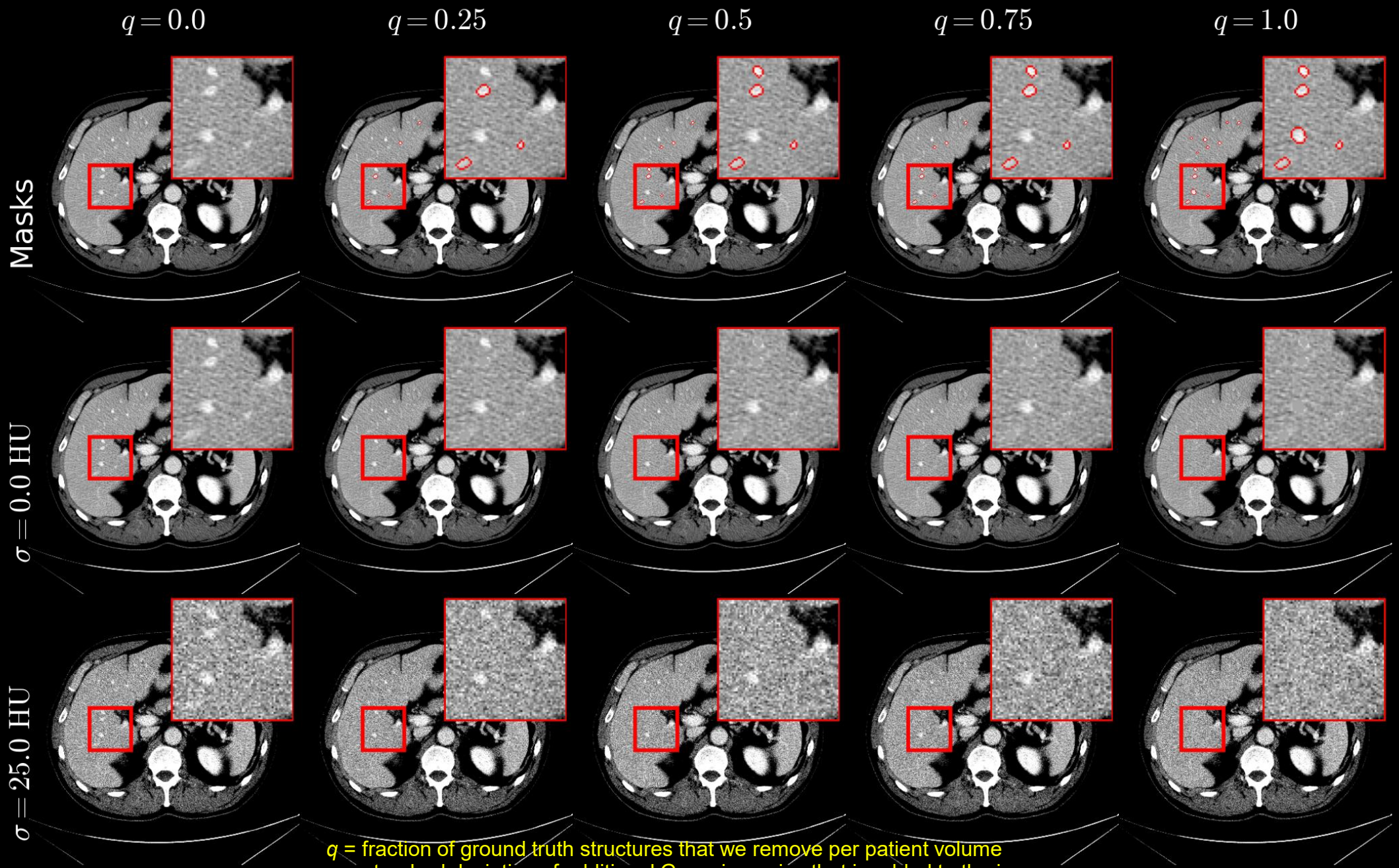
- Evaluate the proposed metric on synthetic datasets where the amount of removed structures is known.
- Utilize three datasets from the Medical Decathlon[1], a collection of ten medical image segmentation tasks with ground truth annotations.



[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.

dkfz.

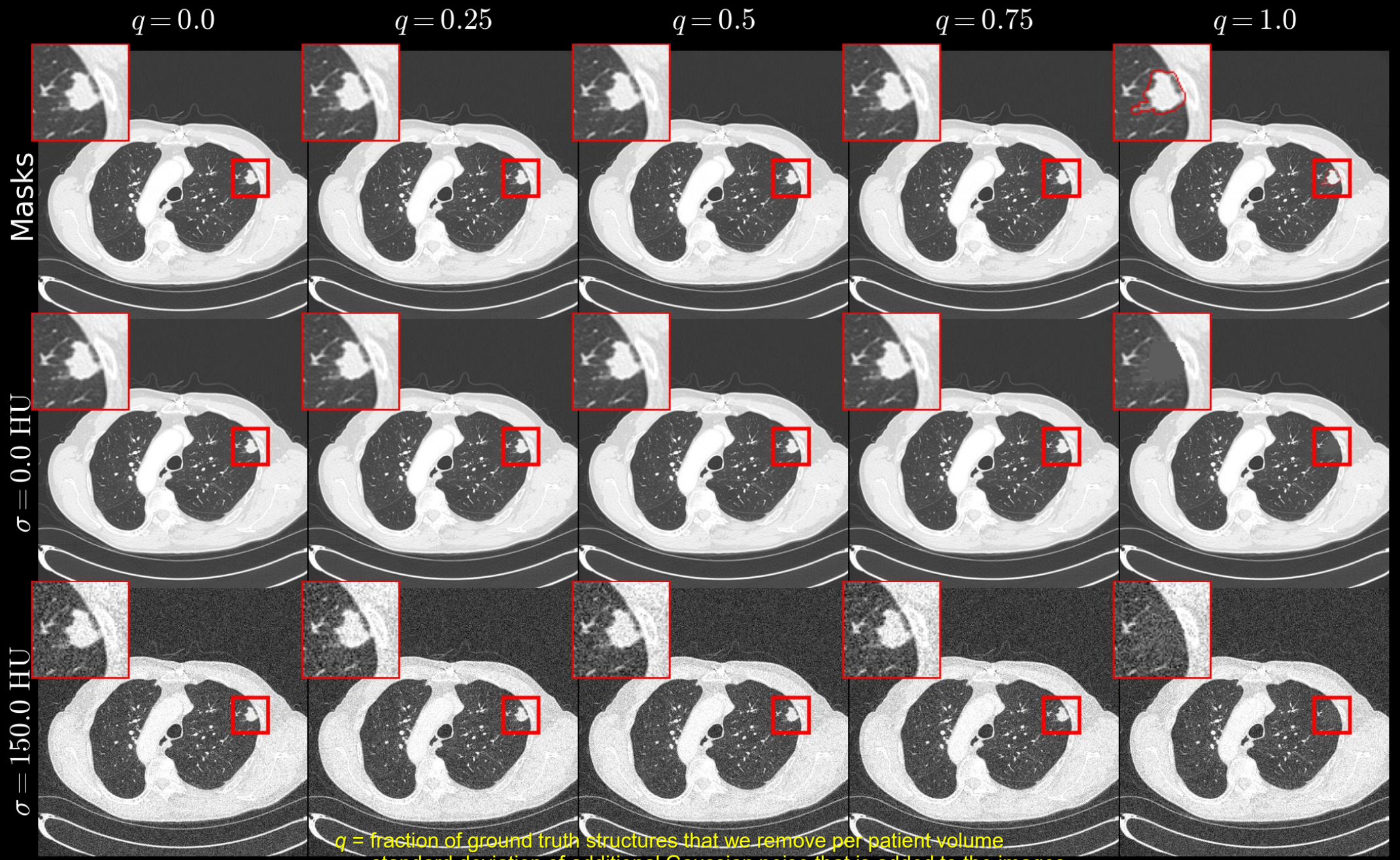
Hepatic Vessels



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

lkfz.

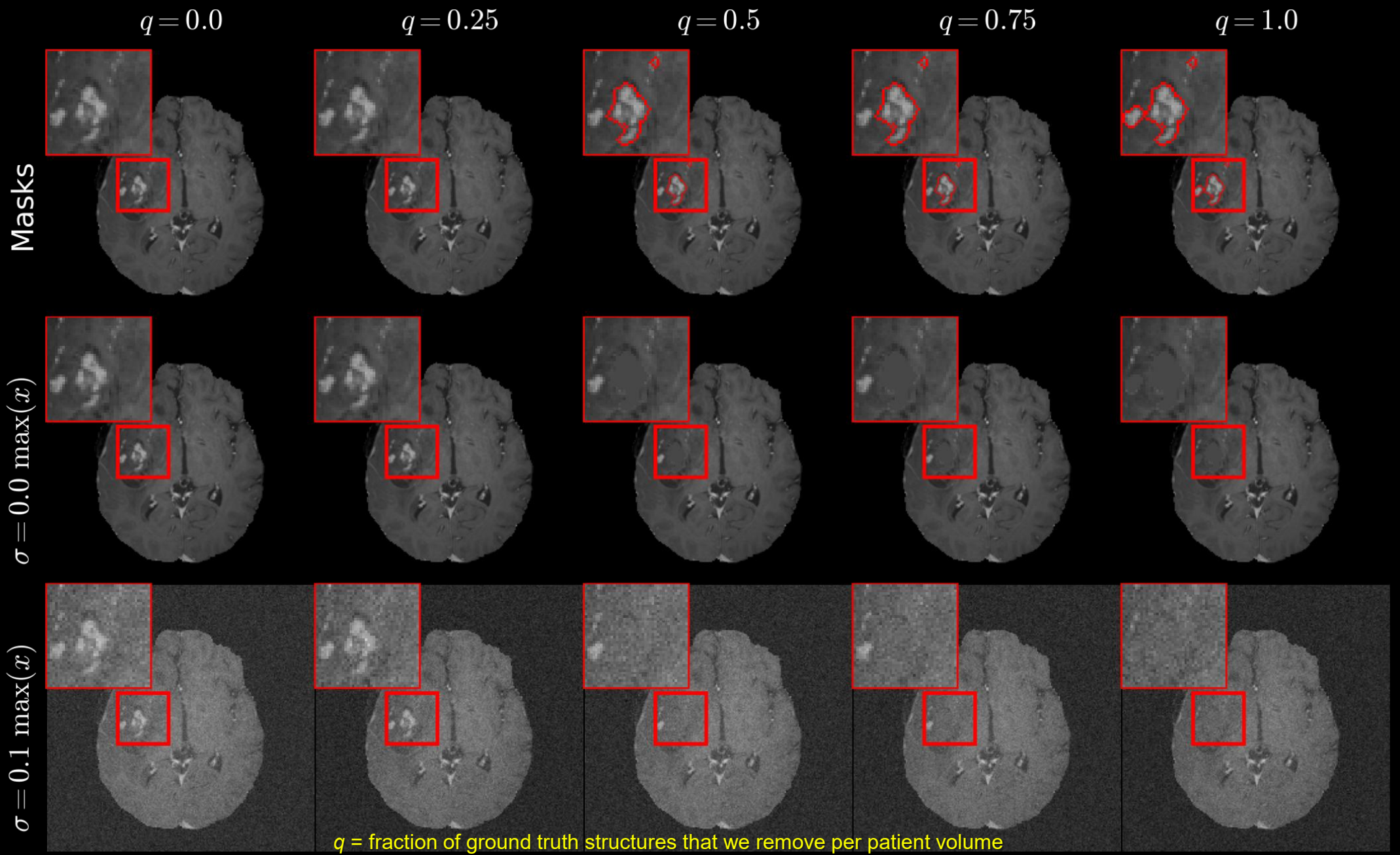
Lung Cancer



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

lkfz.

Brain Tumor

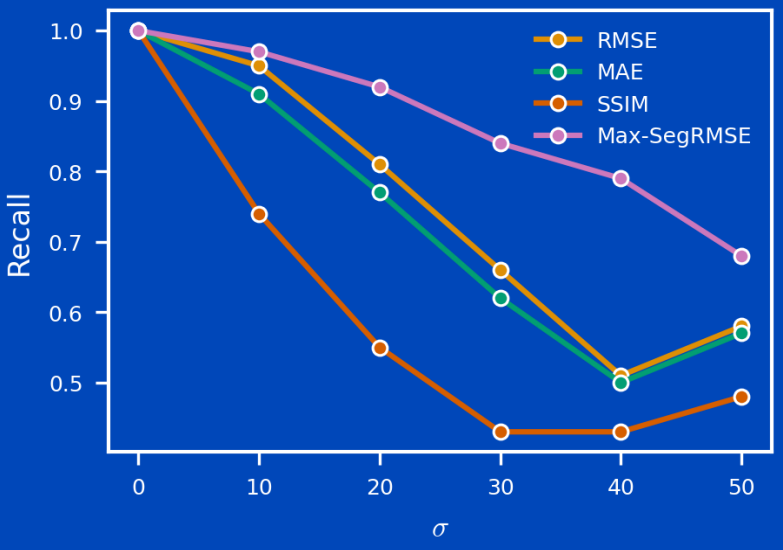


q = fraction of ground truth structures that we remove per patient volume
σ = standard deviation of additional Gaussian noise that is added to the images

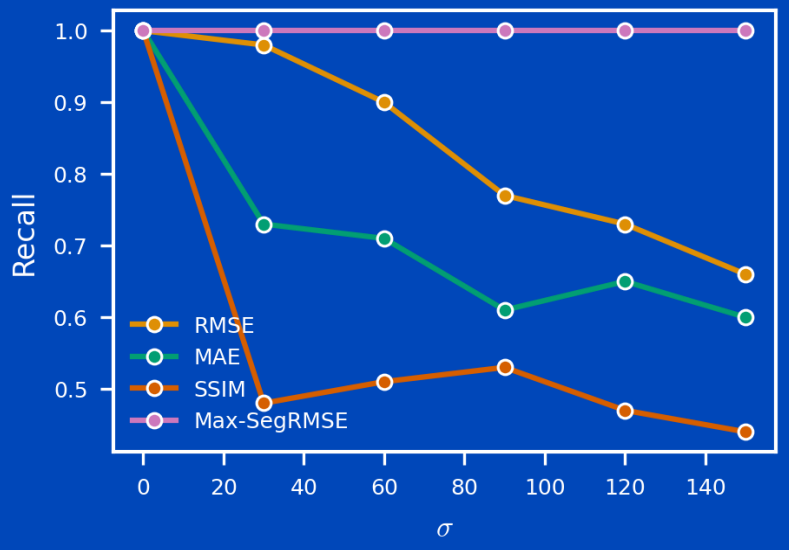
lkfz.

Results: True Positive Fraction (Recall)

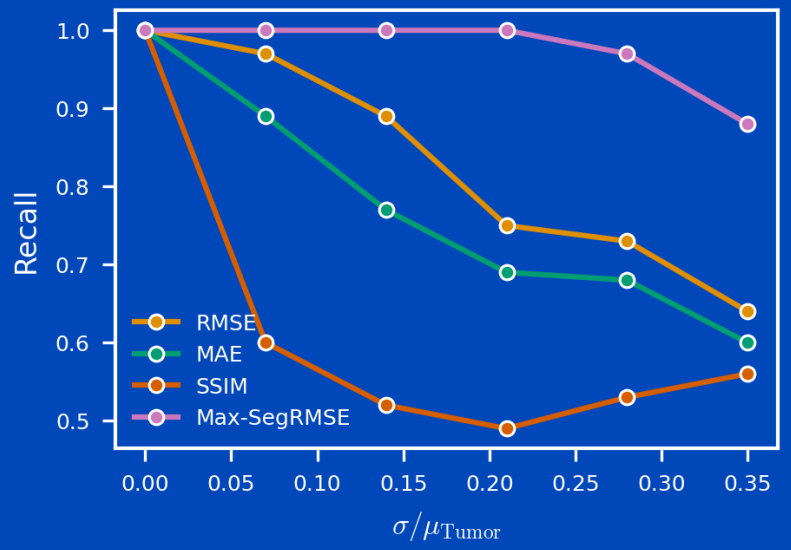
Hepatic Vessels



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.

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
Small perturbations in the input that lead to large alterations in predictions

⟷

Invariances
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

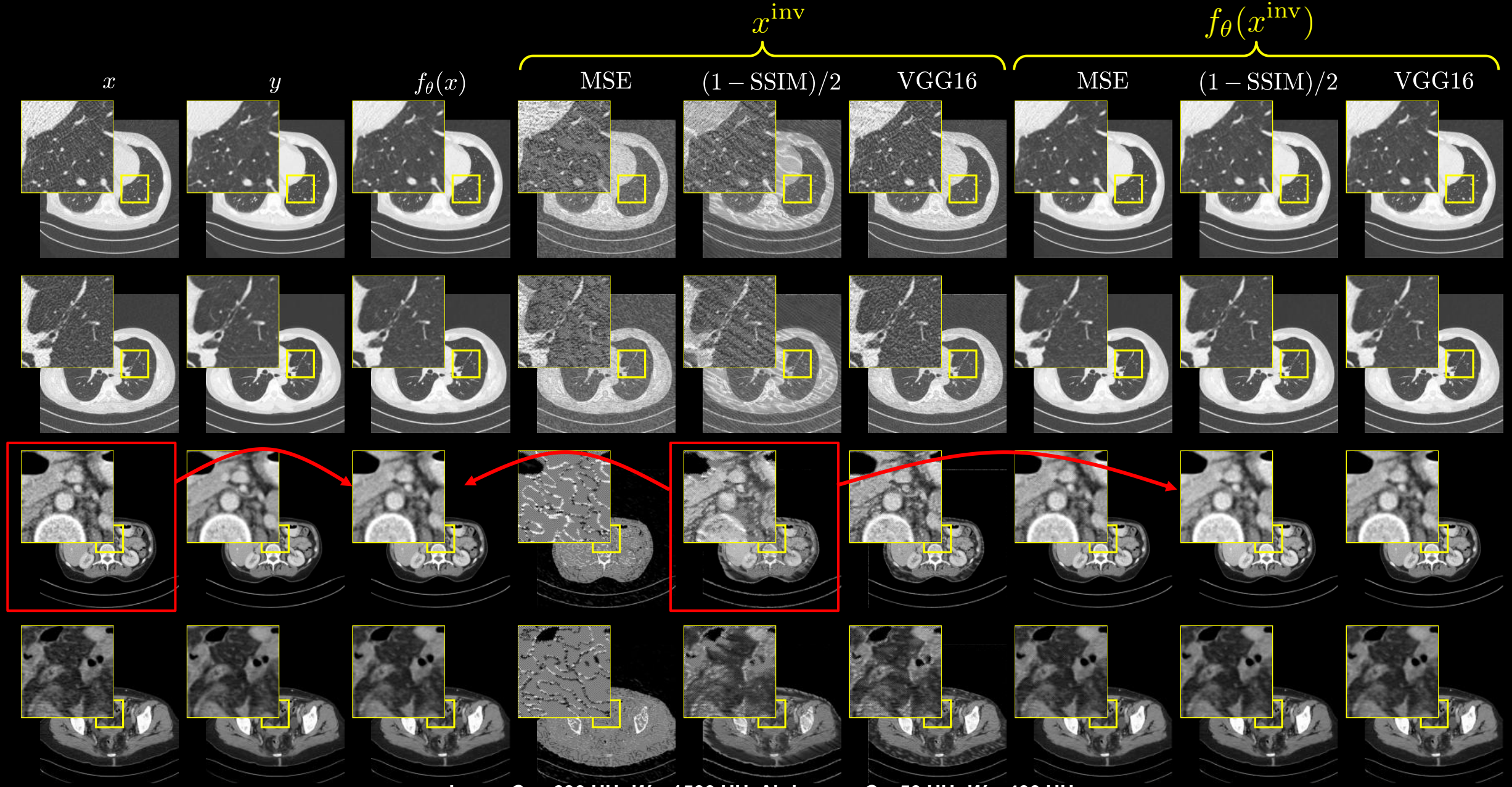
Elias Eulig[a,b,*], Joscha Maier[a], Björn Ommer[c], and Marc Kachelrieß[a,d]

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

ABSTRACT

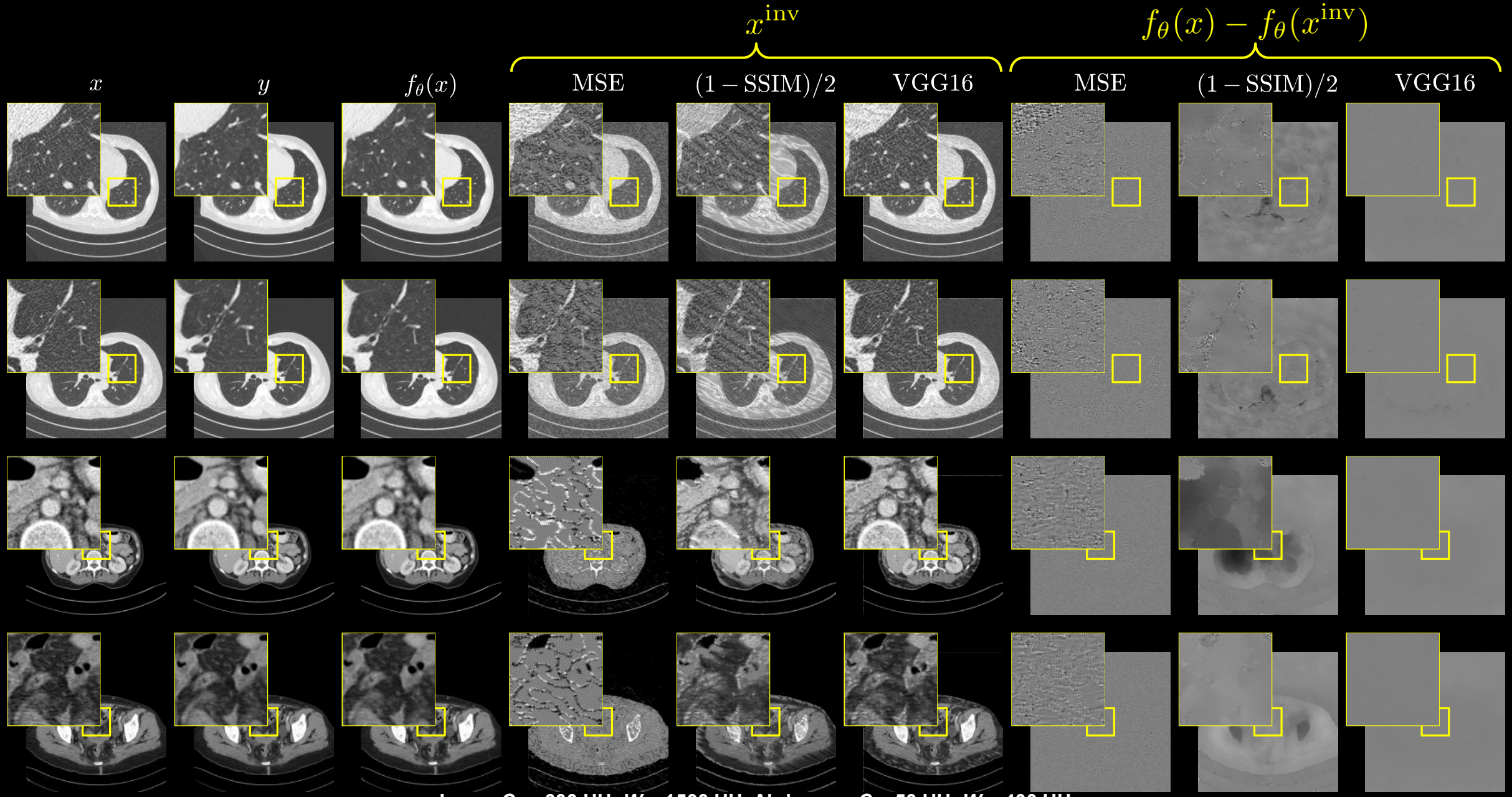
E. Eulig, J. Maier, B. Ommer, and M. Kachelrieß. May denoising remove structures? How to reconstruct invariances of CT denoising algorithms. SPIE Medical Imaging Conference Record, 1292506:1-6, April 2024.

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

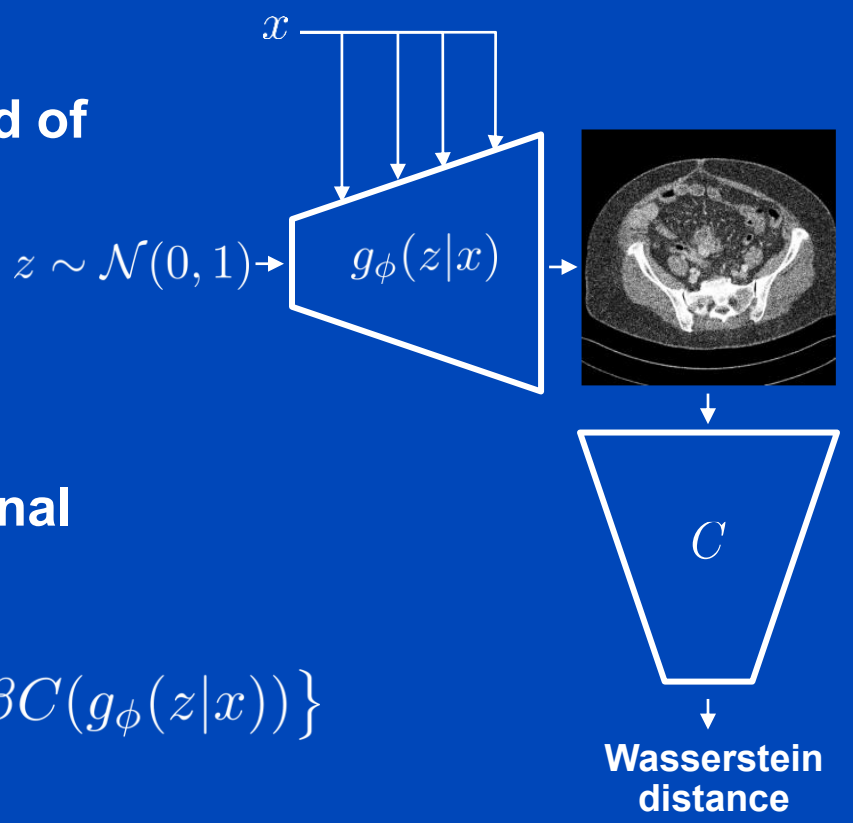


Lung: *C* = -600 HU, *W* = 1500 HU. Abdomen: *C* = 50 HU, *W* = 400 HU.

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.

dkfz.

Generating Invariances

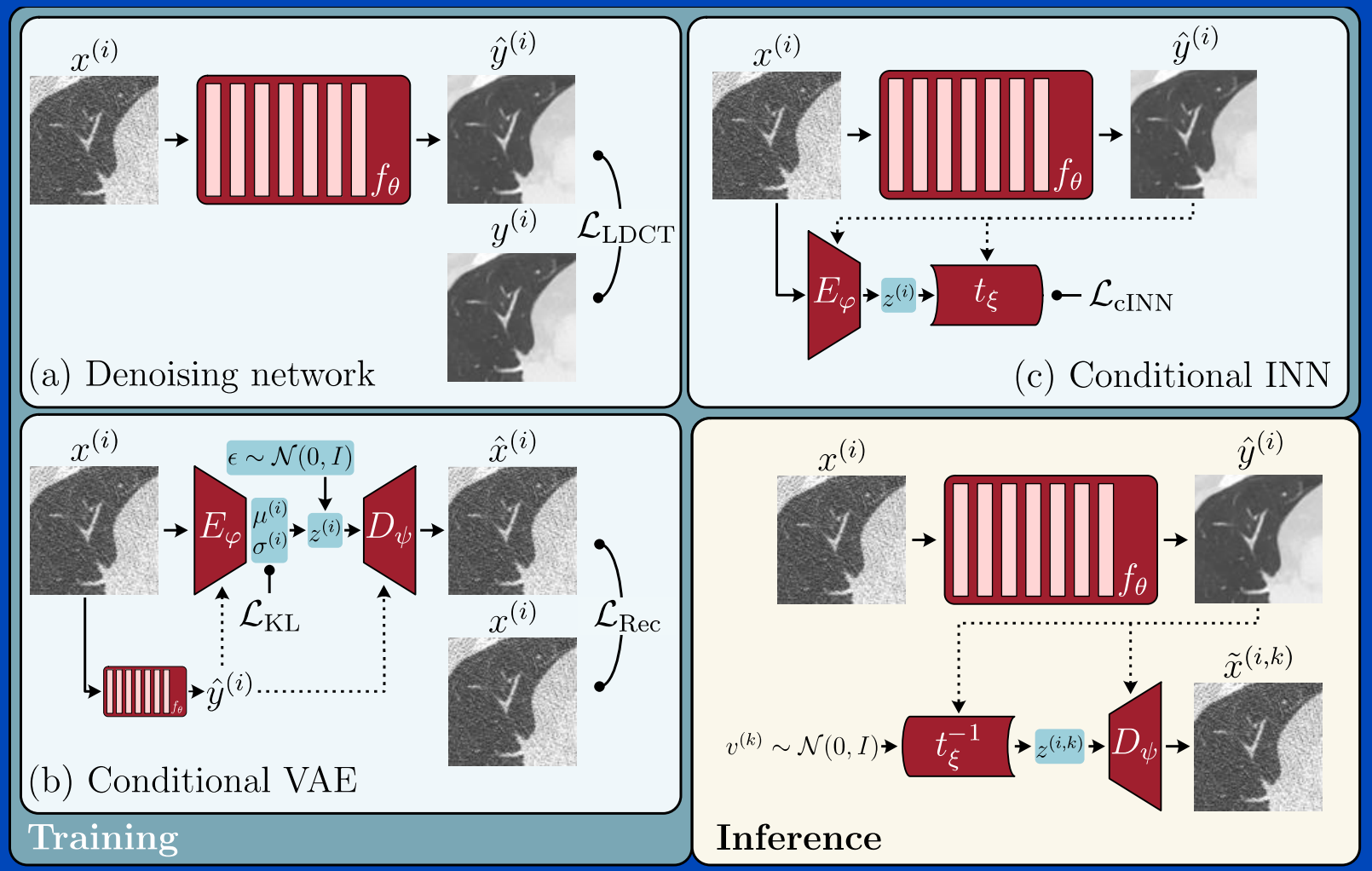
Pipeline to reconstruct invariances:

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.

dkfz.

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 marc.kachelriess@dkfz.de or through DKFZ's PhD program.

Parts of the reconstruction software were provided by RayConStruct® GmbH, Nürnberg, Germany.