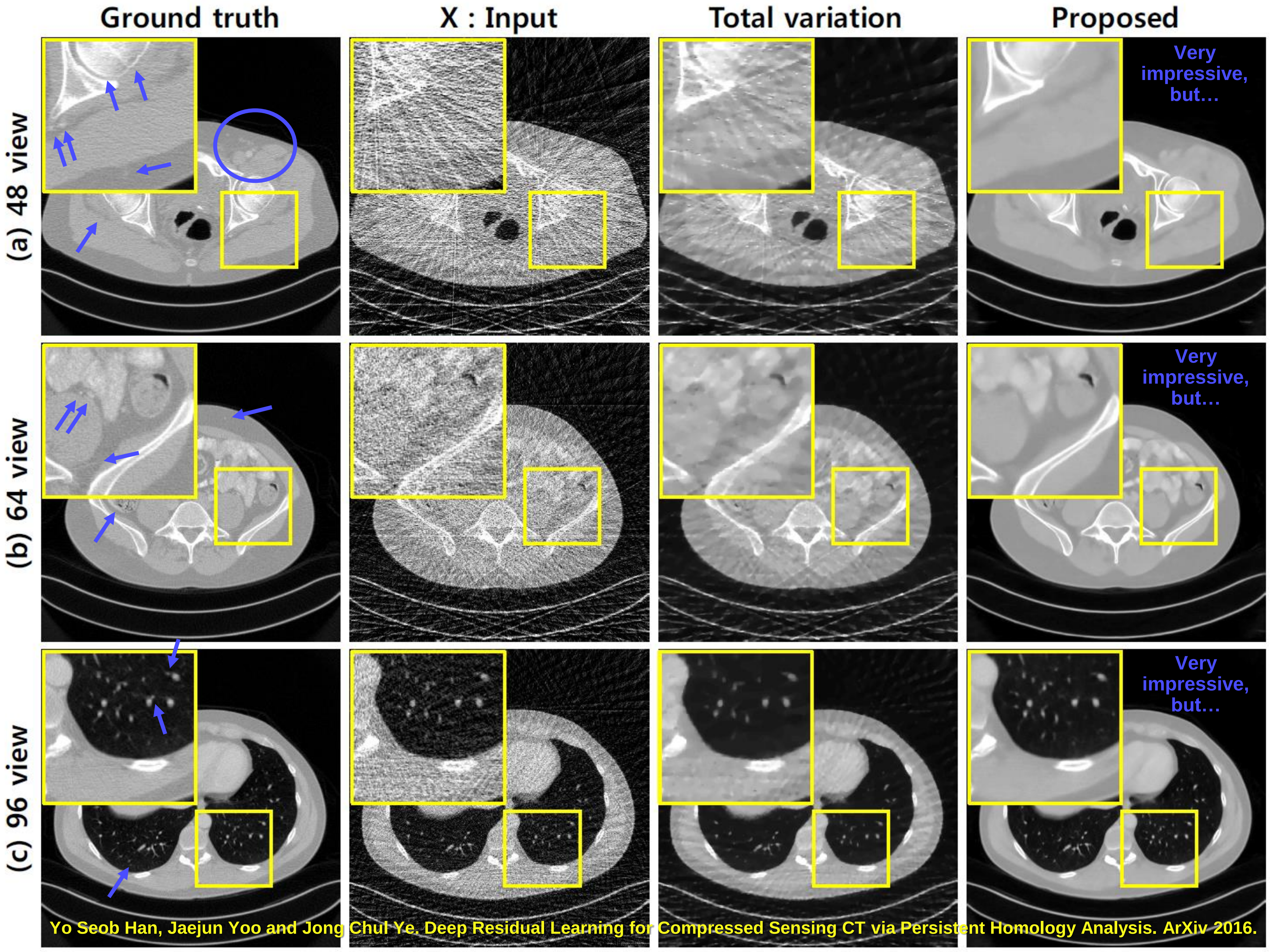
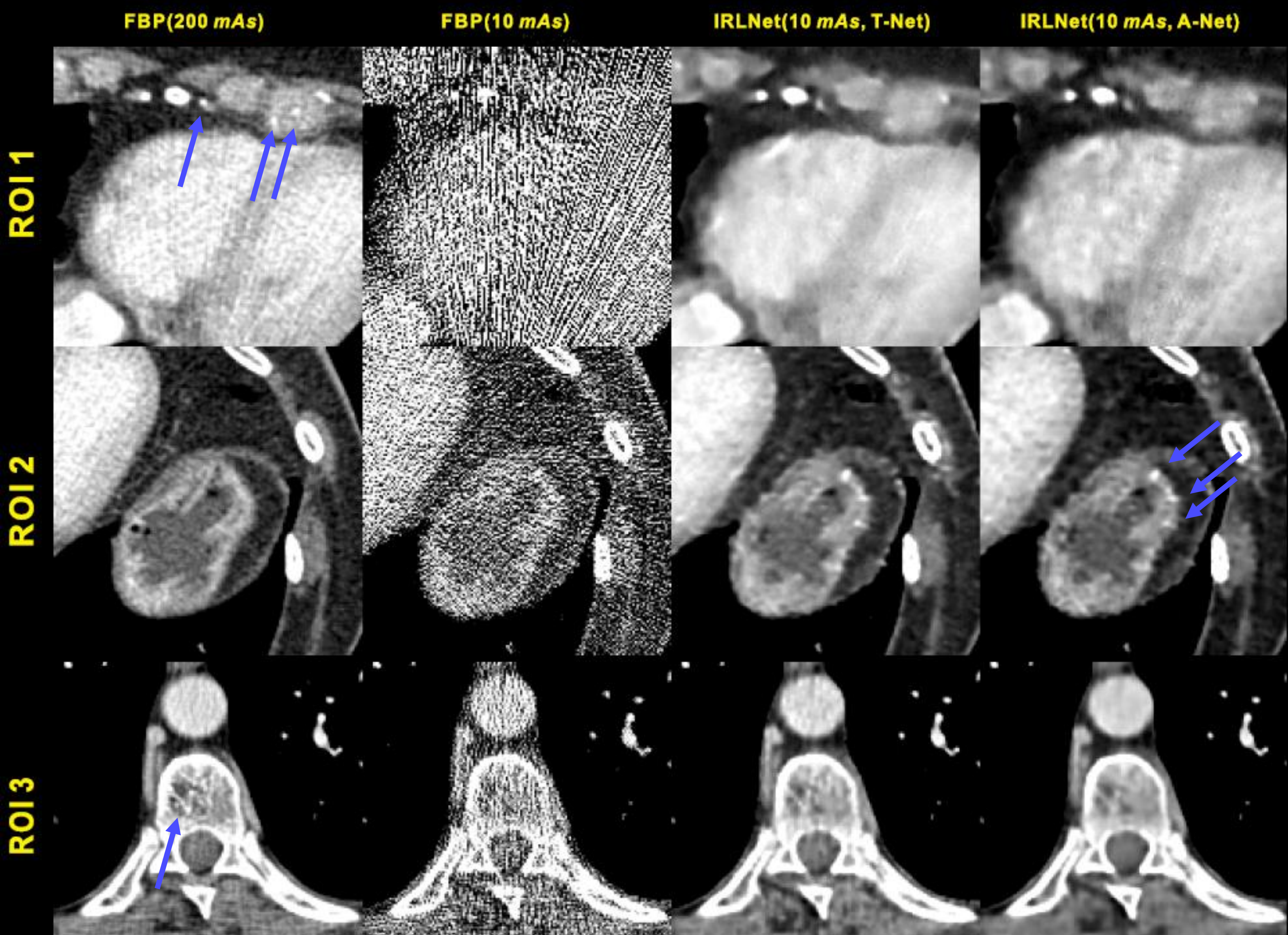
Invarianzen Neuronaler Netze am Beispiel der CT Rauschreduktion

Elias Eulig[1,2], Björn Ommer[2,3], and Marc Kachelrieß[1,2]

[1]German Cancer Research Center (DKFZ), Heidelberg, Germany
[2]Heidelberg University, Heidelberg, Germany
[3]Ludwig Maximilian University, Munich, Germany

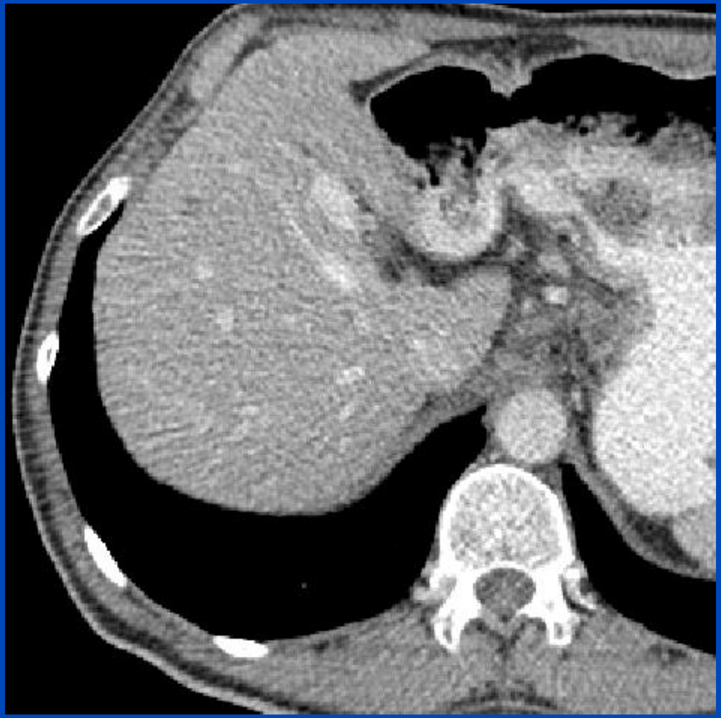
DEUTSCHES KREBSFORSCHUNGSZENTRUM IN DER HELMHOLTZ-GEMEINSCHAFT

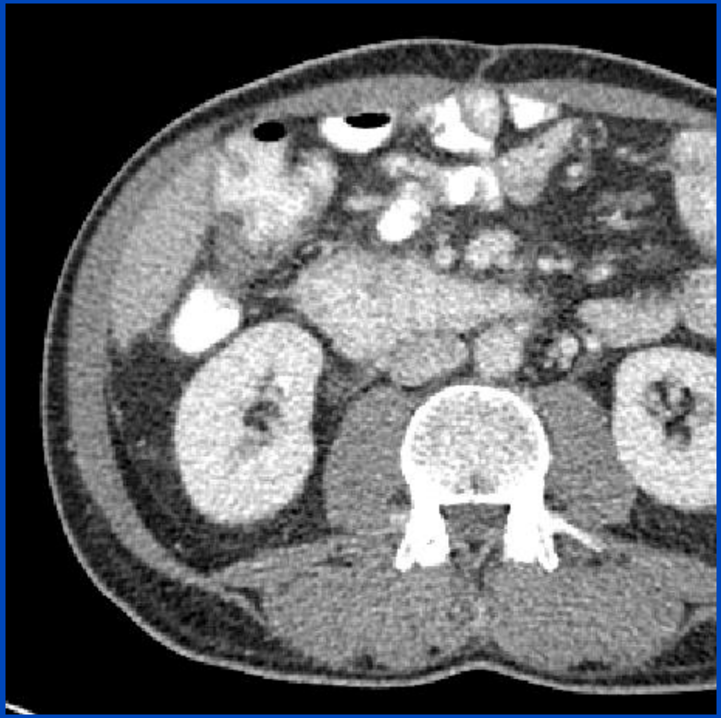

Ground truth
X : Input
Total variation
Proposed
(a) 48 view
(b) 64 view
(c) 96 view
Very impressive, but…
Very impressive, but…
Very impressive, but…
Yo Seob Han, Jaejun Yoo and Jong Chul Ye. Deep Residual Learning for Compressed Sensing CT via Persistent Homology Analysis. ArXiv 2016.



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

Noise Removal Example 3





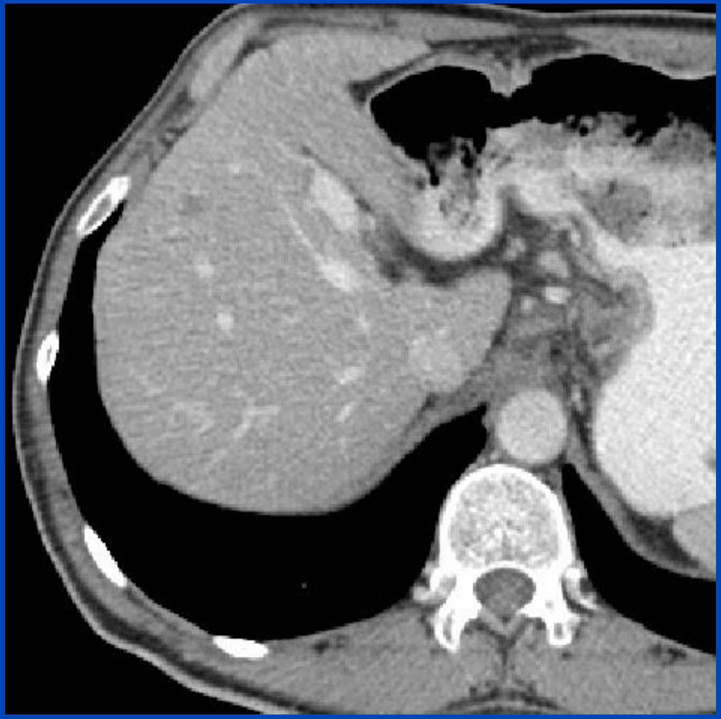
Low dose images (1/4 of full dose)

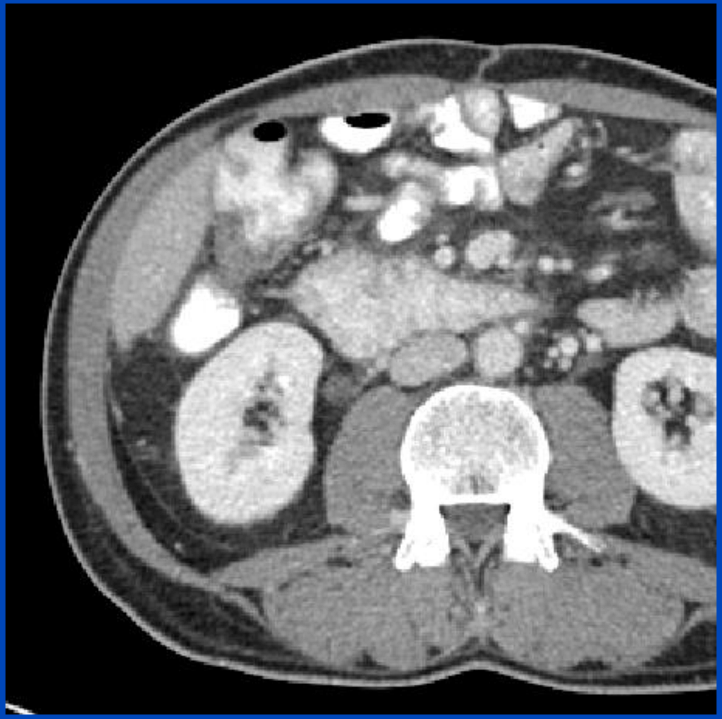
Noise Removal Example 3

Denoised low dose

Andrew D. Missert, Shuai Leng, Lifeng Yu, and Cynthia H. McCollough. Noise Subtraction for Low-Dose CT Images Using a Deep Convolutional Neural Network. Proceedings of the 5th CT-Meeting: 399-402, 2018.

Noise Removal Example 3





Full dose

Andrew D. Missert, Shuai Leng, Lifeng Yu, and Cynthia H. McCollough. Noise Subtraction for Low-Dose CT Images Using a Deep Convolutional Neural Network. Proceedings of the 5th CT-Meeting: 399-402, 2018.

Motivation

In general:

- **Deep learning methods are employed for many problems.**
- **However, they lack interpretability (black-box).**
- **Recent advancement in generative modelling signal false confidence.**

Aim:

- **Lay fundamentals for post-hoc interpretability and robustness analysis of denoising DNNs.**
- **Use two simple denoising networks *f* as initial examples:**
  - **Chen‘s simple 3-layer CNN trained with $\mathcal{L}_2$ loss[1]**
  - **Yang‘s Wasserstein GAN with additional perceptual loss[2]**
- **See what they have learned to represent and what to ignore: For a given output *x´* there are many inputs *x* that produce the same output *x´* = *f*(*x*).**
- **Employ low dose CT image and projection dataset for all studies.[3]**

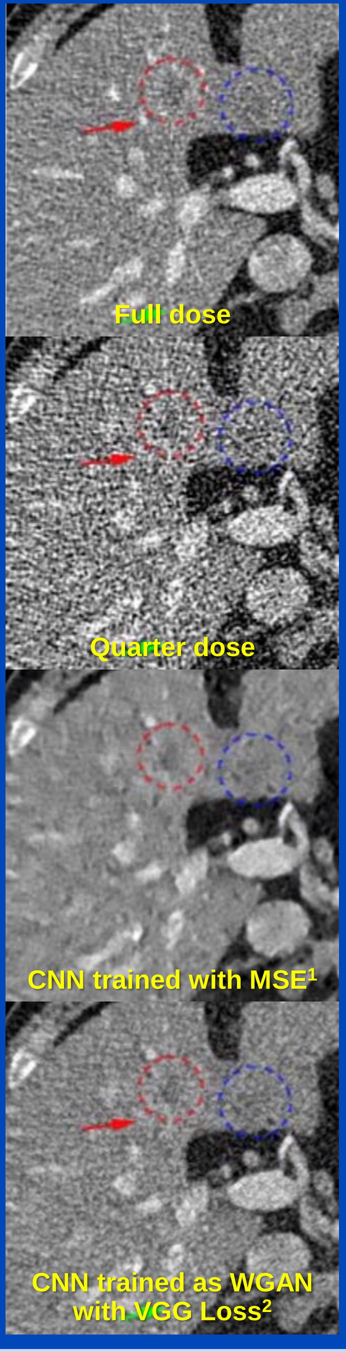
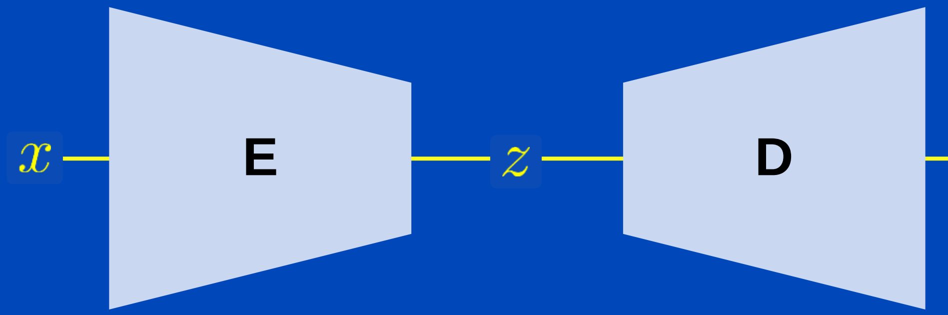


Figure from reference [2]

[1] H. Chen et al., "Low-dose CT denoising with convolutional neural network”, ISBI 2017, 2017.
[2] Q. Yang *et al*., "Low-Dose CT Image Denoising Using a Generative Adversarial Network […]”, in *IEEE TMI*, vol. 37, no. 6, 2018.
[3] C. McCollough et al., “Data from low dose CT image and projection data [data set],” The Cancer Imaging Archive, 2020.

Recap 1: What is an Autoencoder (AE)?

- **In and output domain are the same, here *x*.**
- **Bottleneck *z* enforces the encoder and decoder to do a good job.**

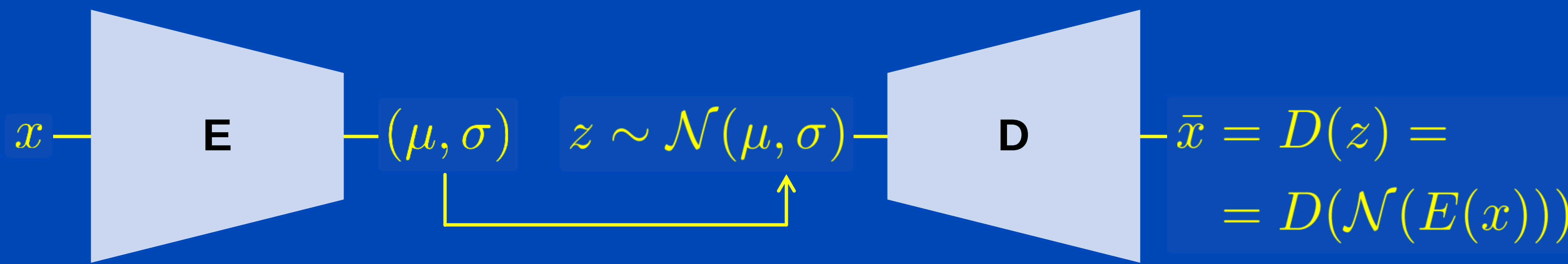


$$\bar{x} = D(z) = \\ = D(E(x))$$

- **Examples:**
  - **Principal component analysis (linear autoencoder), lossless**
  - **PCA with dimensionality reduction (nonlinear due to clipping), lossy**
  - **Image compression and decoding, e.g. jpeg, lossy**
- **Latent space typically not interpretable.**

Recap 2: What is a Variational AE (VAE)?

- Make latent space regular.
- Allow to sample in latent space from a given distribution, here: normal distribution.



- The VAE is a generative model.
- It allows to generate new data by sampling new values from the multivariate normal distribution.

Method

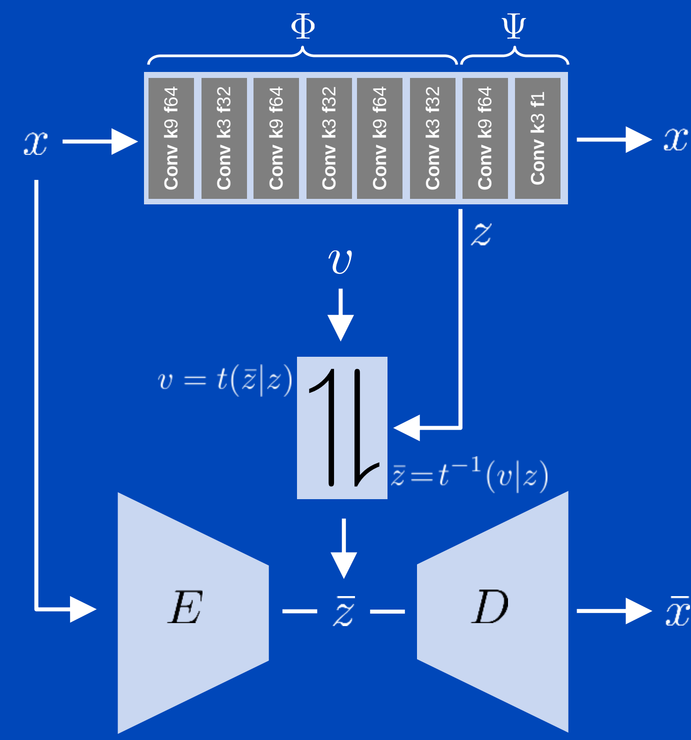
Recovering Invariances

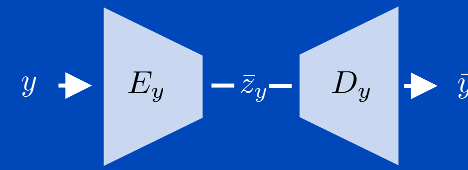
1. **Our work is based on Rombach et al.[1]**
2. **Use denoising methods Chen et al. & Yang et al.**
3. **Train VAE to learn a complete data representation of the low dose images *x*.**
4. **For each denoising method and layer in the network we wish to evaluate, train a cINN to recover the invariances.**
5. **For a given test image, sample 250 invariances *v*, apply the inverse mapping *t* $^{-1}$ and apply the pretrained decoder *D*.**

$$t^{-1} \text{ maps } \mathcal{N}(0,1) \text{ onto } p(\bar{z}|z).$$

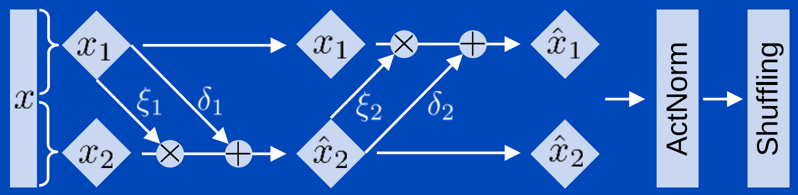
Thus it produces only images that are likely under the training distribution of the AE.

$$f(x) = \Psi(\Phi(x))$$





Alternative: Use VAE in high dose domain, i.e. VAE_y, to visualize the invariances.



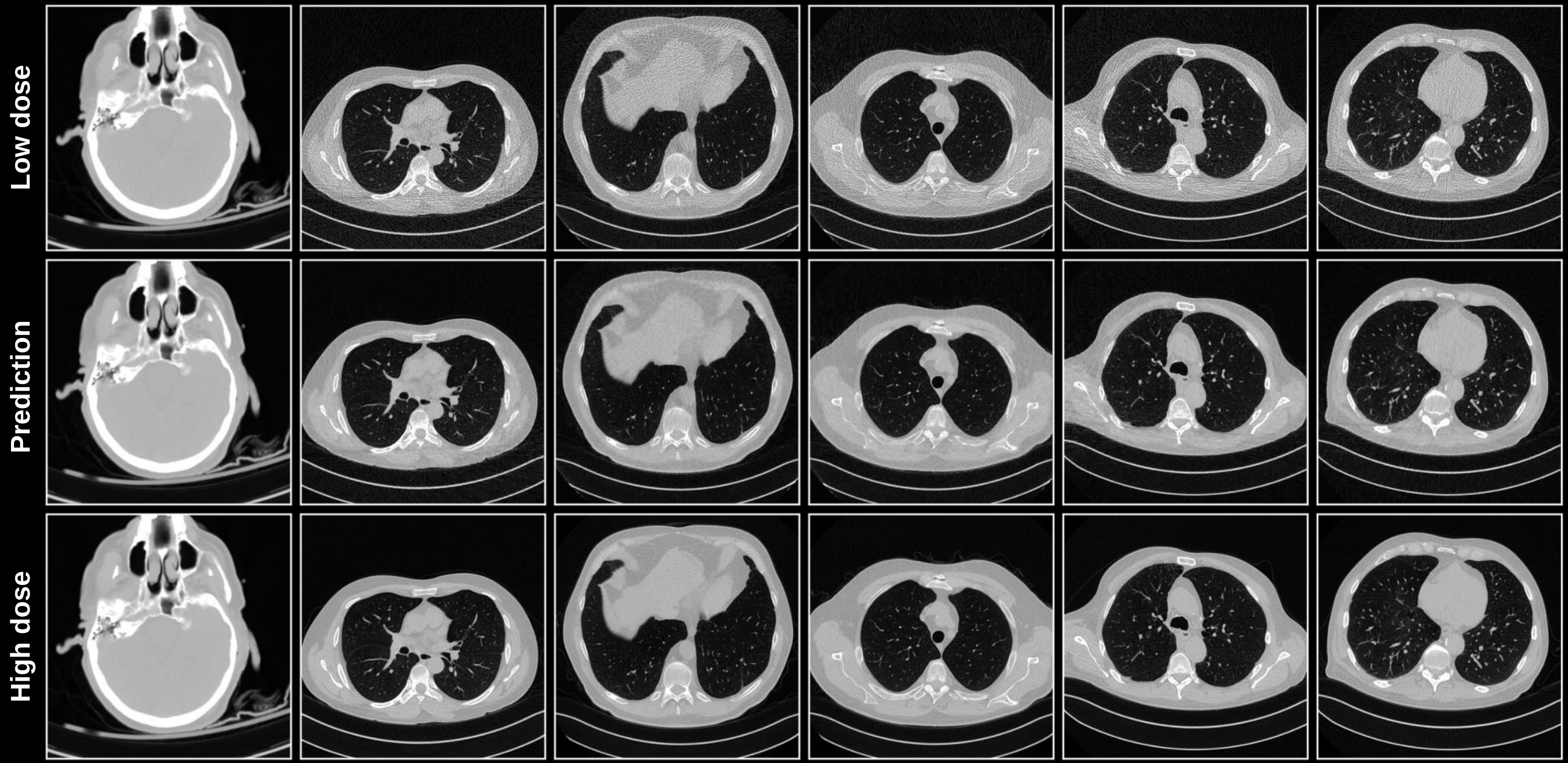
Building block of INN: Invertible block, ξ_{12} and δ_{12} are CNNs or NNs

$$x_1 \exp(\xi_2(\hat{x}_2)) + \delta_2(\hat{x}_2) = \hat{x}_1$$
$$x_2 \exp(\xi_1(x_2)) + \delta_1(x_1) = \hat{x}_2$$

[1]Rombach et al. "Making sense of CNNs: Interpreting deep representations and their invariances with INNs", ECCV 2020.

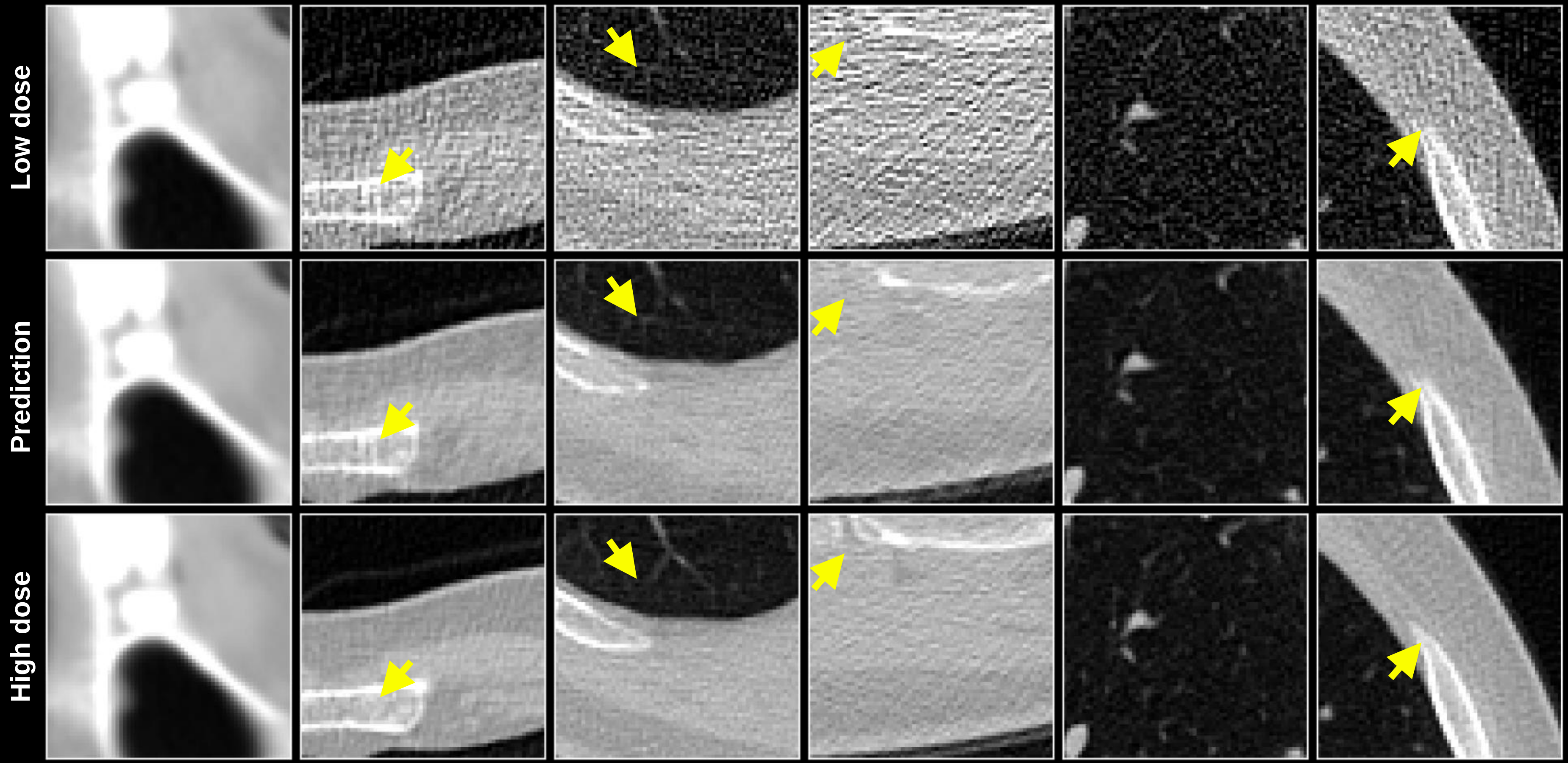
Results

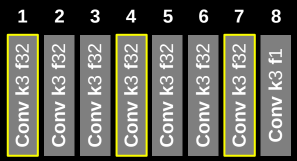
Denoising (Yang et al.) $f = \Psi \circ \Phi$



Results

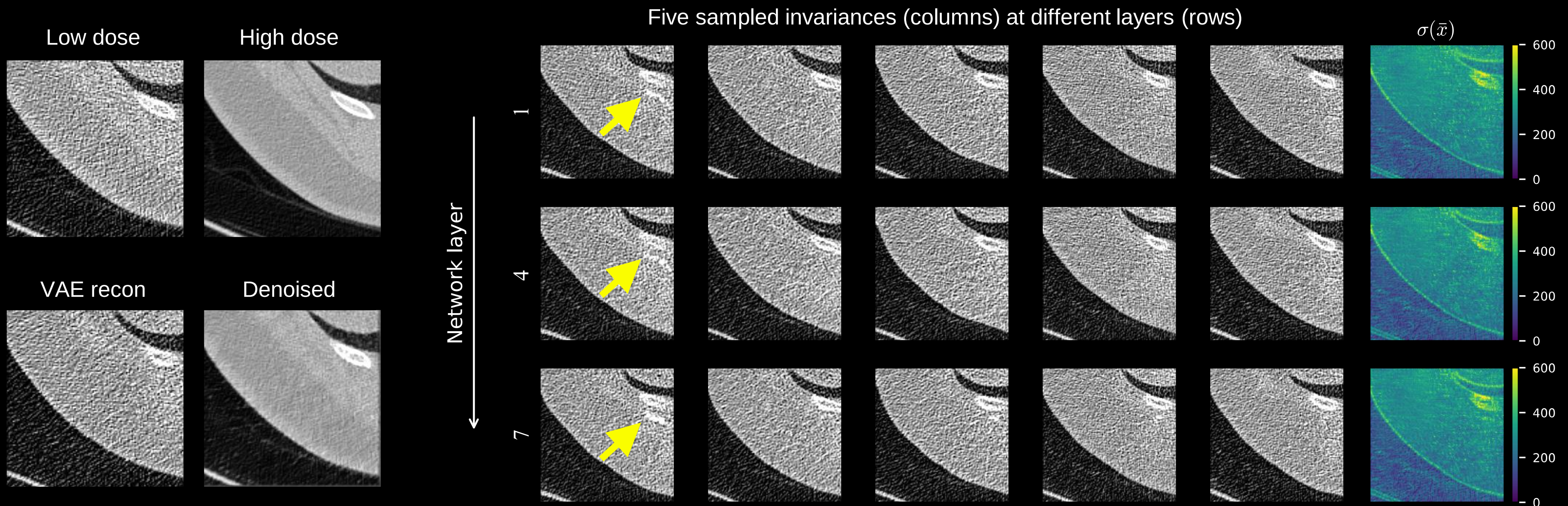
Denoising (Yang et al.) $f = \Psi \circ \Phi$

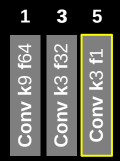


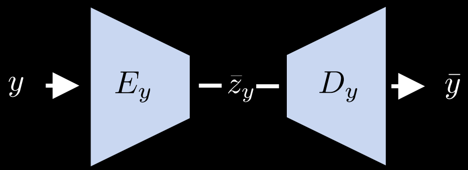


Results

Sampling Invariances (Yang et al.)

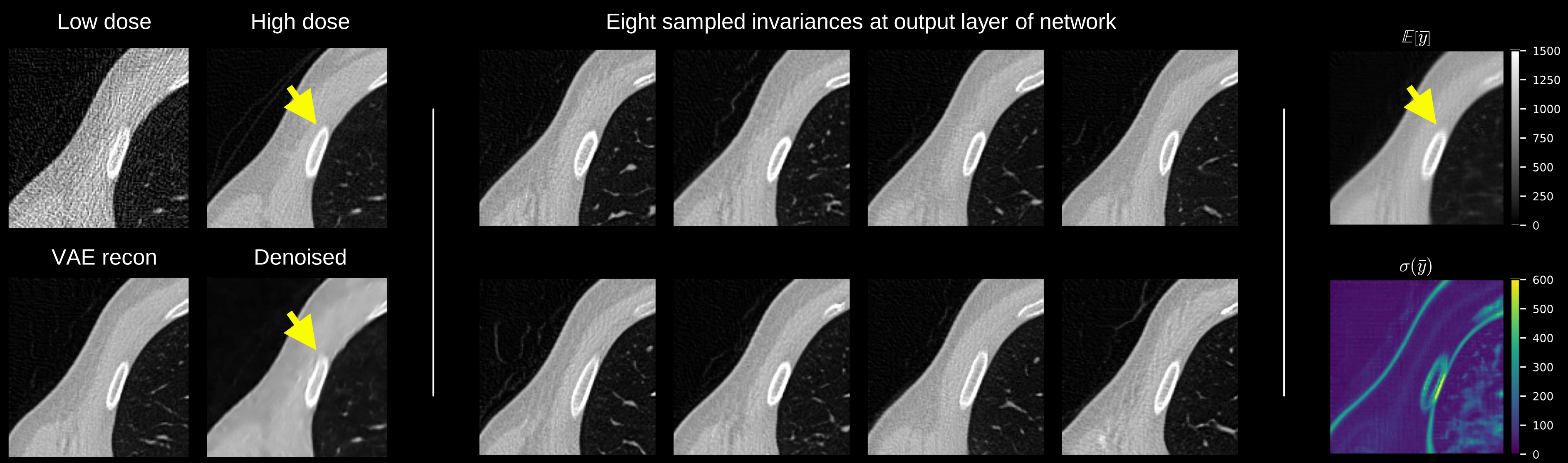






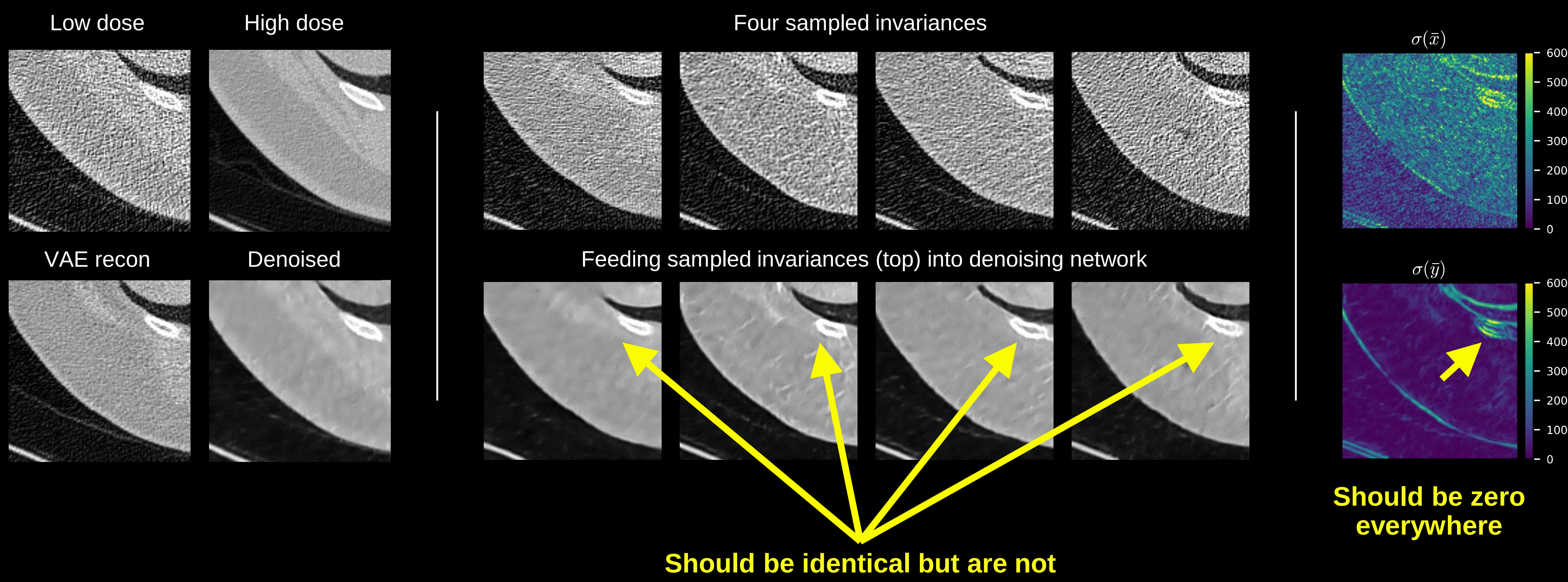
Results

Sampling Invariances in Target Domain (Chen et al.)



Test

Feed sampled invariances back into the network



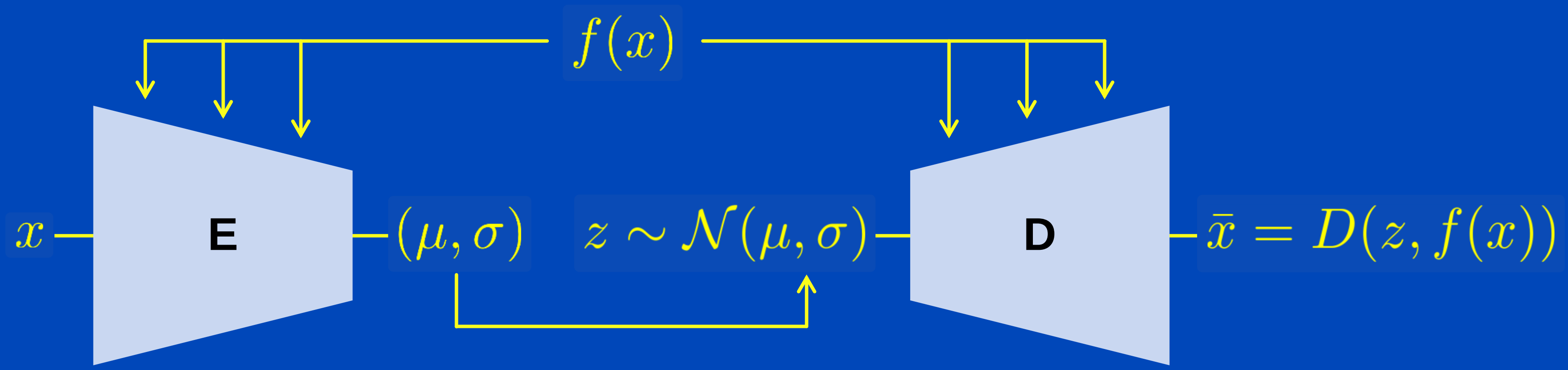
Conclusion & Outlook

- Novel method to highlight invariances of a given denoising network
- Architecture-agnostic (also works for MRI, PET, …)
- Feeding invariances back leads to different outputs $\rightarrow$ VAE is a severe limitation

- Outlook
  - Improve VAE (use conditional VAE)
  - Further analyze sampled invariances
  - Only show “interesting” invariances to the reader

Idea: Condition VAE on *f*(*x*)

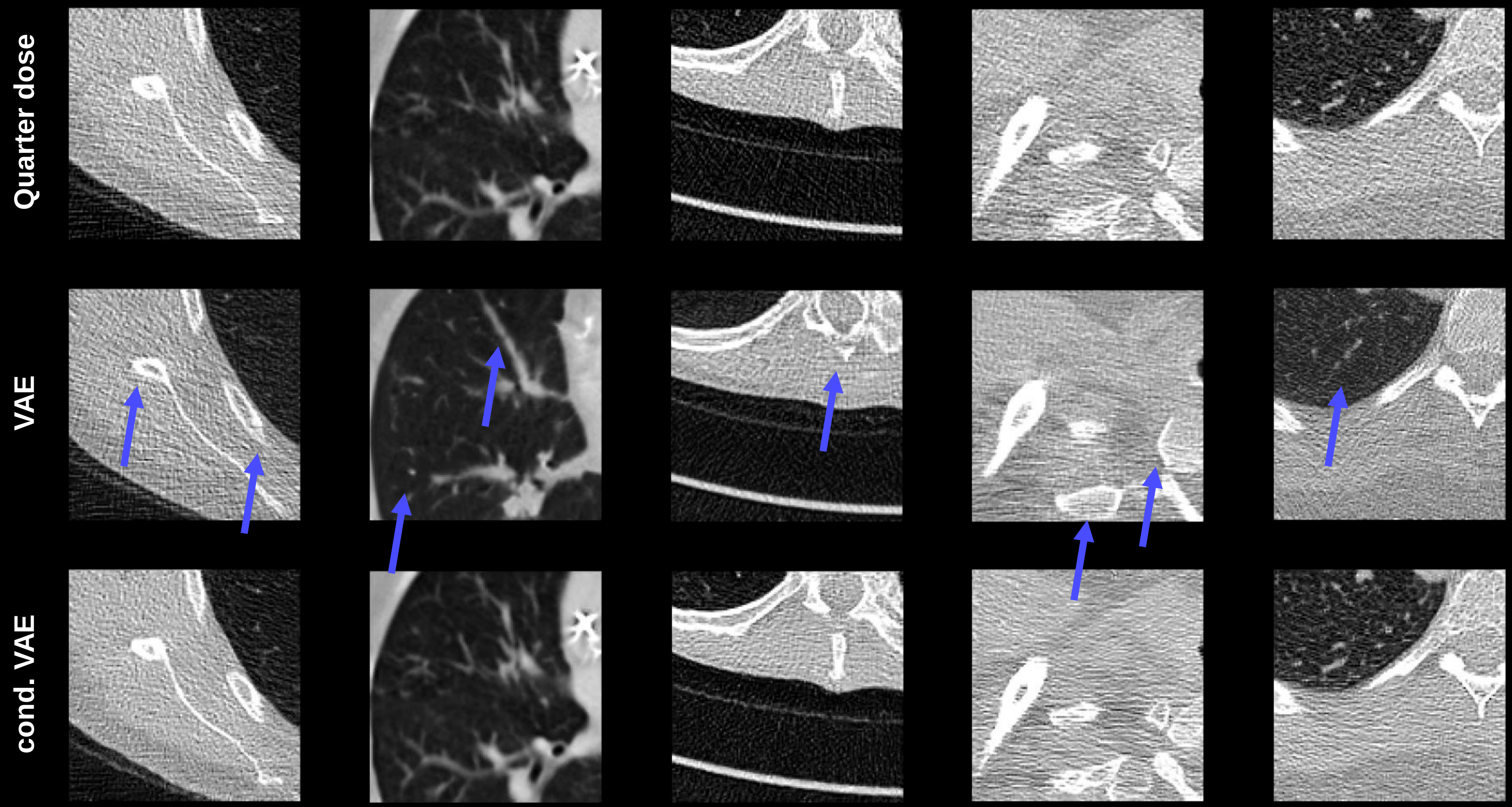
- **No need to encode what's in the denoised image *f*(*x*).**



- **Note: This also eliminates the need for a cINN as we can now directly sample from $p(x|z \sim \mathcal{N}(E(x)), f(x))$.**

Results

From VAE to conditional VAE



Thank You!

This presentation will soon be available at www.dkfz.de/ct.
This work was supported in part by the Helmholtz International Graduate School for Cancer Research, Heidelberg, Germany.

The 8th International Conference on

Image Formation in X-Ray Computed Tomography

Conference Chair

Marc Kachelrieß, German Cancer Research Center (DKFZ), Heidelberg, Germany