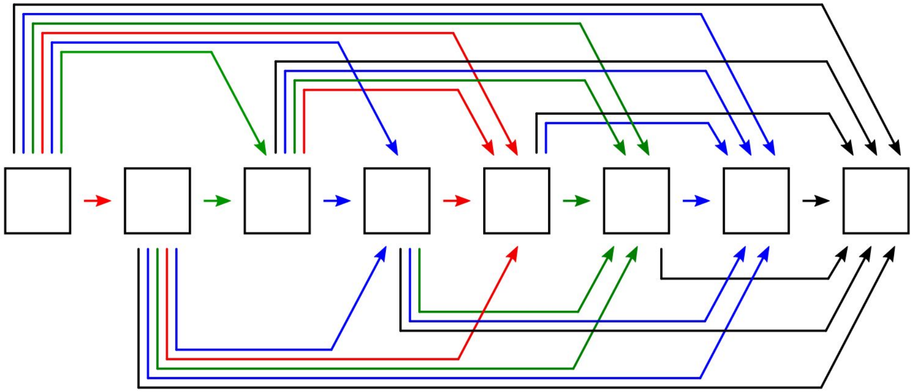
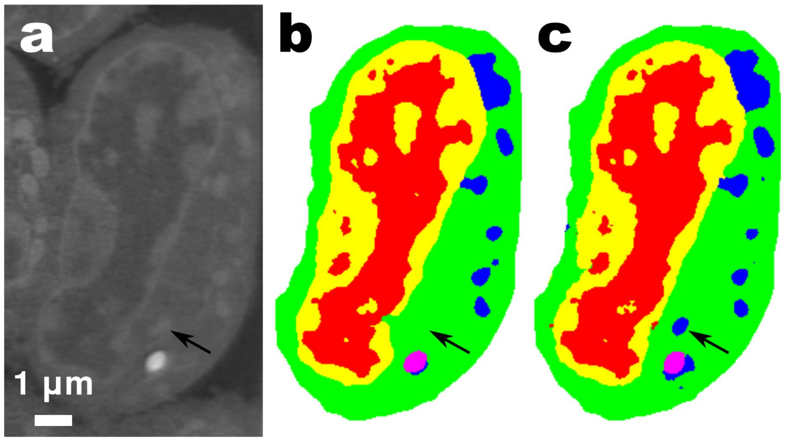




Machine learning for large scientific images

D.M. Pelt

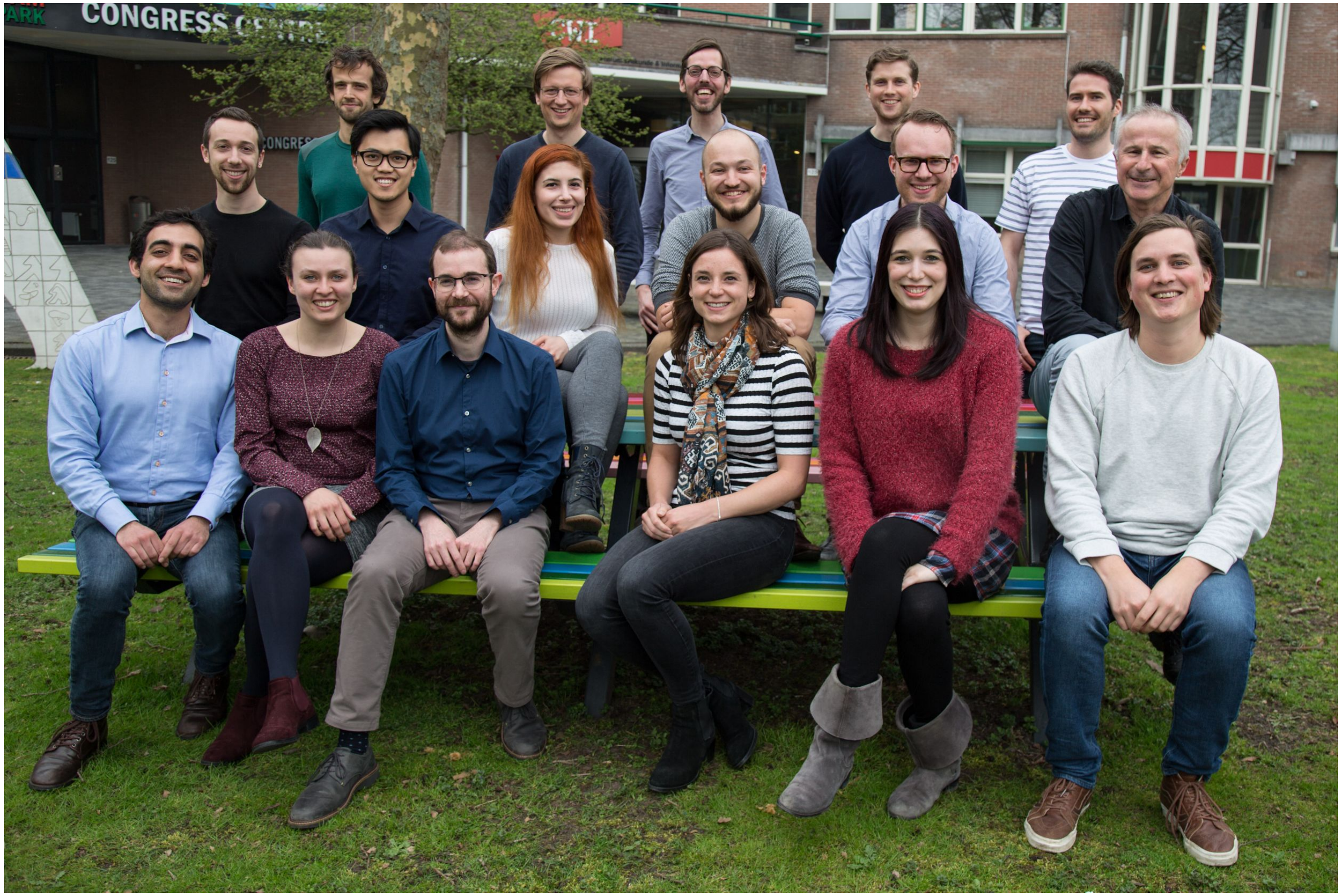
Computational Imaging Group, CWI

2019 Woudschoten Conference

About me

- Performed PhD research at CWI (2016)
  - Supervised by prof. dr. K. J. Batenburg
  - Topic: efficient tomographic reconstruction algorithms
- After PhD spent a year as post-doc at Berkeley Lab
  - Supervised by prof. J. A. Sethian
  - Focus on developing machine learning algorithms
- Currently post-doc at Computational Imaging group
- Recently awarded NWO VENI grant (2018)
  - Title: Machine learning for large 3D tomographic images

Computational Imaging Group



- Growing group of researchers focused on imaging
  - Broad range of topics and expertises
  - https://www.cwi.nl/research/groups/computational-imaging

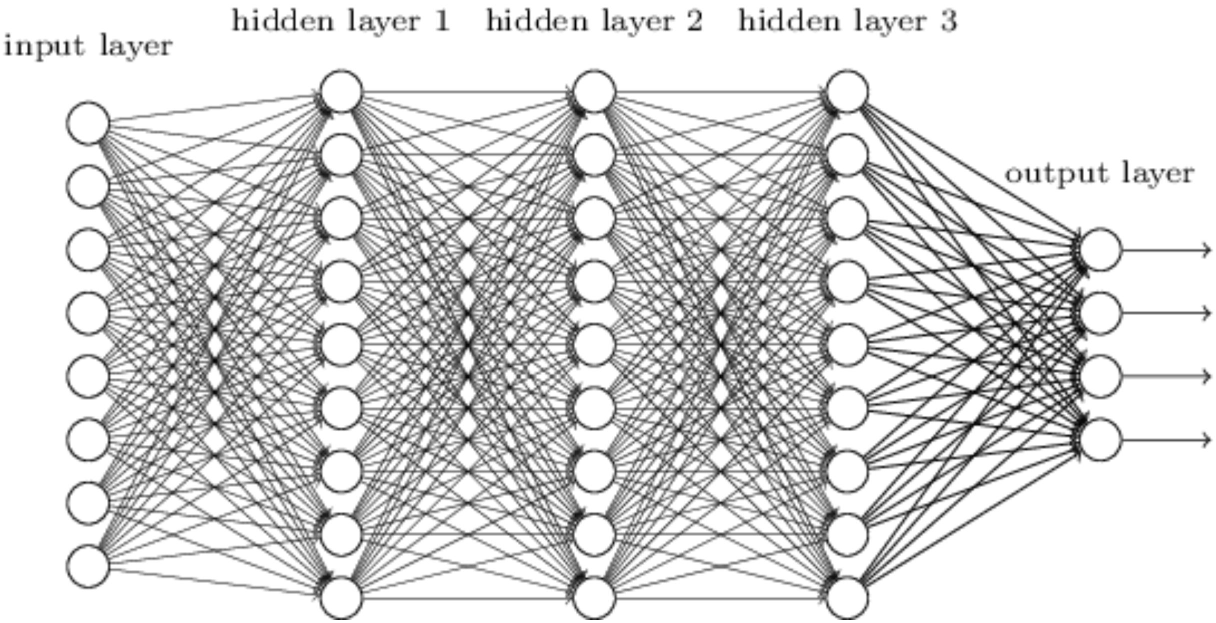
Overview

Recent developments in ML for scientific images

- Brief introduction to machine learning for images
- Problems with popular existing approaches
- Recently proposed new approach for scientific images
- Results for various scientific problems
- Conclusions and outlook

Neural networks

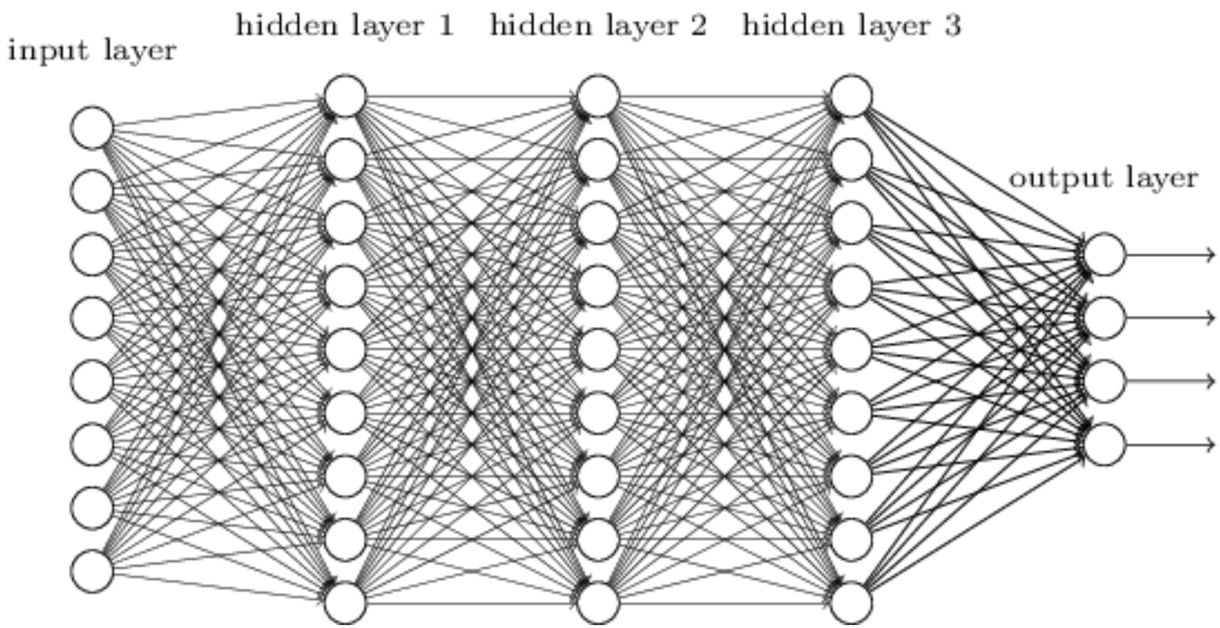
- Model some unknown function $f : \mathbb{R}^m \to \mathbb{R}^n$ using a nonlinear network g_ϕ with parameters $\phi \in \mathbb{R}^{N_\phi}$



- f and ϕ are unknown

Training networks

- Training: find ϕ such that $g_\phi(\mathbf{x}) \approx f(\mathbf{x})$
- Supervised learning: pairs of $(\mathbf{x}, f(\mathbf{x}))$ are known

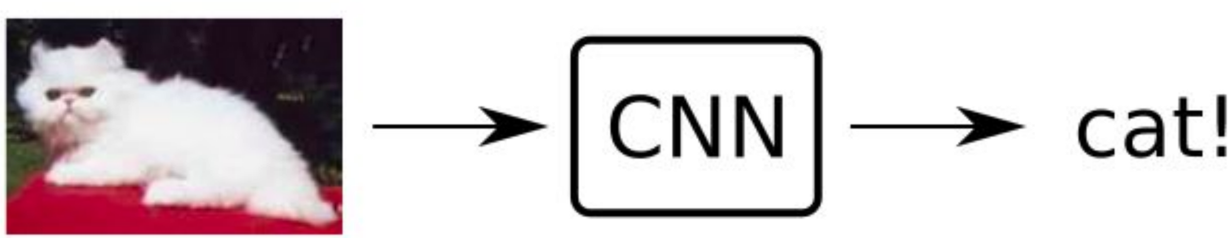


- Partial derivatives are easy and fast to compute
  - Backpropagate error function derivative using chain rule

CWI

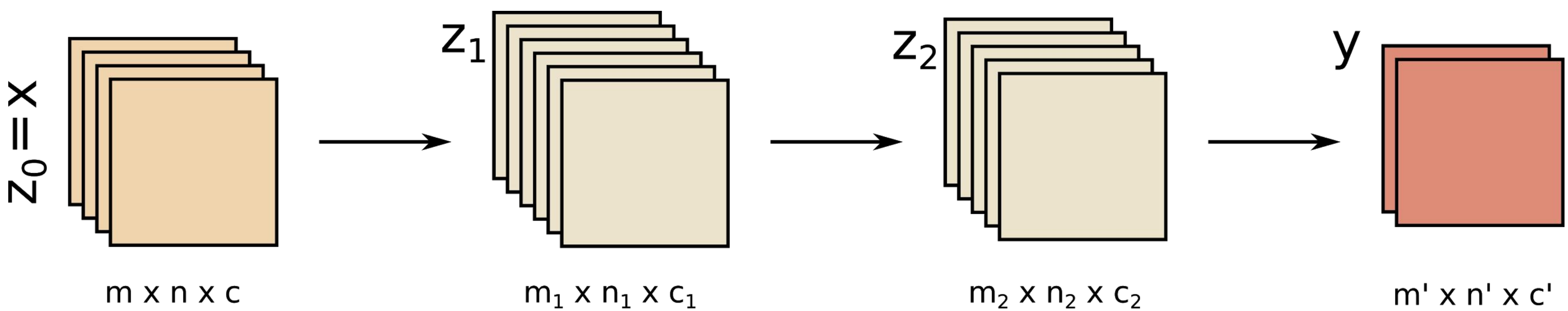
Machine learning for images

- Rapid increase in popularity after 2012
  - Even though math is much older!
- 2018 Turing award for Bengio, Hinton, and LeCun
- Used daily by companies (Google, Facebook, etc)
- Typically, use Convolutional Neural Networks (CNN)

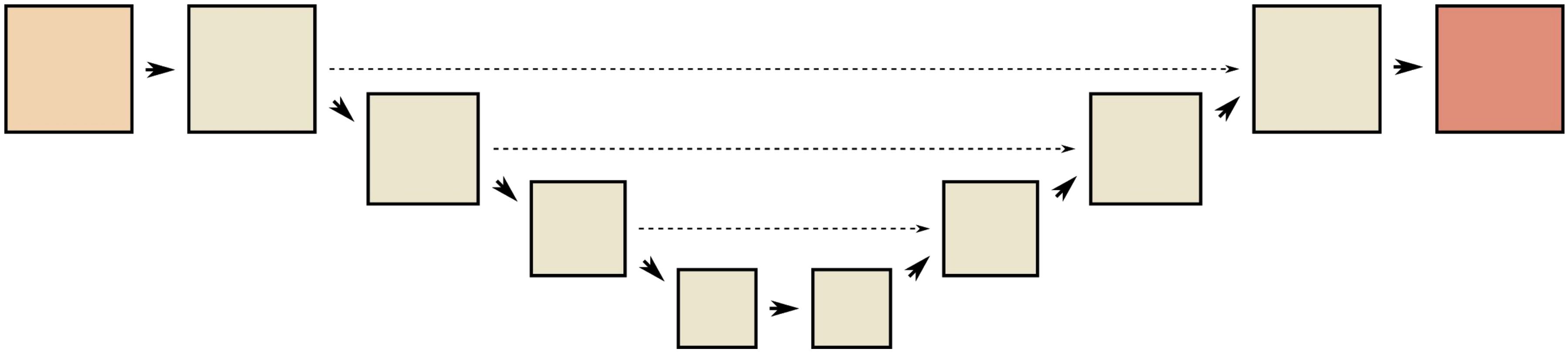


Convolutional neural networks

- Learn mapping from input image x to output image y
- Each layer convolves images of previous layer
- Often small filters (e.g. 3x3), but many layers/images
- Learn filters by presenting input/target pairs



Encoder-decoder networks



- Deep networks have many layers
- Typically, use downscaling and upscaling to capture features at different scales
- First encoder, then decoder

Applications of CNNs

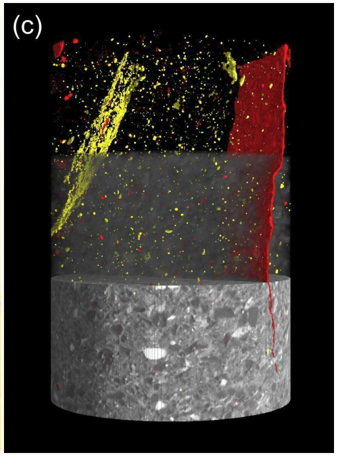
Aerial to Map

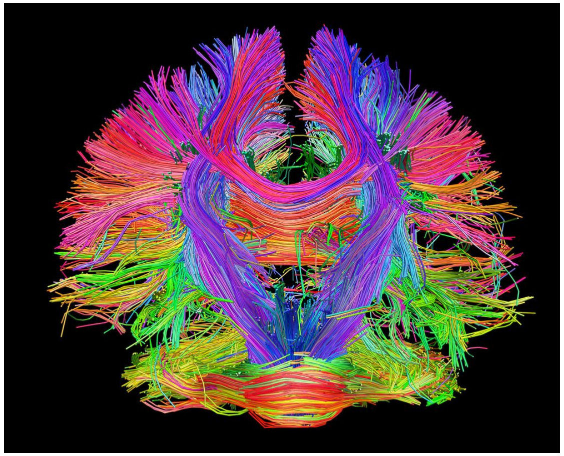
Photograph Monet Van Gogh Cezanne Ukiyo-e

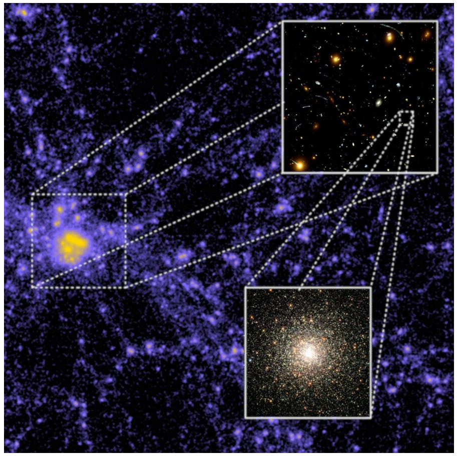
Sources from GitHub: junyanz/CycleGAN, phillipi/pix2pix, pathak22/context-encoder

Potential for scientific images

- ML has proven successful for photographic images
- Has large potential for scientific images
  - For example, ML could improve image quality, enable automatic segmentation, perform analysis, …

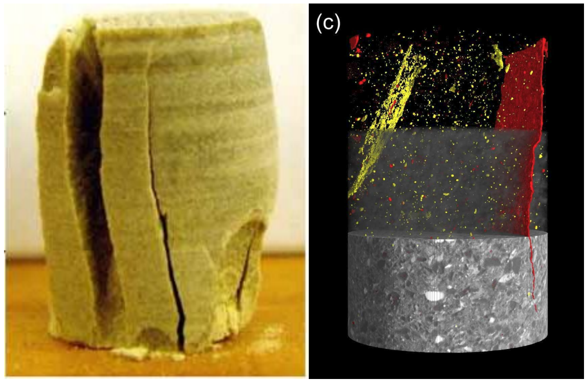


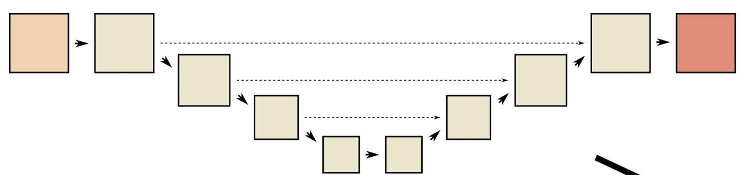




How can machine learning be applied in science?

Obvious approach

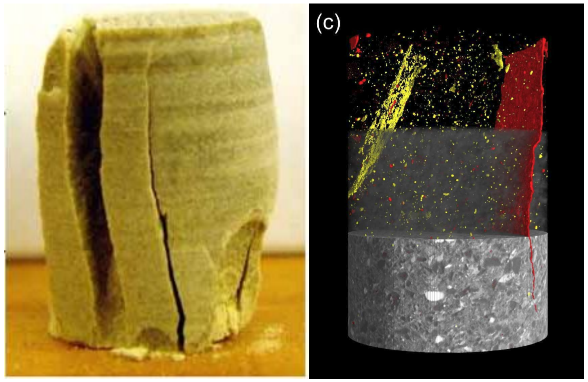


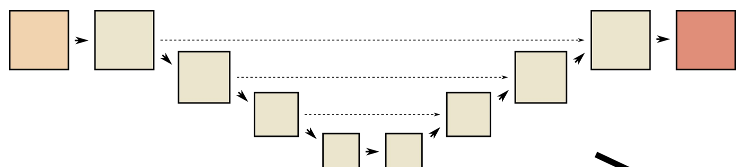


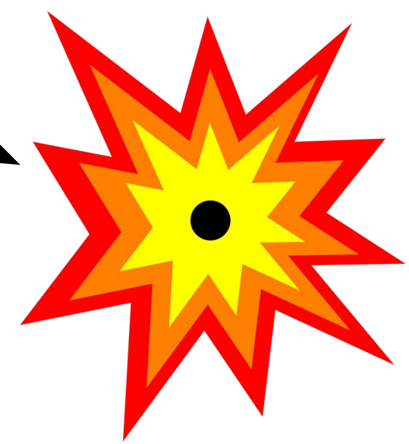
?

CWI

Obvious approach

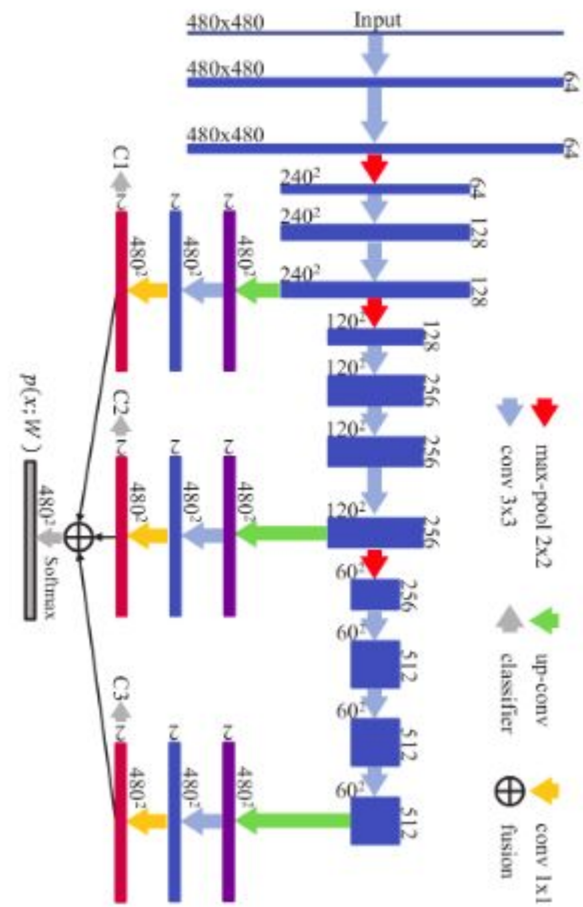






Problem 1: practical

- Complicated networks
  - How to choose which operations to use and how to combine them?
- Implementing the networks requires expert knowledge
- What works in one problem often does not work well for others
  - Changes are needed that require expert knowledge as well



Problem 2: mathematical

- Networks require many parameters to achieve accurate results (e.g. millions)
- Problem: networks prone to overfitting training set
  - Large networks are often pre-trained with a huge training set (e.g. ImageNet >10 million images)

- In scientific problems, training data is often limited (e.g. only a few images)
  - Training large networks with limited data is difficult

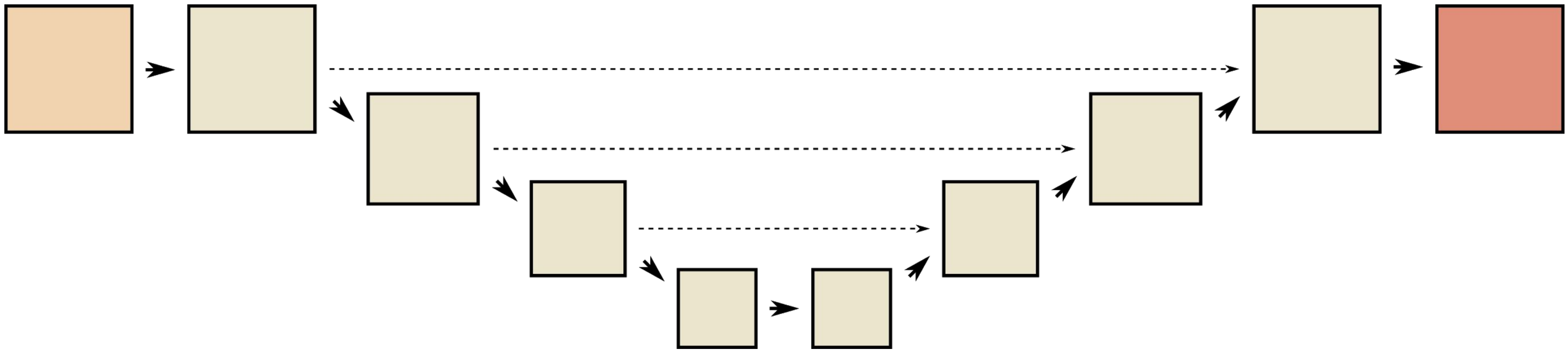
Problem 3: computational

- Networks require many layers and intermediate images to achieve accurate results
- Applying to large images requires prohibitively large computational costs
  - GPU memory, computation time

- Networks often tested on small images (e.g. 28x28 up to 256x256)
  - Typical synchrotron CT image is 2560x2560

New approaches specific to scientific images are needed!

Enc-Dec inefficiencies

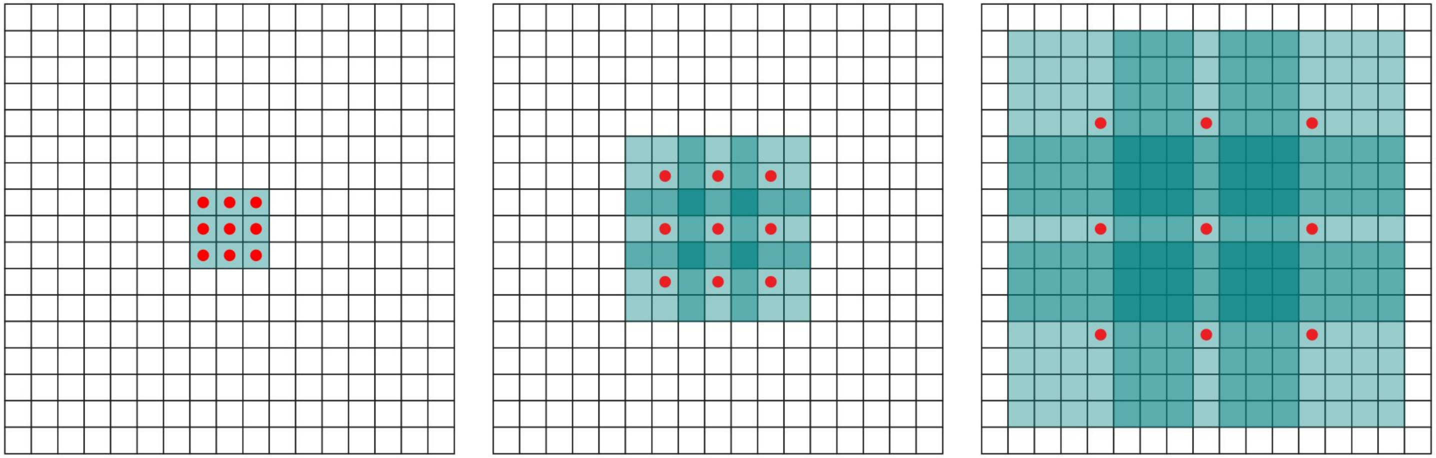


- Scaling operations impose structure
- Detected features have to be copied to deeper layers
- Decoder cannot be used to improve encoder

- Better: remove scaling, reuse features, mix decoder and encoder

Solution 1: remove scaling

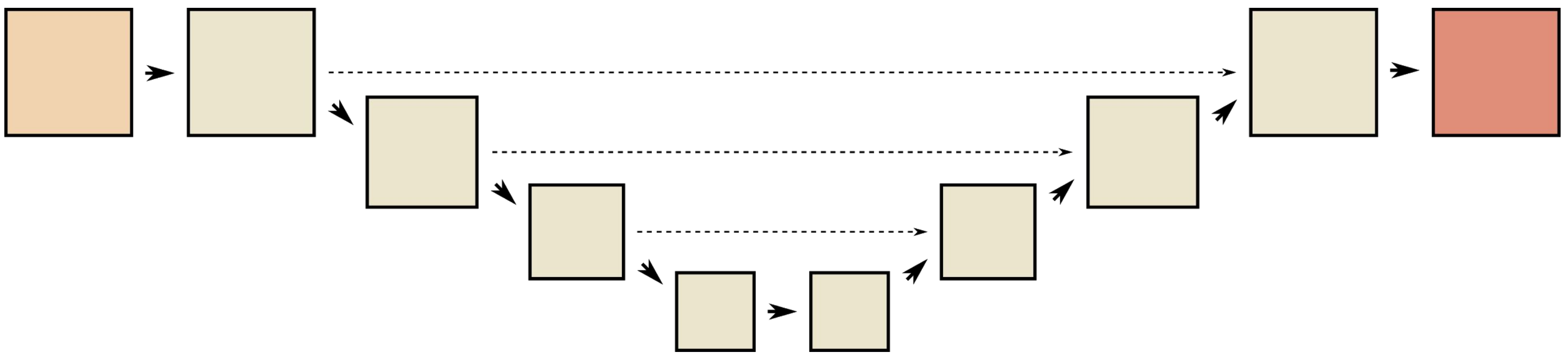
- Use dilated convolutions instead of scaling to capture features at larger scales

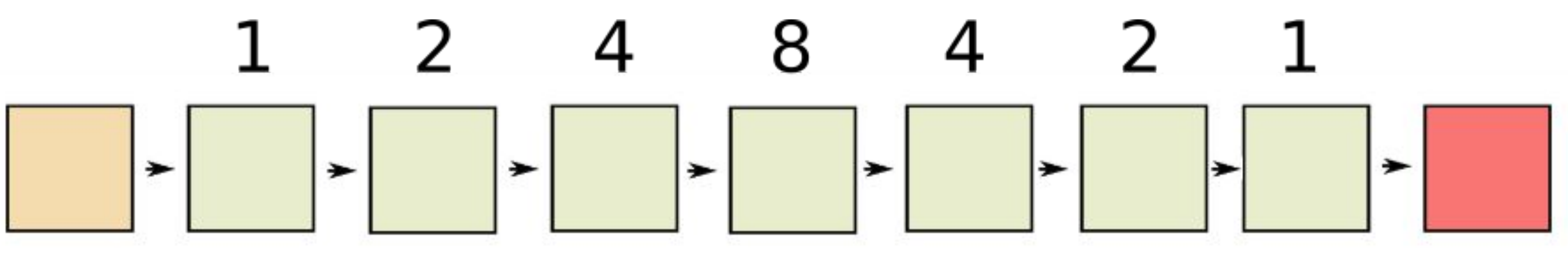


(a) (b) (c)

Yu, Fisher, and Vladlen Koltun. "Multi-scale context aggregation by dilated convolutions." arXiv preprint arXiv:1511.07122 (2015).

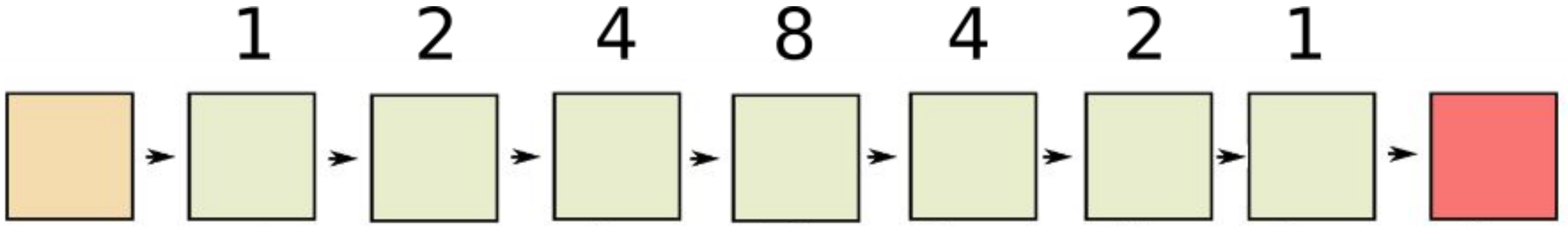
Solution 1: remove scaling





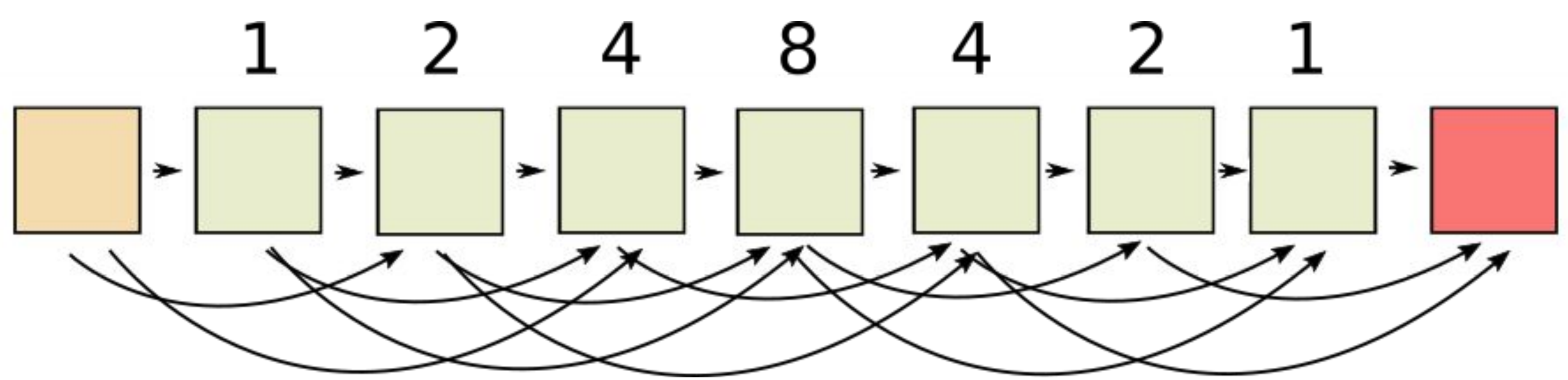
Solution 2: dense connections

- No scaling: all intermediate images are same size
- Why limit input to previous layer only?
- In fact, can use all previous images as input!
  - Including input images
- To compute output, can use all intermediate images



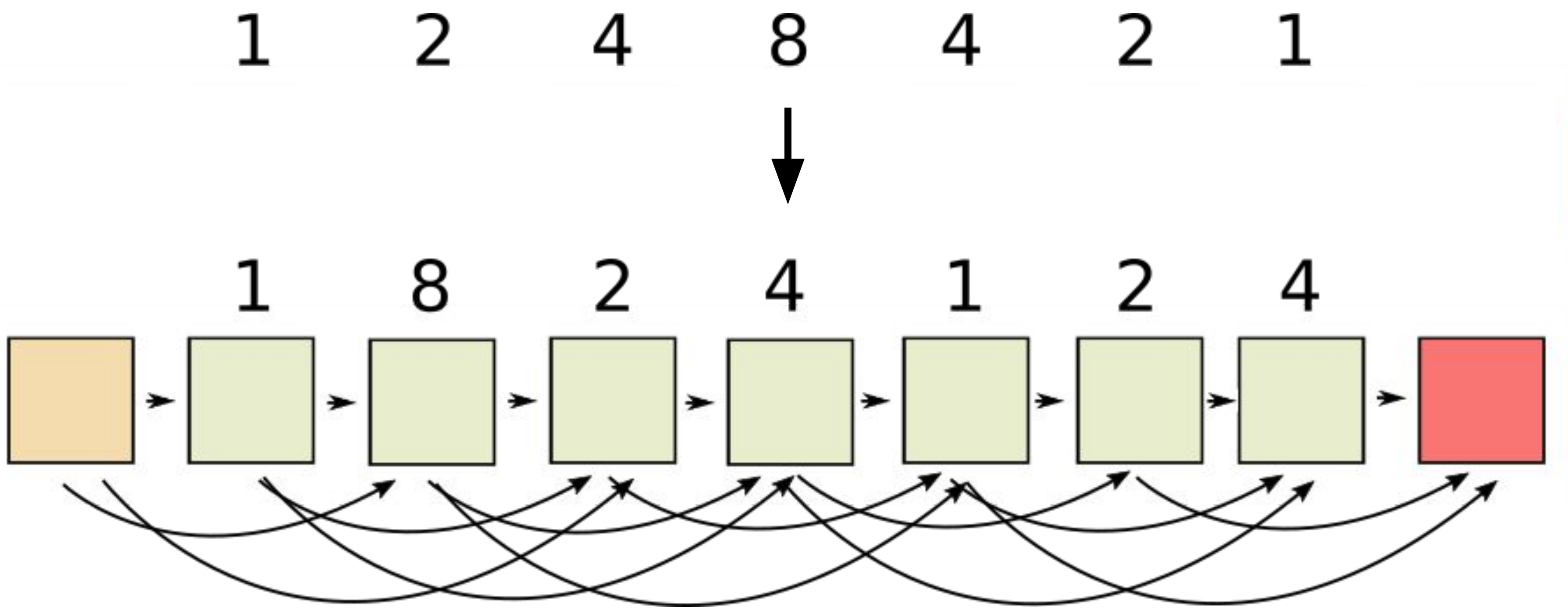
Solution 2: dense connections

- No scaling: all intermediate images are same size
- Why limit input to previous layer only?
- In fact, can use all previous images as input!
  - Including input images
- To compute output, can use all intermediate images



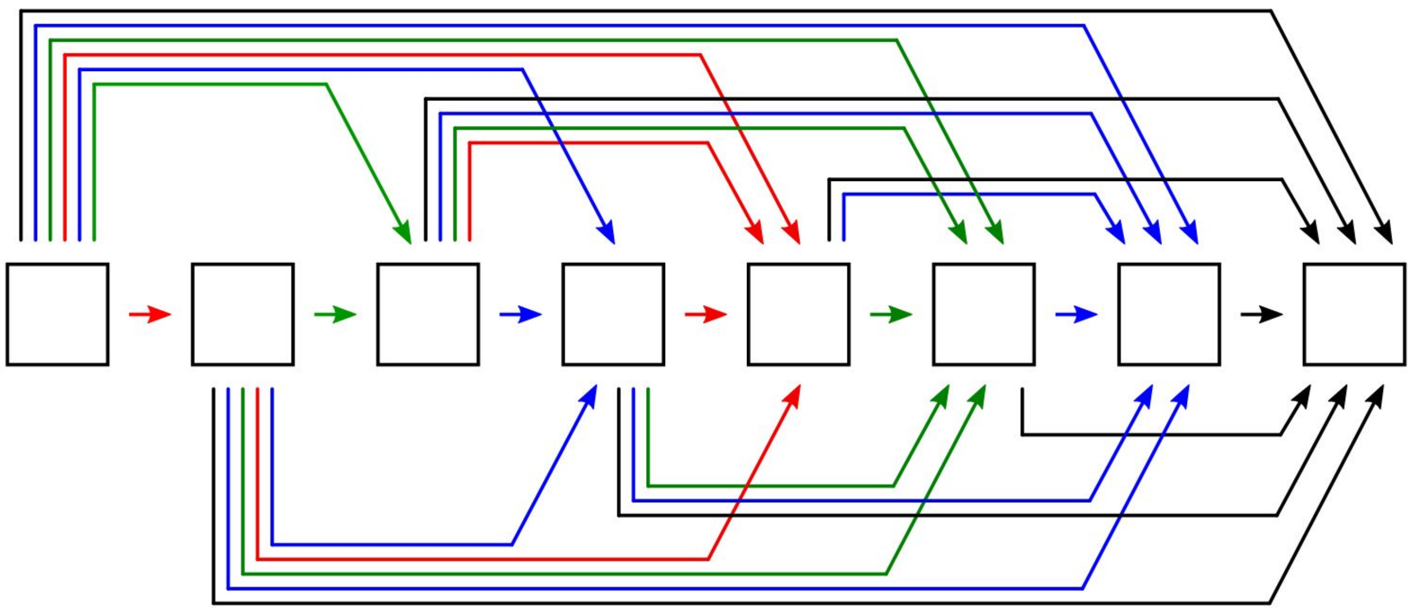
Solution 3: mix scales

- Use small and large dilations throughout the network
- Result: encoder and decoder are effectively mixed



Mixed-scale dense network

- Dilated convolutions instead of scaling
- Each layer is a single image
- All previous images (including input) are used

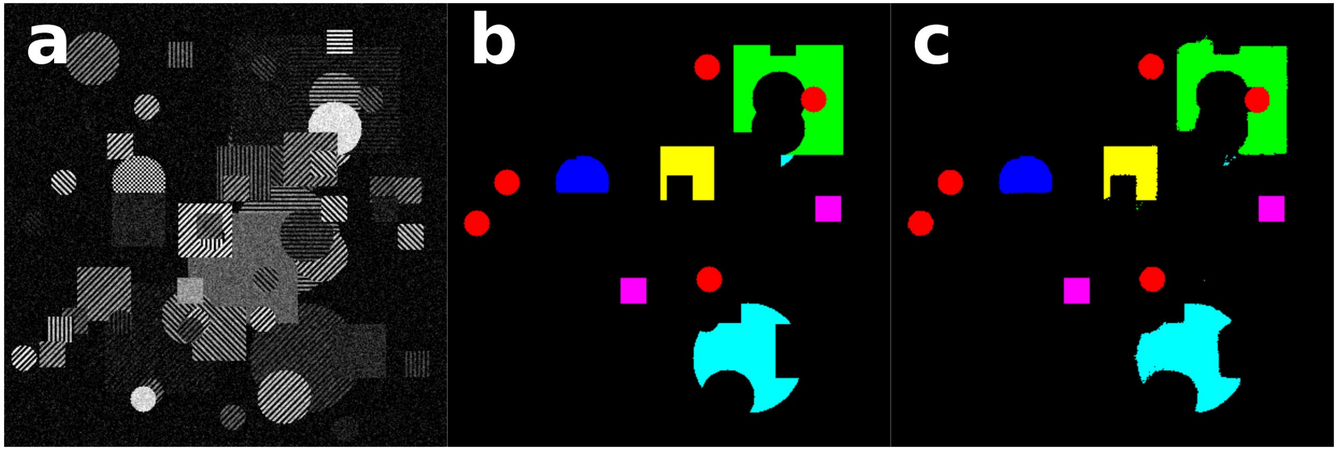


(Pelt & Sethian, PNAS 2018)

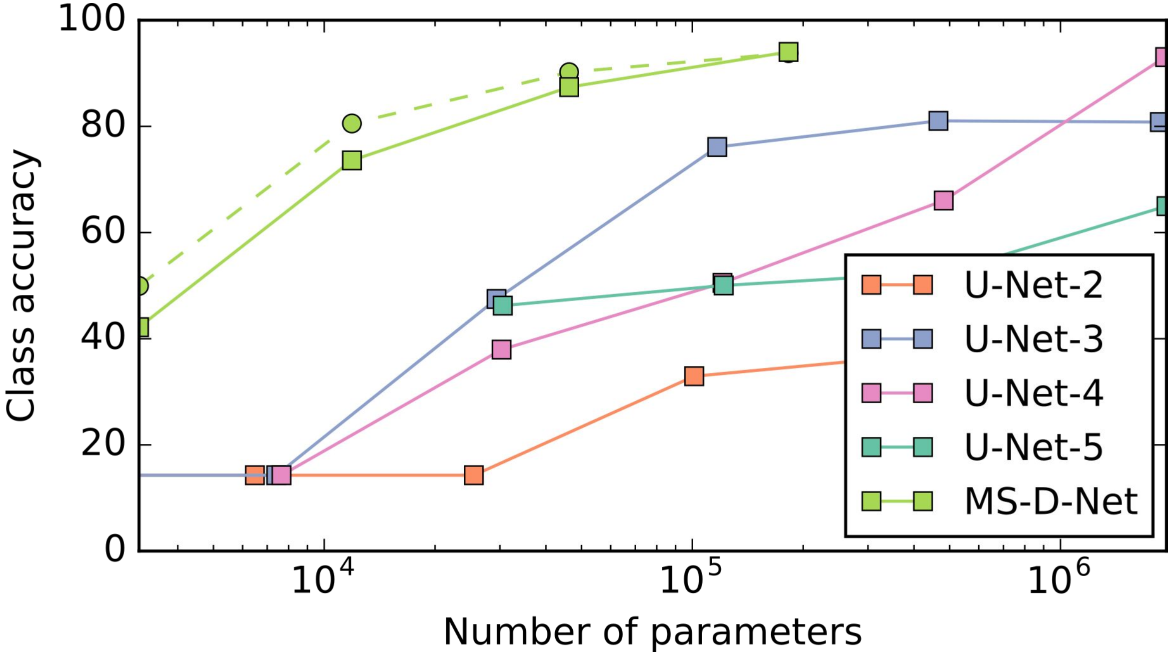
Advantages

- Maximum reuse of features
  - Detected feature is directly usable in deeper layers
- Fewer intermediate images are needed
  - Larger images can be efficiently processed
- Fewer parameters are required
  - Accurate training with few training examples
- Network can learn how to combine different dilations
  - Same network can be applied to different problems

Results - simulations

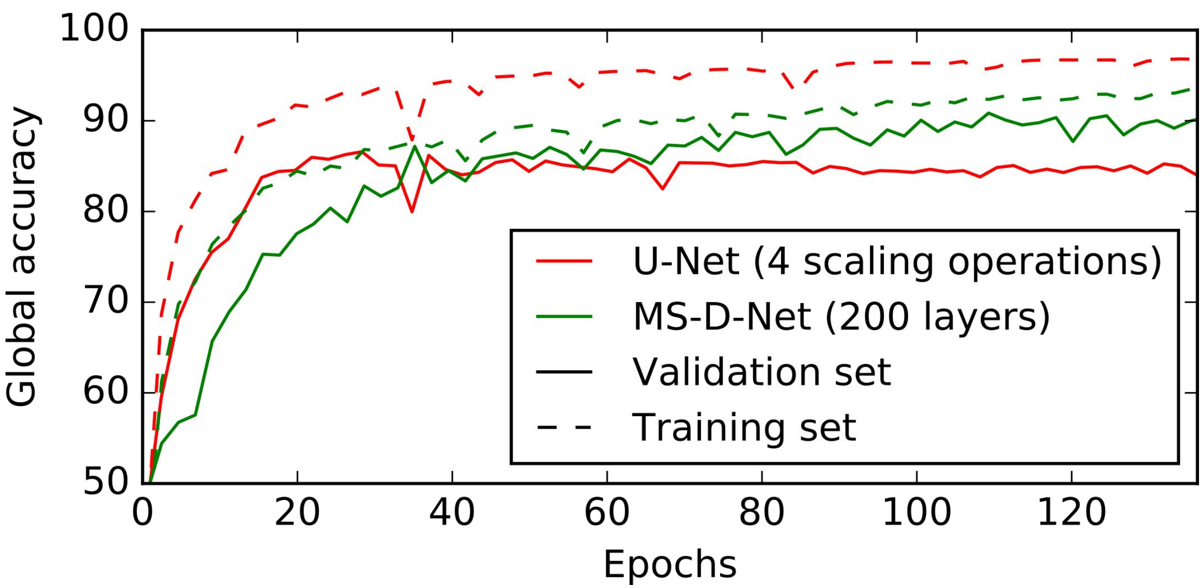


- 512x512 pixels
- 36 combinations
  - shape, size, texture
- Detect 6 combinations



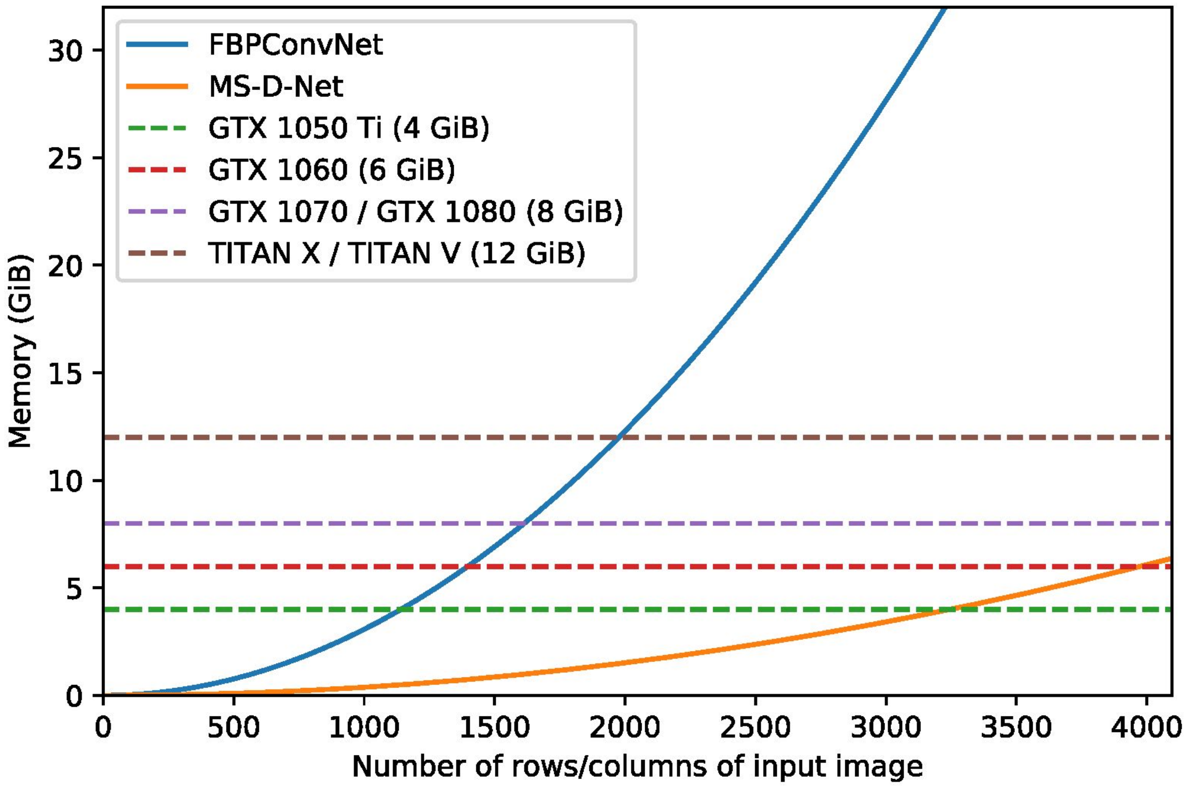
Results - road scene segmentation

Method	Pars (M)	GA	CA
MS-D-Net (100 layers)	**0.048**	85.1	56.8
MS-D-Net (200 layers)	0.187	**87.0**	**63.9**
U-Net (3 scaling operations) (5)	1.863	83.2	50.4
U-Net (4 scaling operations) (5)	1.926	85.5	48.4
SegNet-Basic-EncoderAddition (4)	1.425	84.2	56.5
SegNet-Basic (4)	1.425	84.0	54.6
Boosting + Detectors + CRF (31)		83.8	62.5
Super Parsing (32)		83.3	51.2

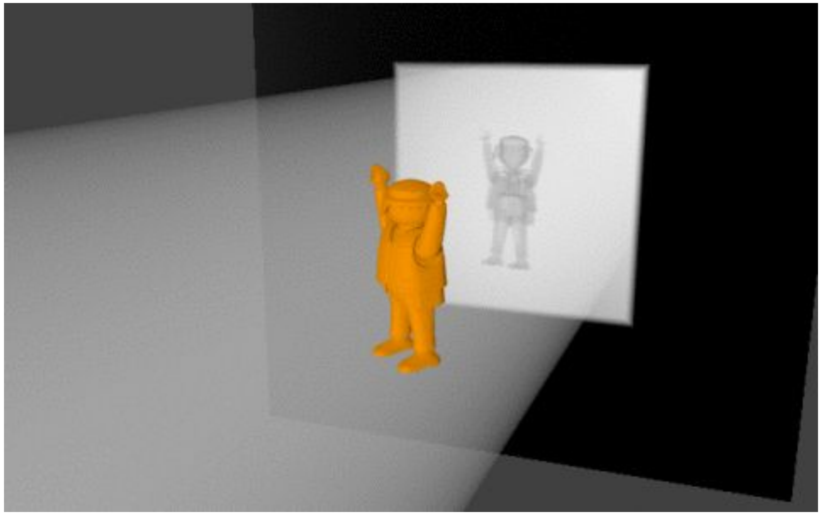


Computational requirements

- All layers are computed in the same way
  - No additional layers/operations (e.g. batchnorm, dropout)
  - Implementation is relatively easy

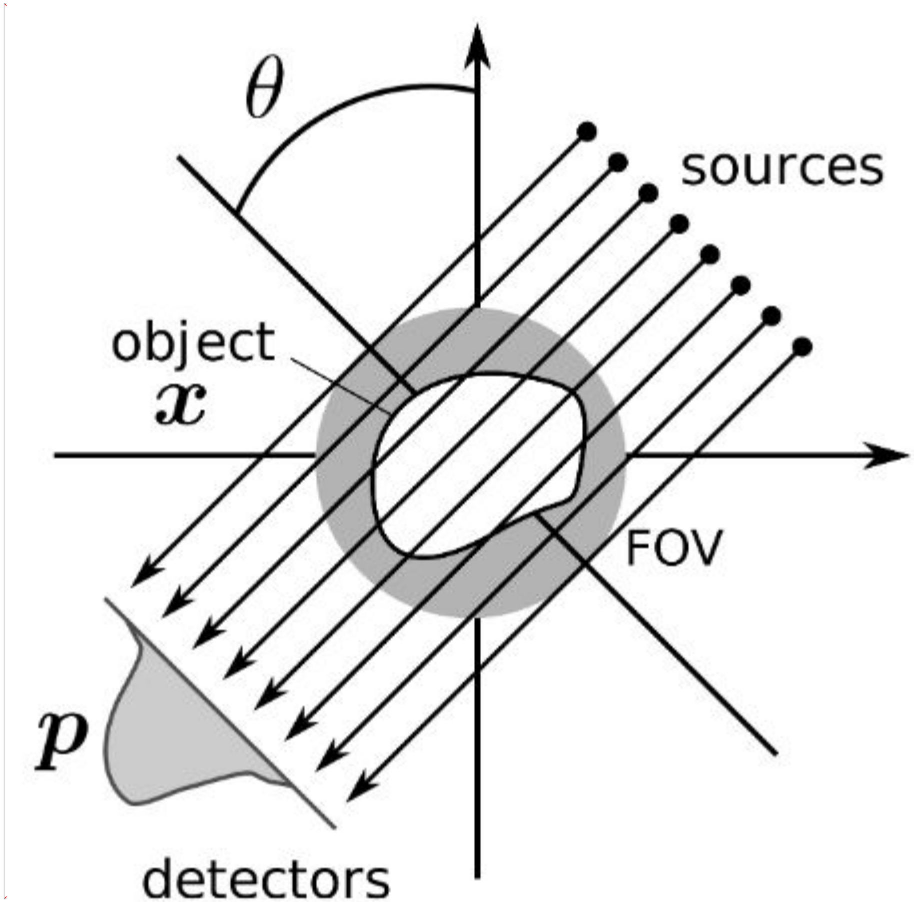


Tomography



Goal: Recover 3D interior from 2D X-ray images

Tomographic reconstruction



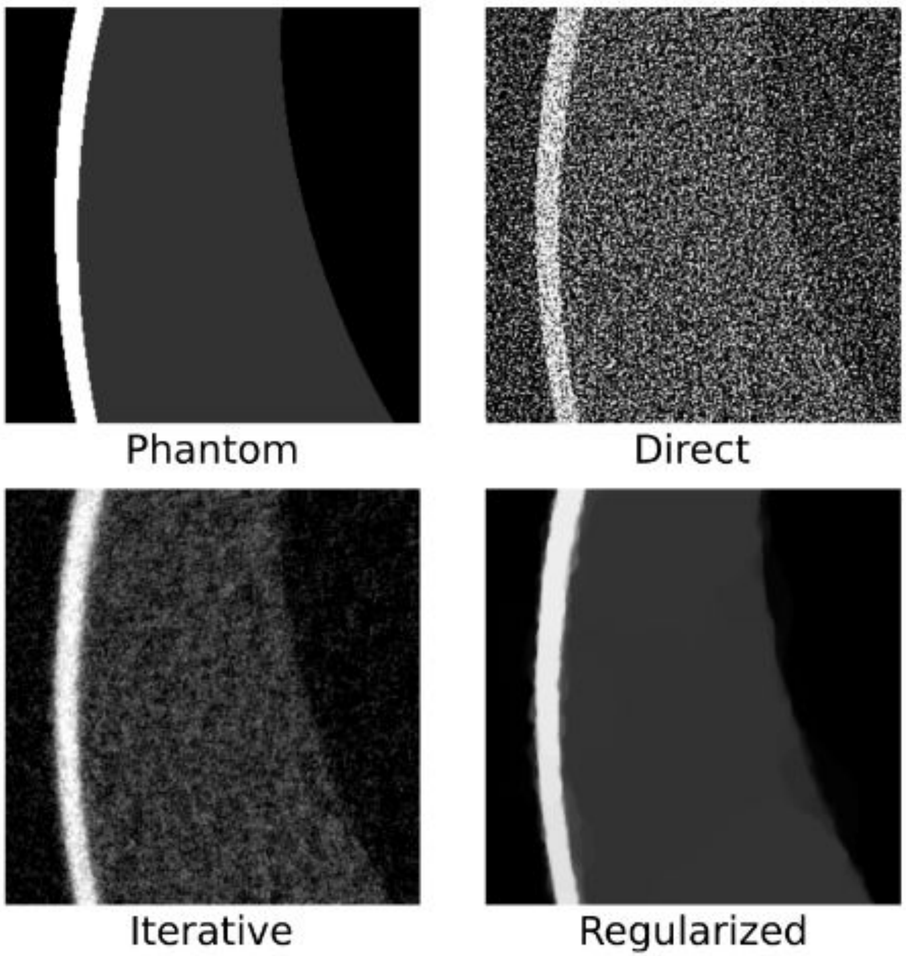
- Unknown object x is scanned by penetrating waves
- Projections p are measured
- Using p, we aim to reconstruct x
- Acquisition is modeled by system matrix W

$$W x = p$$

- Problems: system is ill-posed, underdetermined, and huge

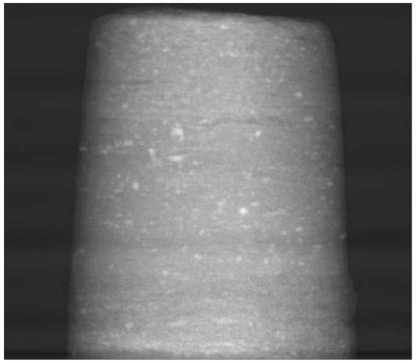
CWI

Tomographic reconstruction



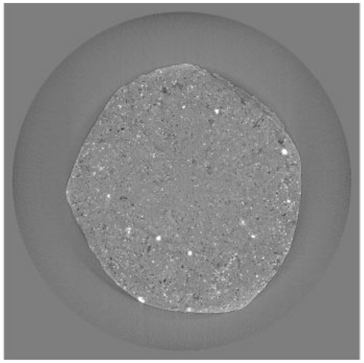
- Direct algorithms
  - FBP, FDK, ...
  - Analytical inversion
- Iterative algorithms
  - SIRT, ART, CGLS, ...
  - Solve linear system
- Regularized methods
  - TV-min, wavelets, …
  - Add regularization term

Problem statement



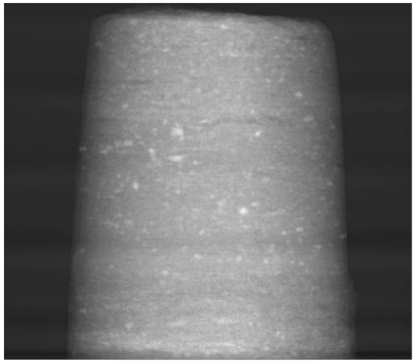
HQ data

easy



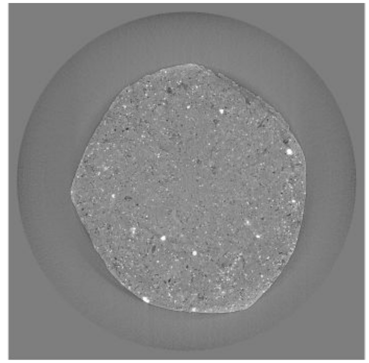
HQ image

Problem statement



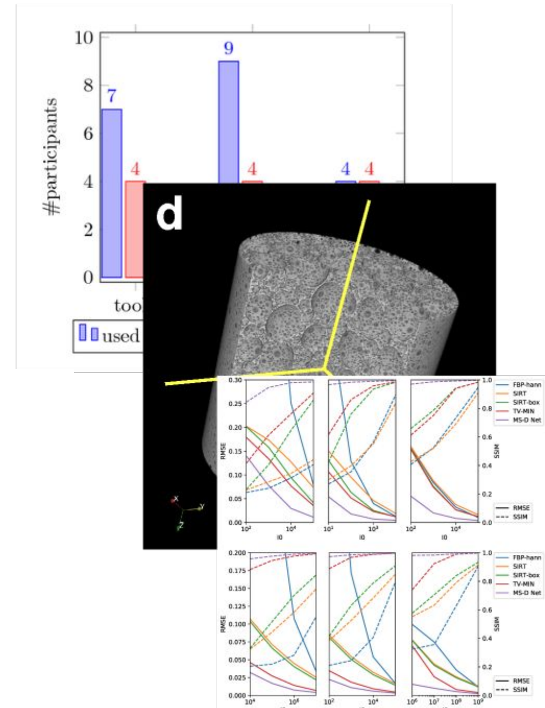
HQ data

easy



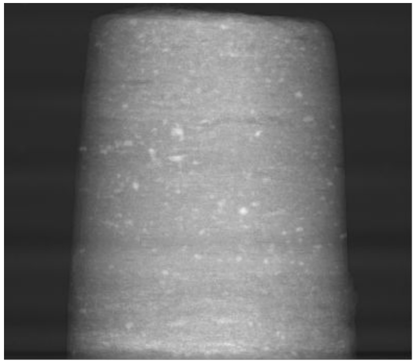
HQ image

hard



Analysis

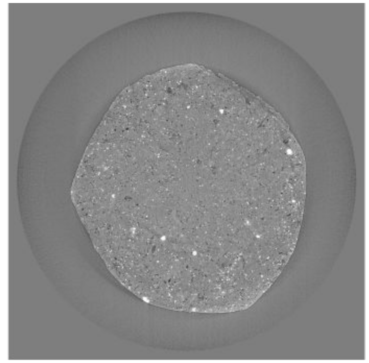
Problem statement



HQ data

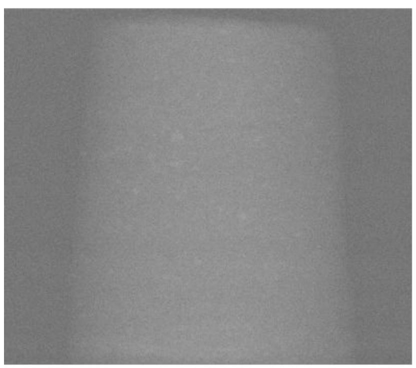
easy

hard

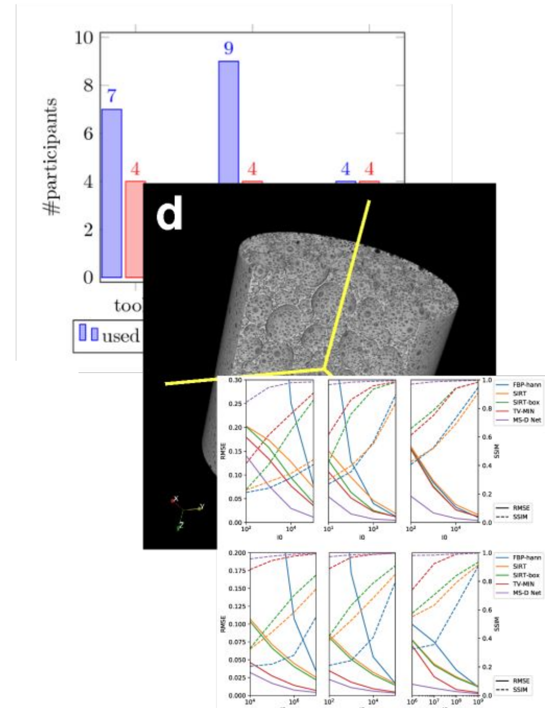


HQ image

hard

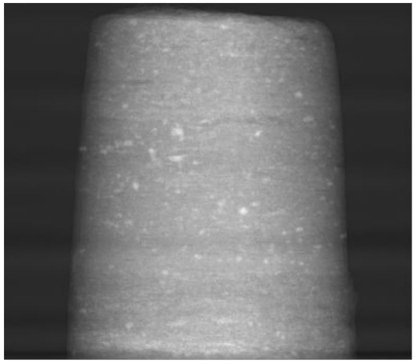


LQ data



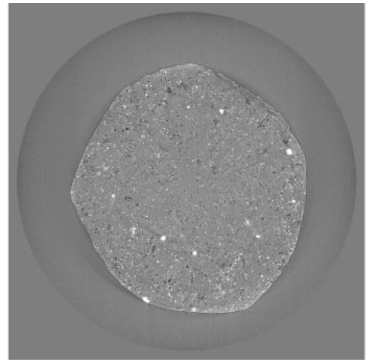
Analysis

Problem statement



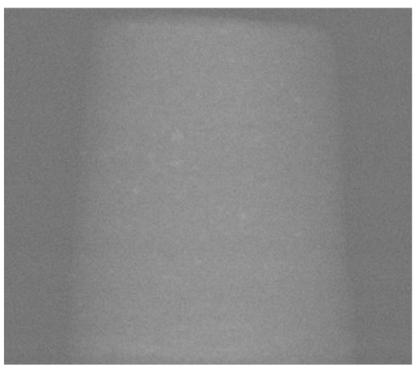
HQ data

easy



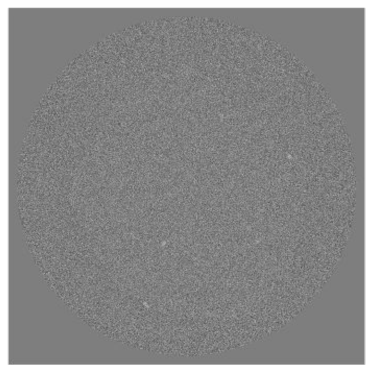
HQ image

hard



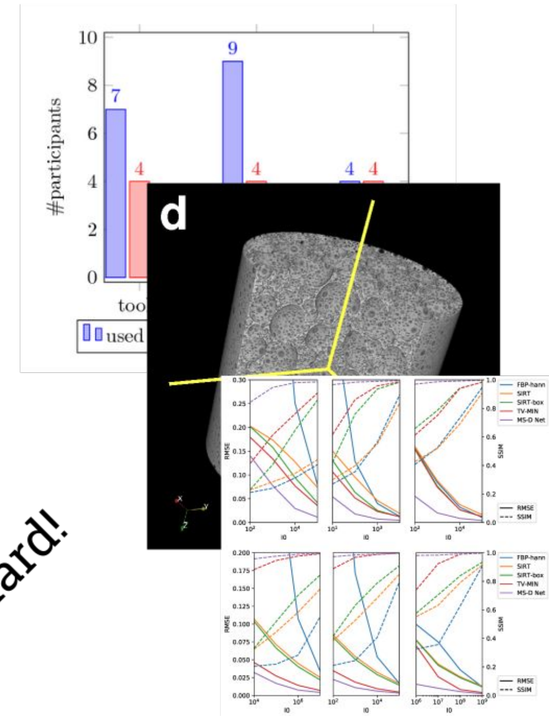
LQ data

easy



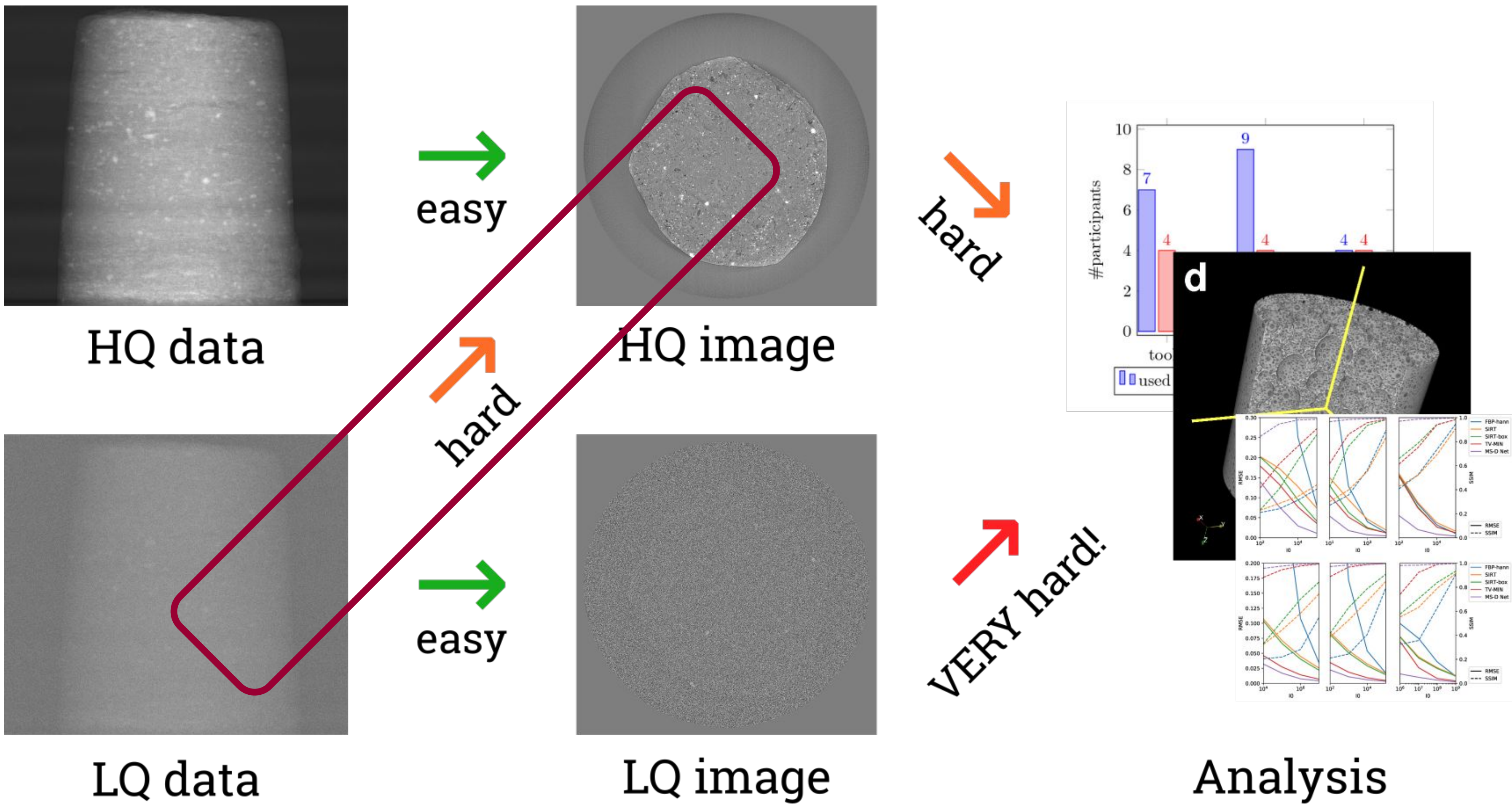
LQ image

VERY hard!

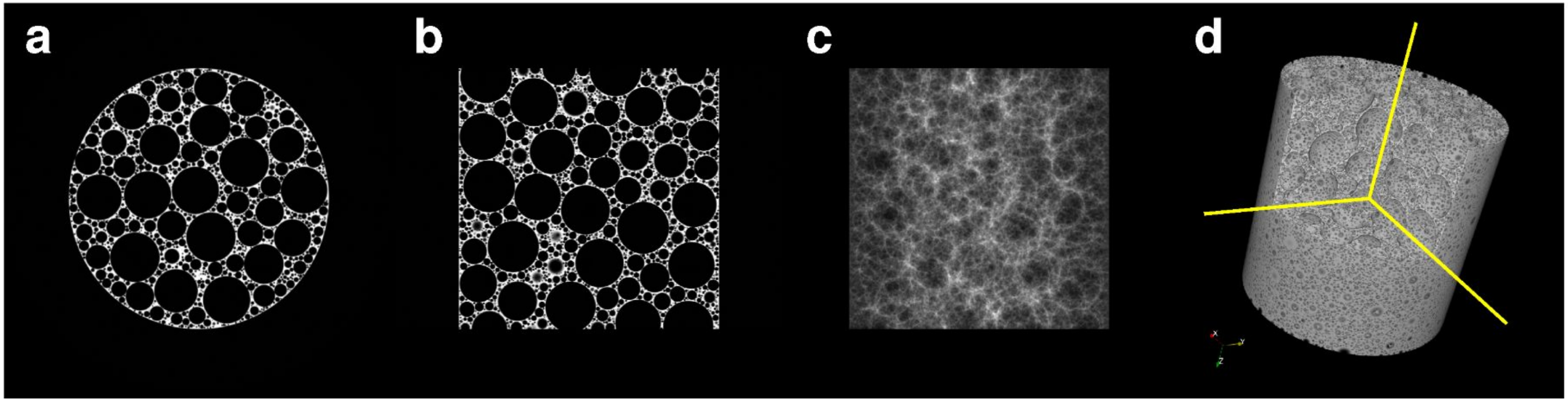


Analysis

Problem statement

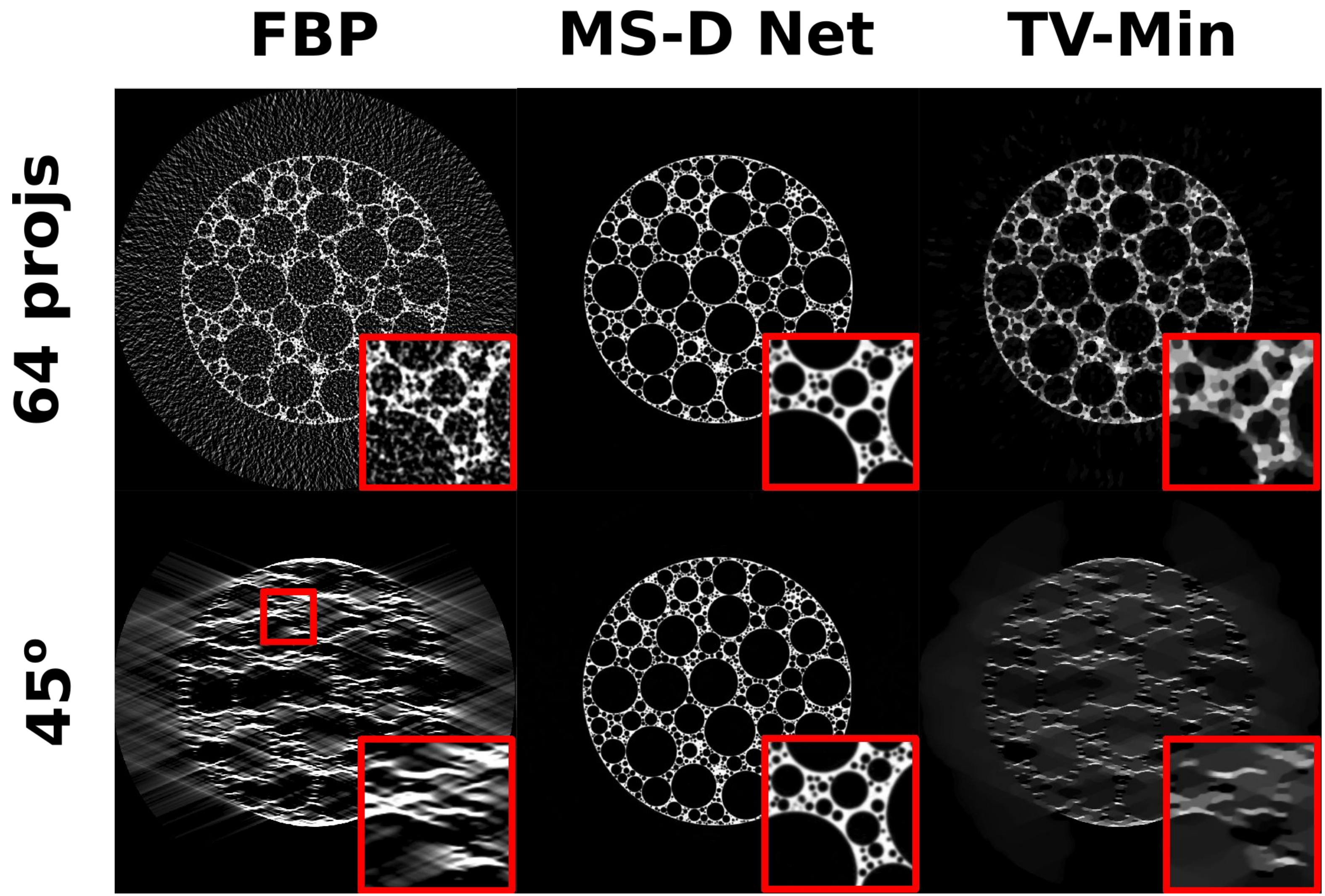


Experiments



- Foam-like simulated objects
- Generate three 3D objects (1024^3 voxels):
  - 1 for training, 1 for validation, 1 for testing
- Use high-quality rec as target for training
- Use low-quality rec as network input

Projections & angular range



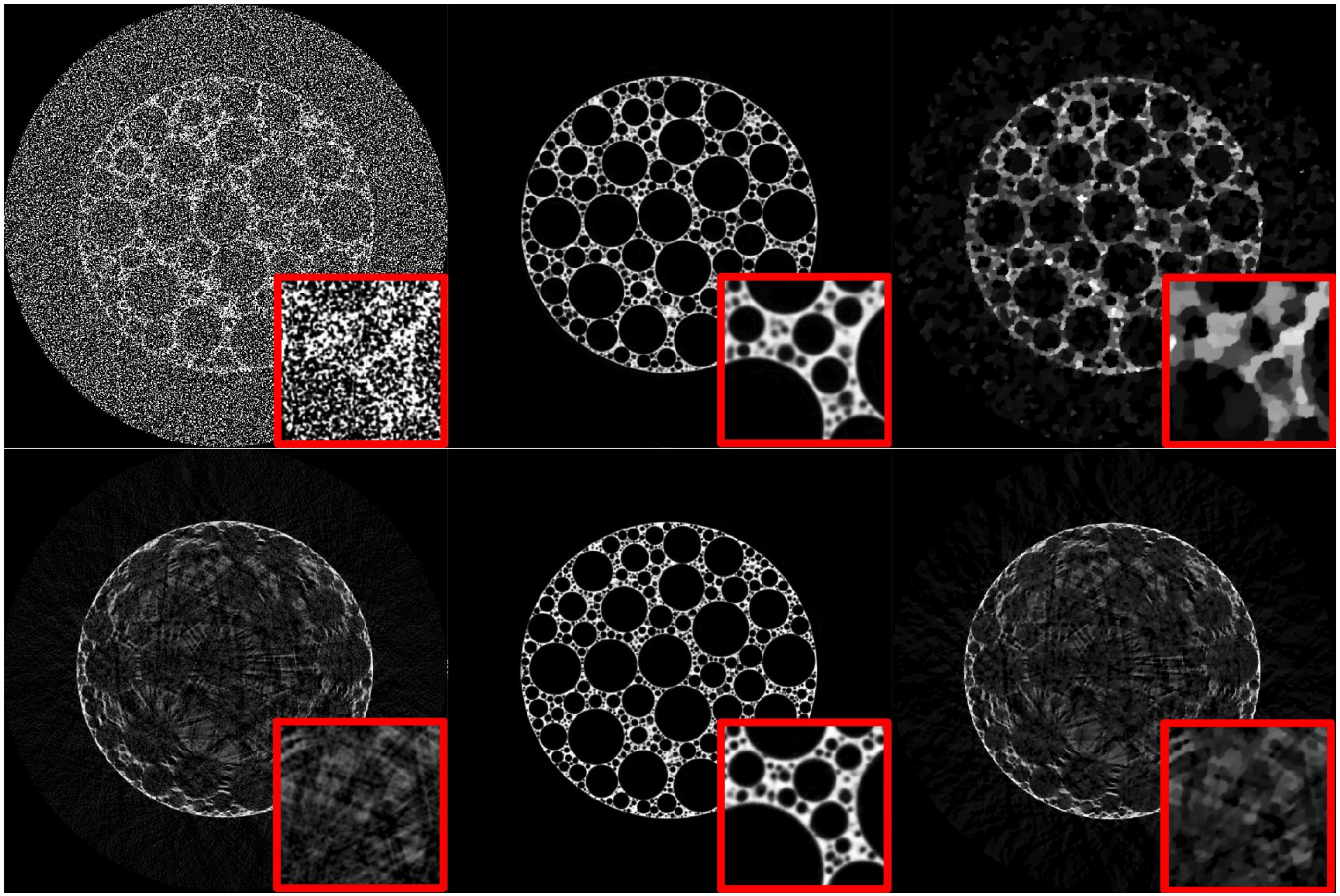
(Pelt et al, J. Imaging 2019)

Noise

FBP | MS-D Net | TV-Min

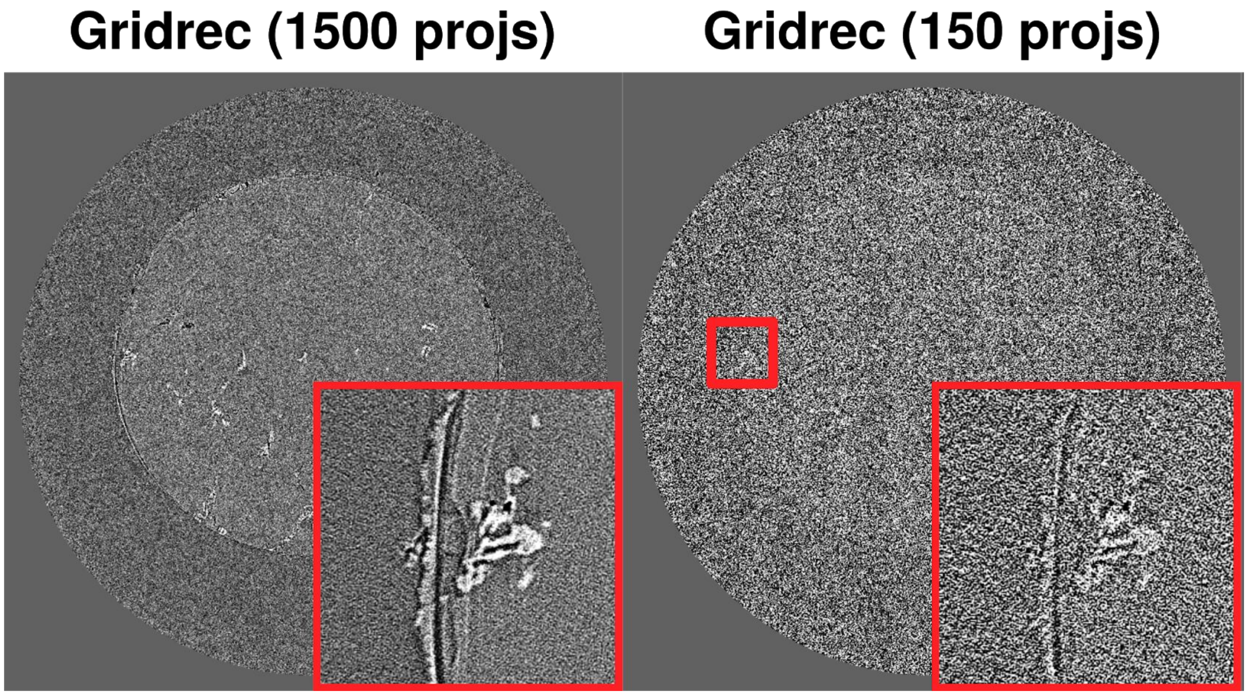
Low abs.

High abs.



(Pelt et al, J. Imaging 2019)

Real-world data

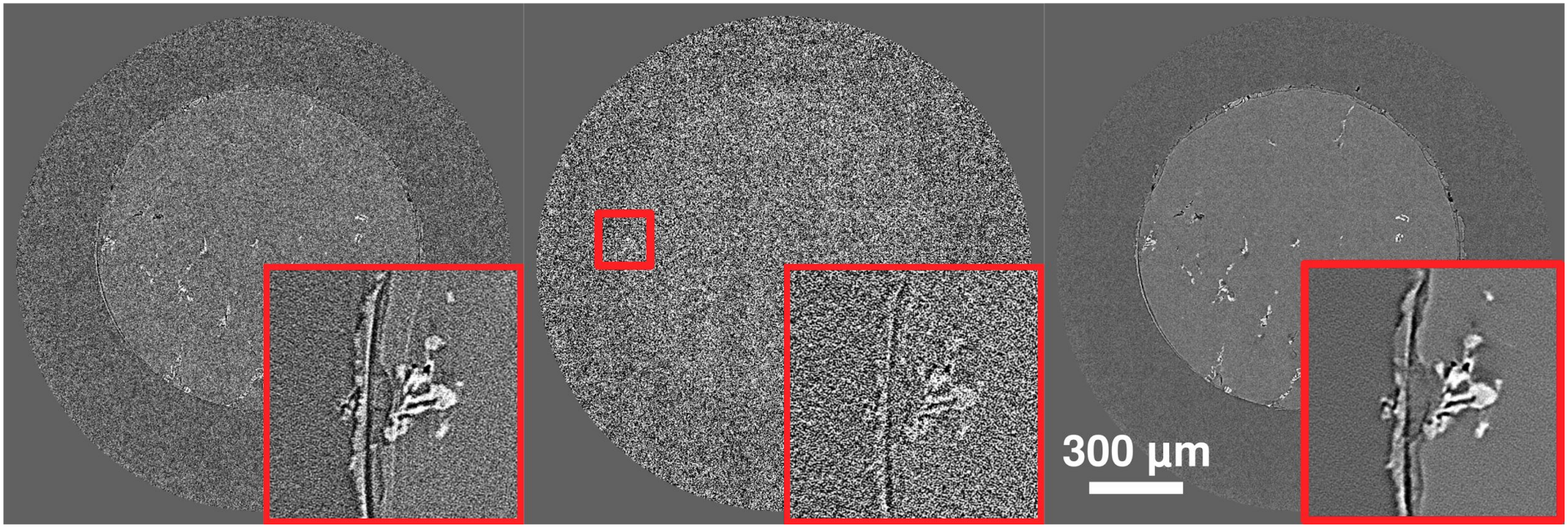


- Fatigue-corrosion data (2160x2560x2560 voxels)
- Use first and last scans as training data
- Shown is an intermediate scan

(Pelt et al, J. Imaging 2019)

Real-world data

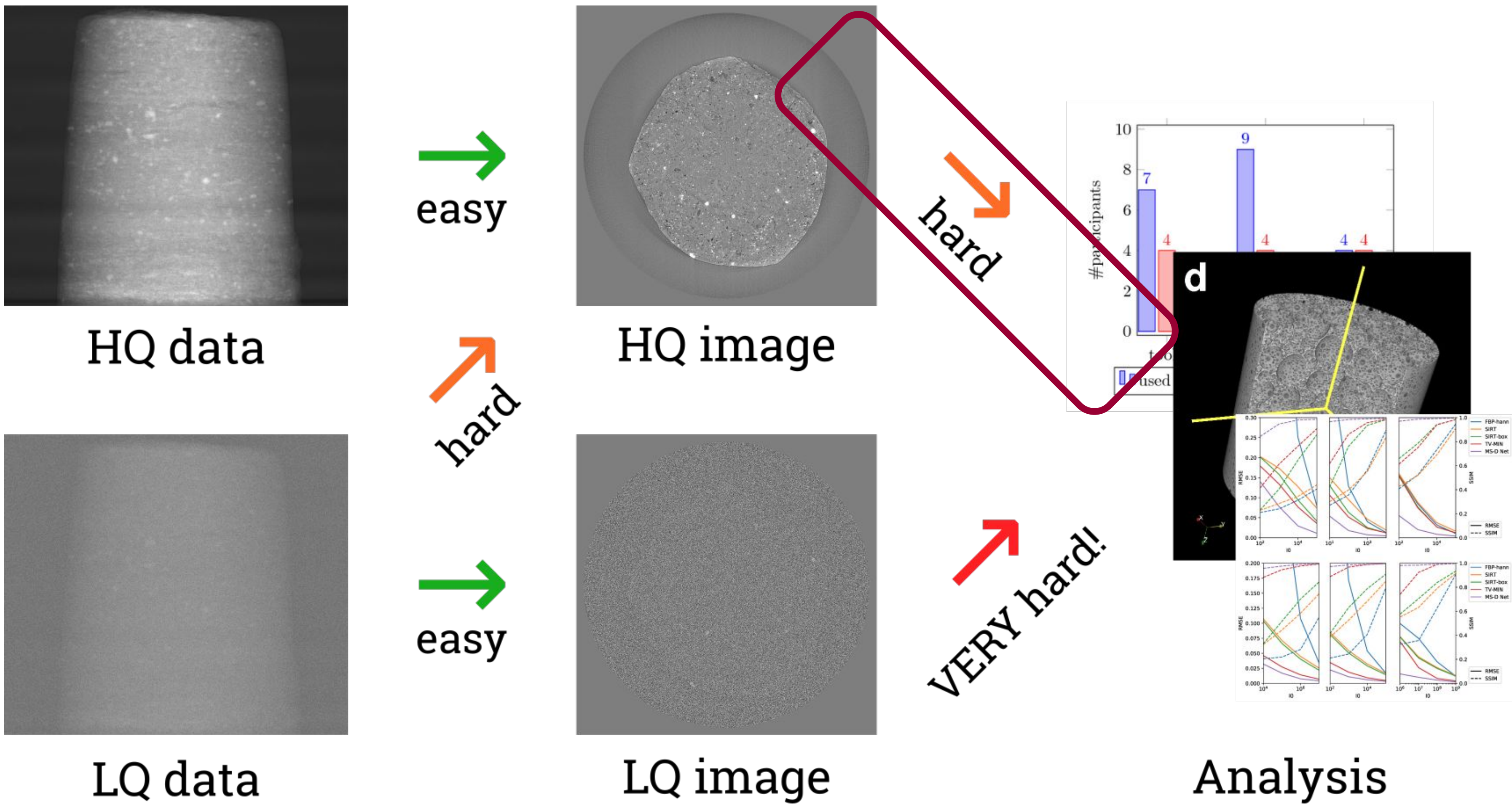
Gridrec (1500 projs) **Gridrec (150 projs)** **MS-D Net (150 projs)**



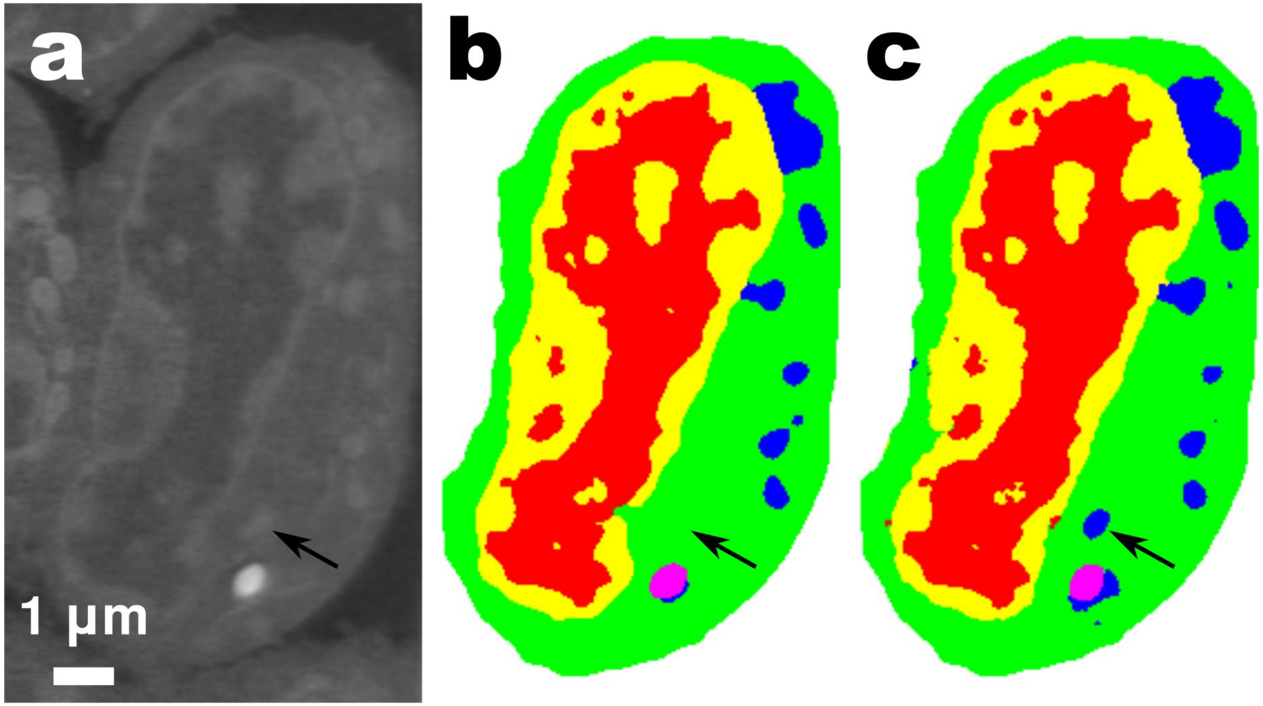
- Fatigue-corrosion data (2160x2560x2560 voxels)
- Use first and last scans as training data
- Shown is an intermediate scan

(Pelt et al, J. Imaging 2019)

Problem statement



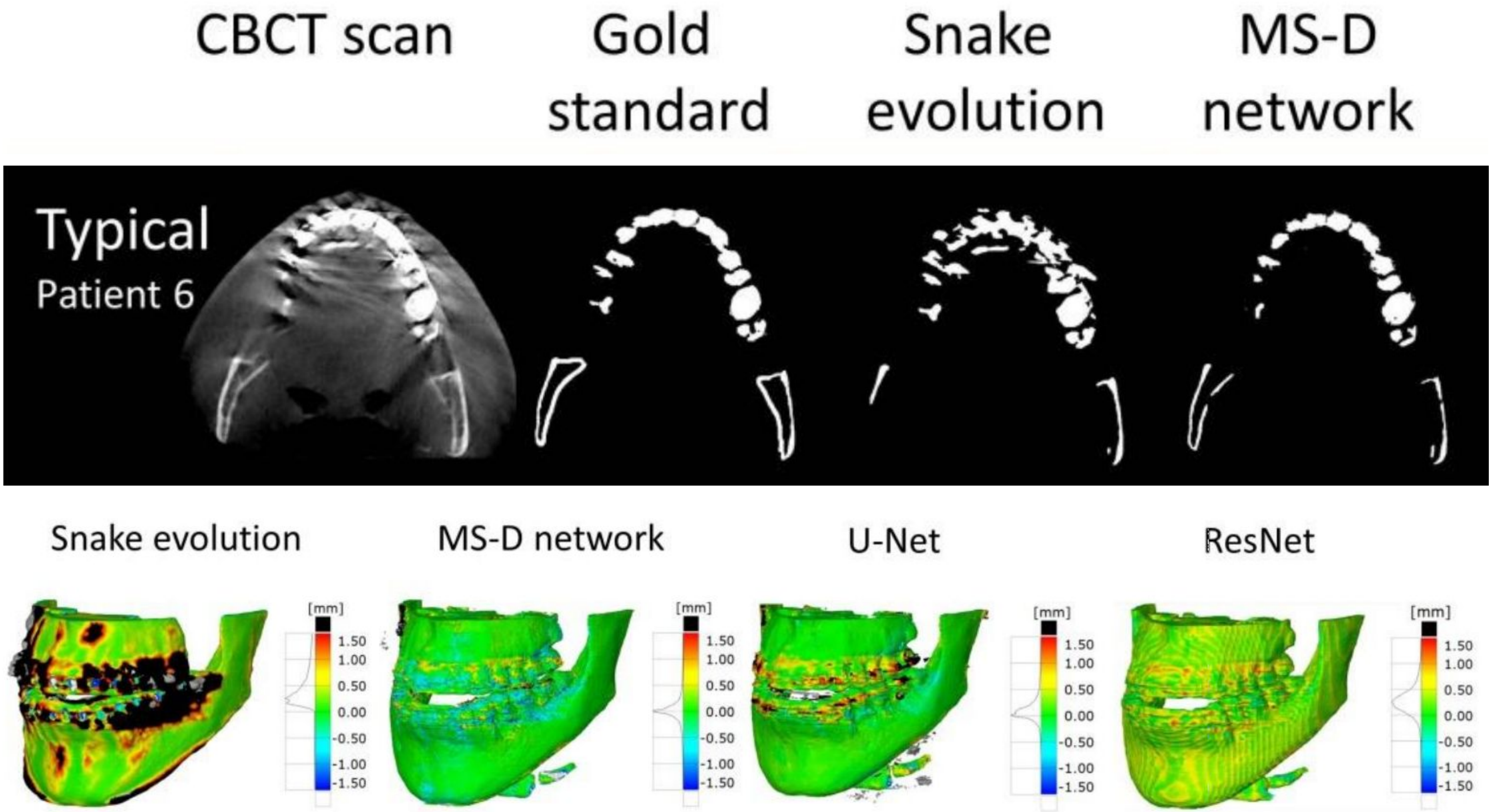
Analysis - cell labeling



- Goal: detect cell structures
  - Nuclear envelope, mitochondria, …
- Use eight manually labeled 512^3 volumes
  - 6 for training, 1 for validation, 1 for testing

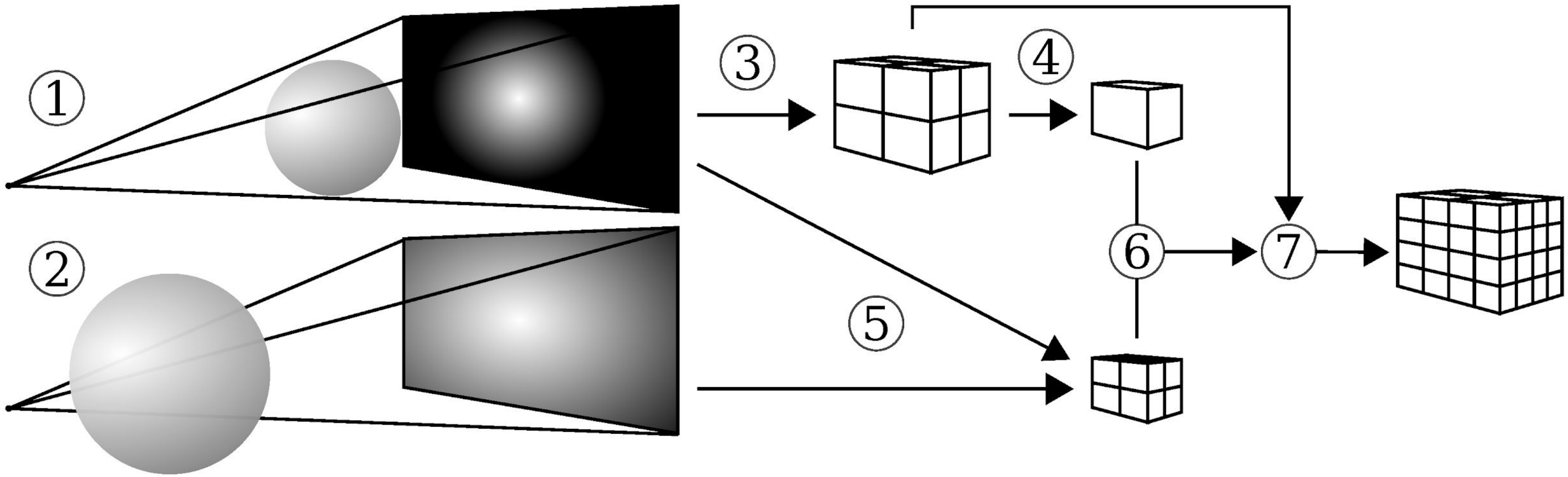
(Pelt & Sethian, PNAS 2019)

Analysis - dental CT



(Minnema et al, Med. Phys. 2019)

Single-object ML

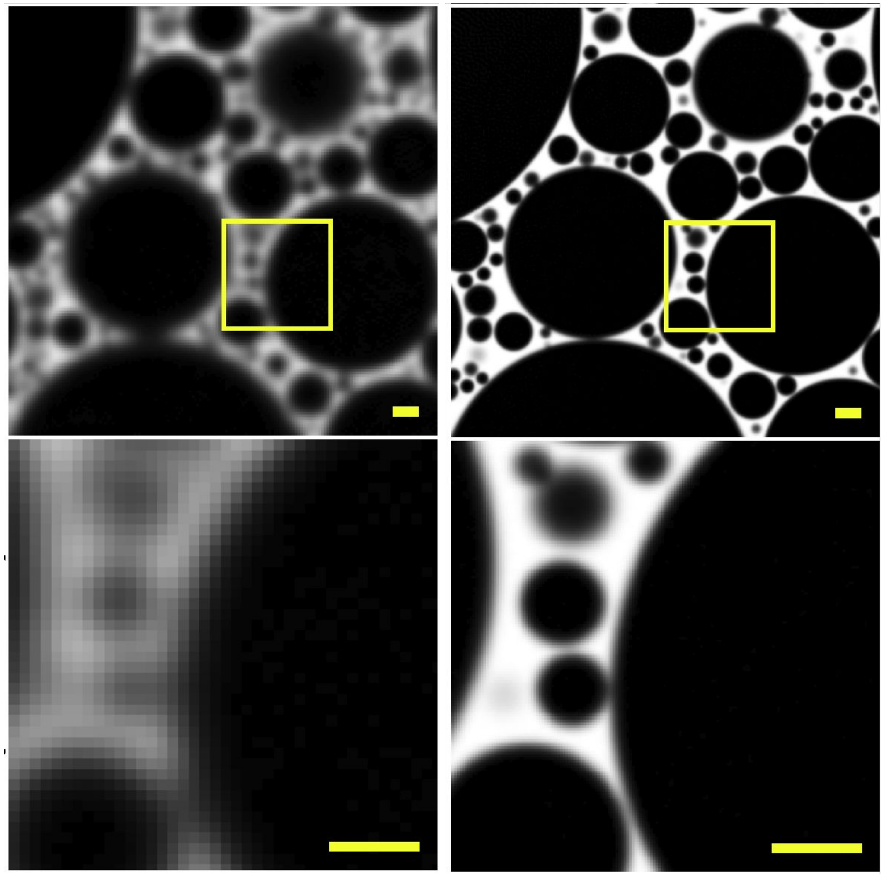


(1) Normal scan	(3) Low-res rec (4) ROI crop	(6) Train network
(2) Zoomed ROI scan	(5) High-res ROI rec	(7) Apply to low-res rec

(Hendriksen et al. 2019)

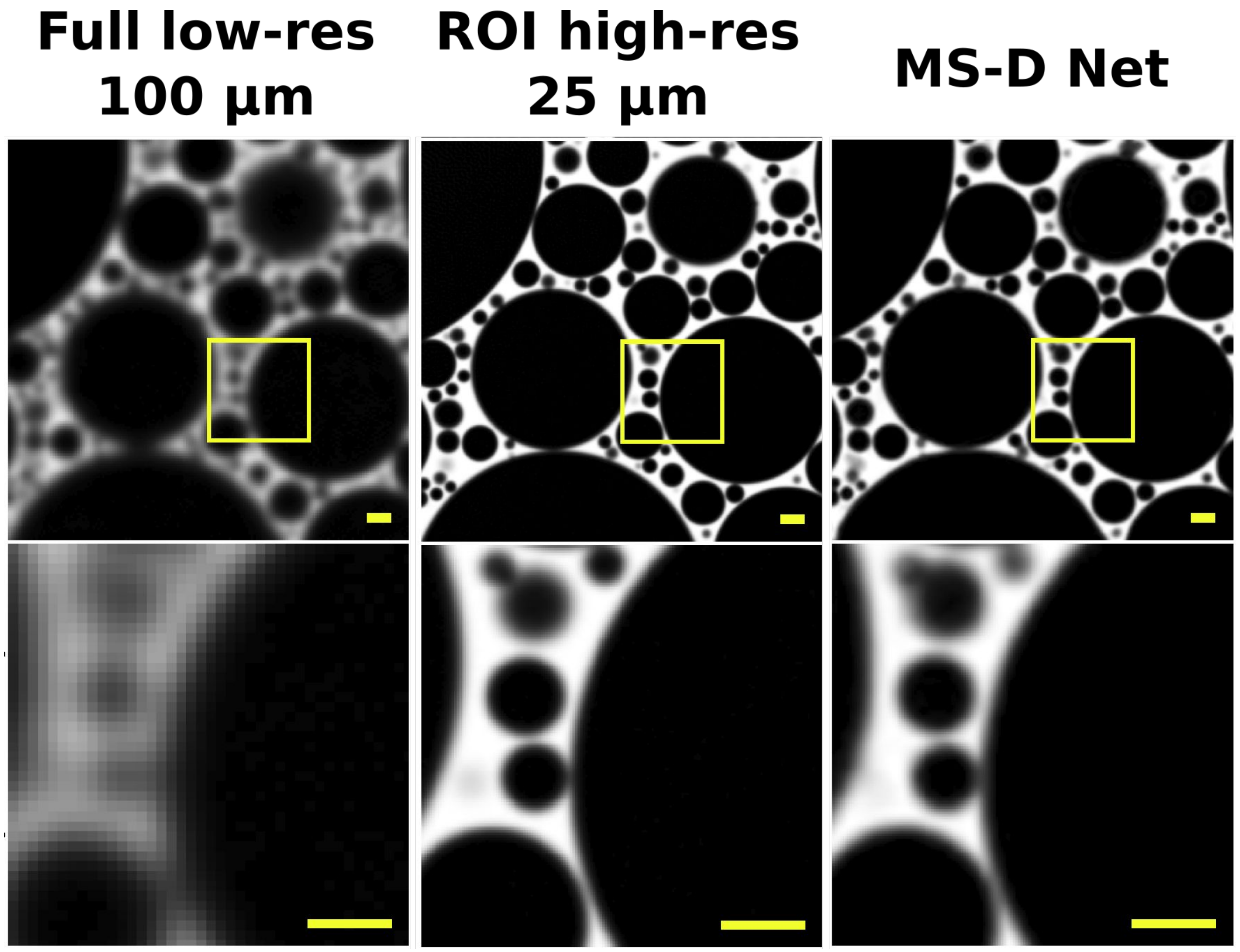
Single-object ML

Full low-res 100 µm | **ROI high-res 25 µm**



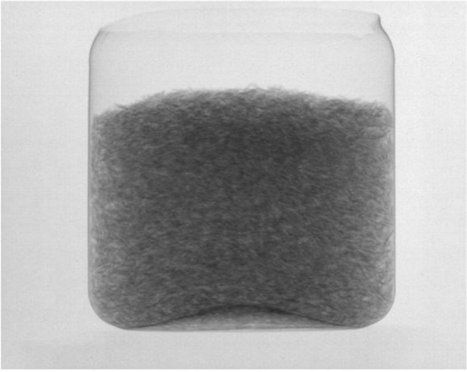
(Hendriksen et al. 2019)

Single-object ML



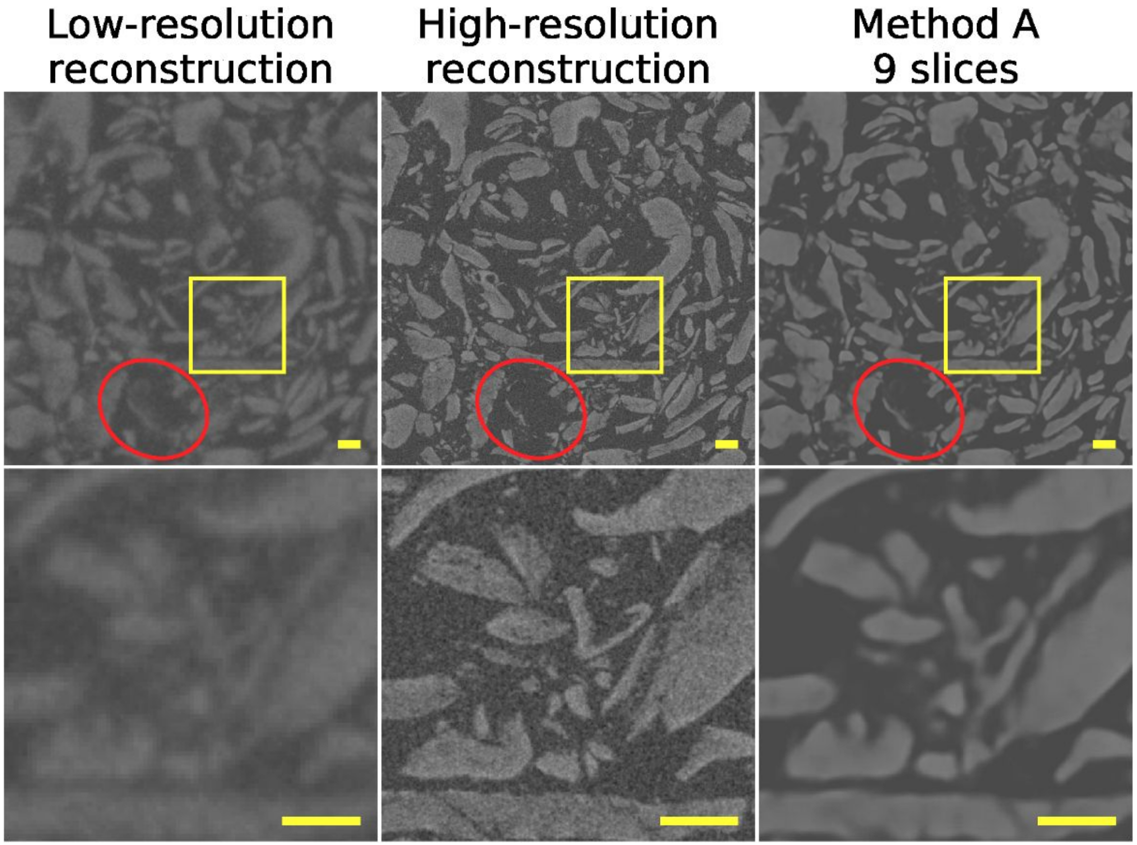
(Hendriksen et al. 2019)

Single-object ML



Low-res: 68.26 µm

High-res: 17.07 µm



(Hendriksen et al. 2019)

Conclusions

- Deep learning has large potential for scientific images
- Popular existing approaches have several problems
- Mixed-Scale Dense network for large scientific images
- Good results for various imaging problems
- Many interesting and exciting challenges remain!

Thank you for your attention!

d.m.pelt@cwi.nl

Pelt, Daniël M., and James A. Sethian. "A mixed-scale dense convolutional neural network for image analysis." Proceedings of the National Academy of Sciences 115.2 (2018): 254-259.

Pelt, Daniël M., Kees J. Batenburg, and James A. Sethian. "Improving tomographic reconstruction from limited data using mixed-scale dense convolutional neural networks." Journal of Imaging 4.11 (2018): 128.

Hendriksen, Allard A., et al. "On-the-Fly Machine Learning for Improving Image Resolution in Tomography." Applied Sciences 9.12 (2019): 2445.

Minnema, Jordi, et al. "Segmentation of dental cone-beam CT scans affected by metal artifacts using a mixed-scale dense convolutional neural network." Medical physics (2019).

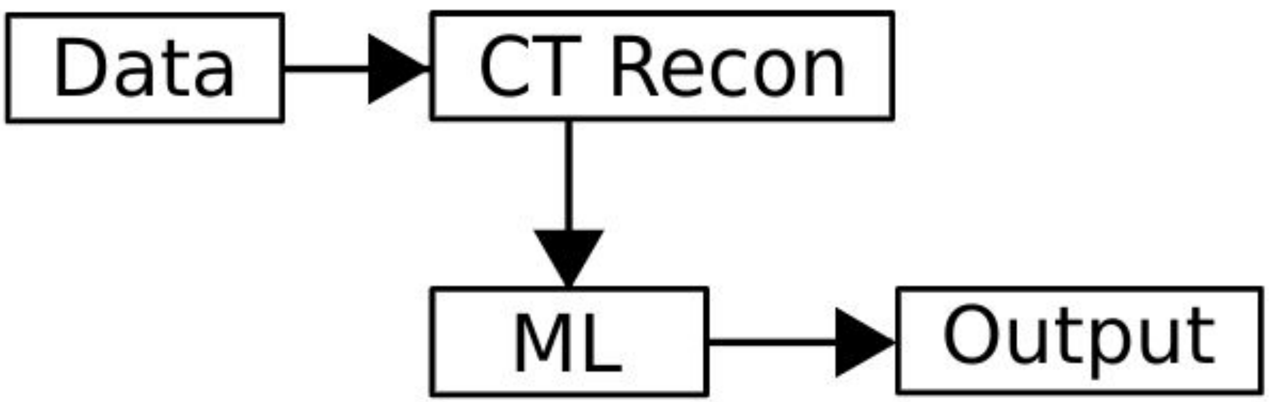
Training example

$\mathbf{x}$ g_ϕ $f(\mathbf{x})$

- Training with 100 similar images

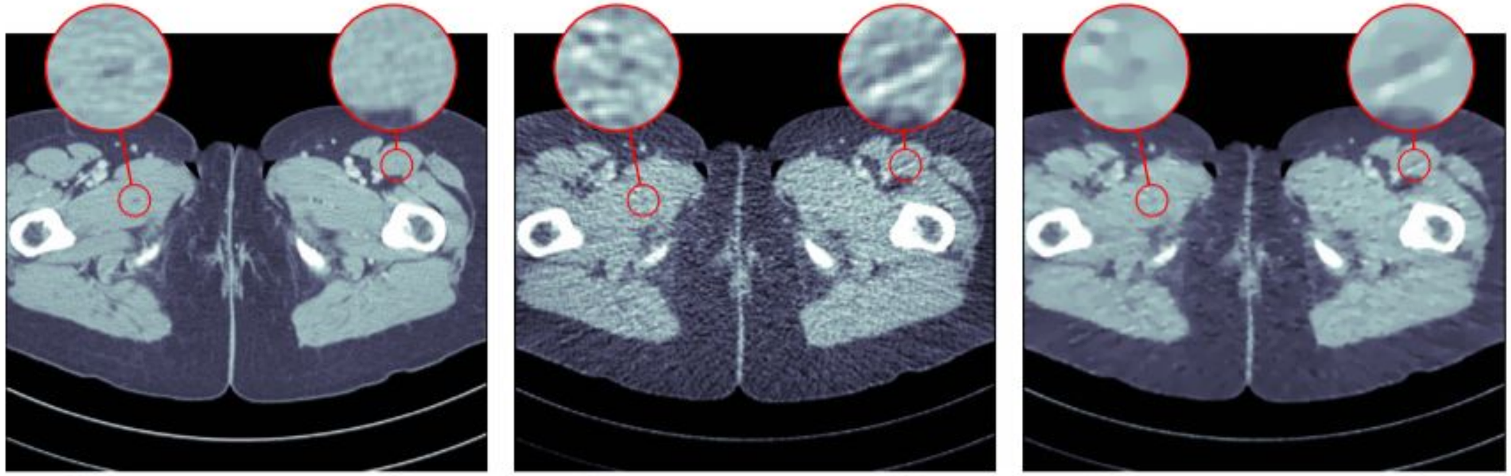
Deep learning for tomography

- Learn parts of existing algorithms (Pelt & Batenburg 2013)
- Learned primal-dual algorithms (Adler & Oktem 2018)
- Use CNN as a post-processing operation (Jin et al 2017)



State-of-the-art: Learned primal-dual

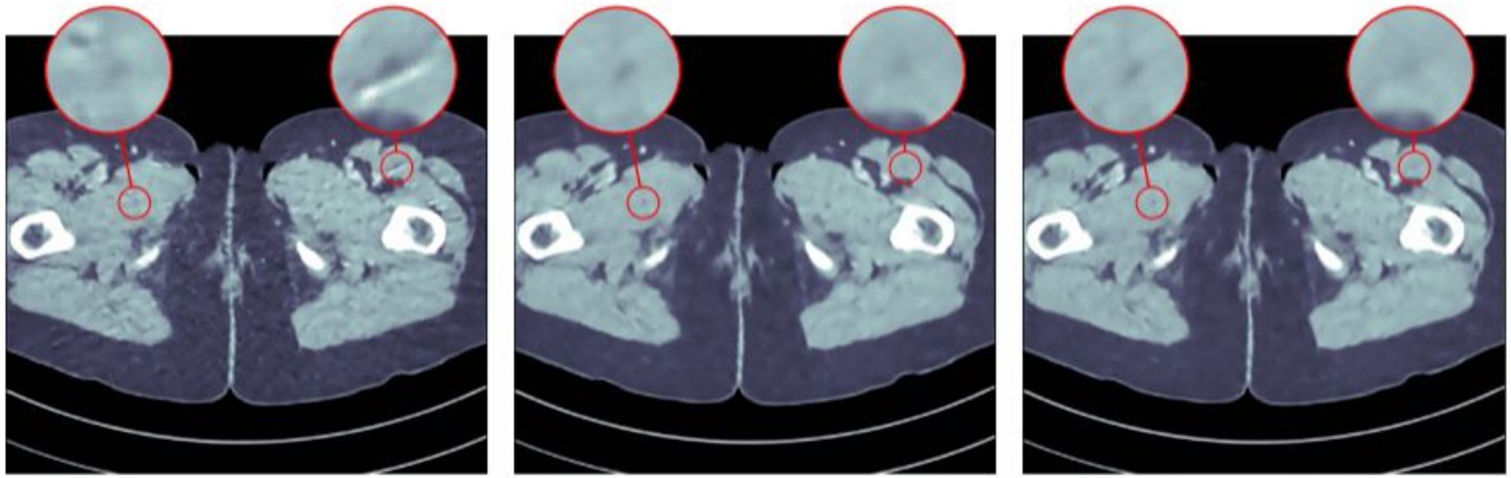
(Adler & Öktem 2018)



(a) 512 × 512 pixel human phantom

(b) Filtered back-projection (FBP)
PSNR 33.65 dB, SSIM 0.830, 423 ms

(c) Total variation (TV)
PSNR 37.48 dB, SSIM 0.946, 64 371 ms

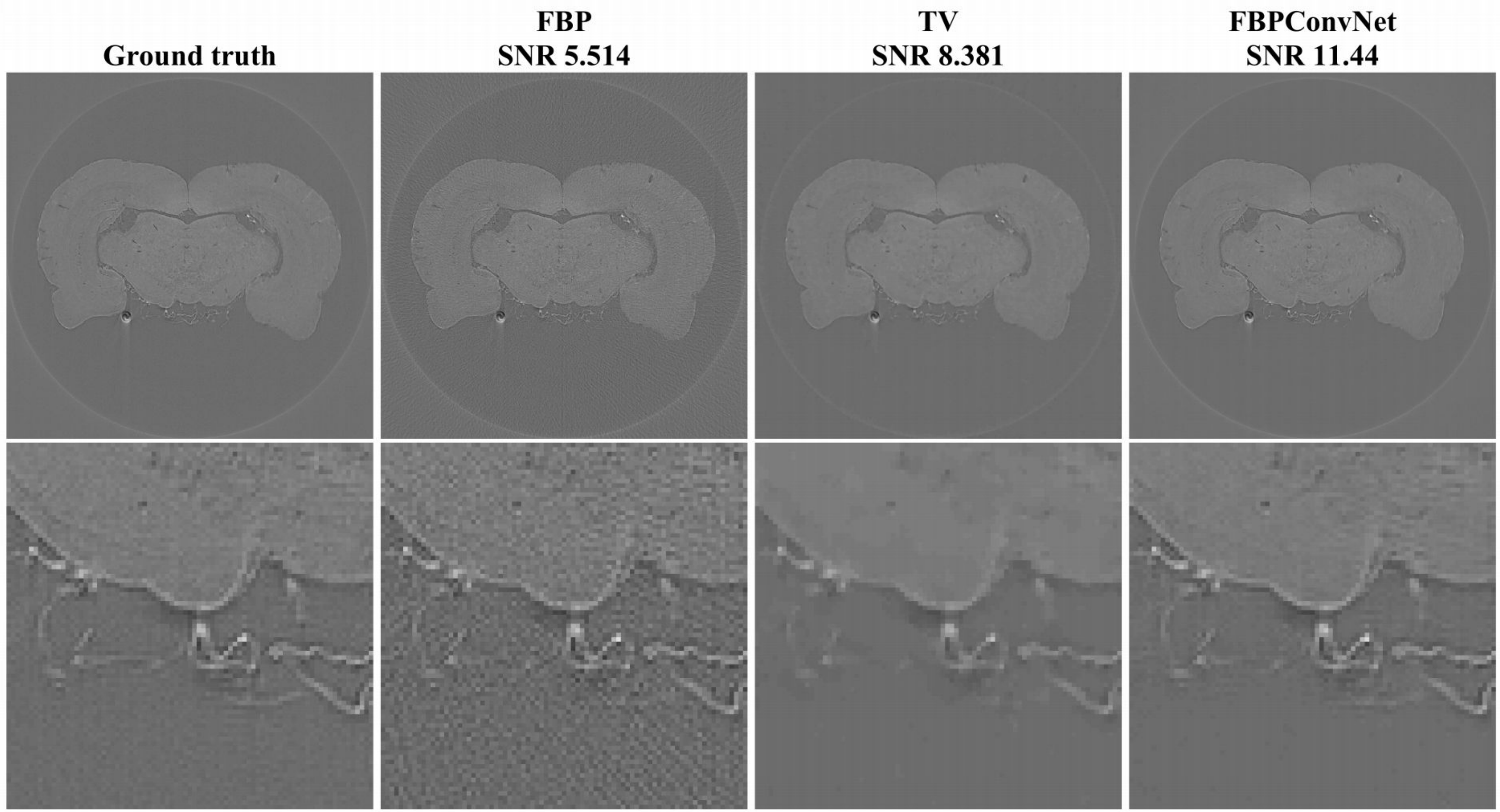


(d) FBP + U-Net denoising
PSNR 41.92 dB, SSIM 0.941, 463 ms

(e) Primal-Dual, linear
PSNR 44.10 dB, SSIM 0.969, 620 ms

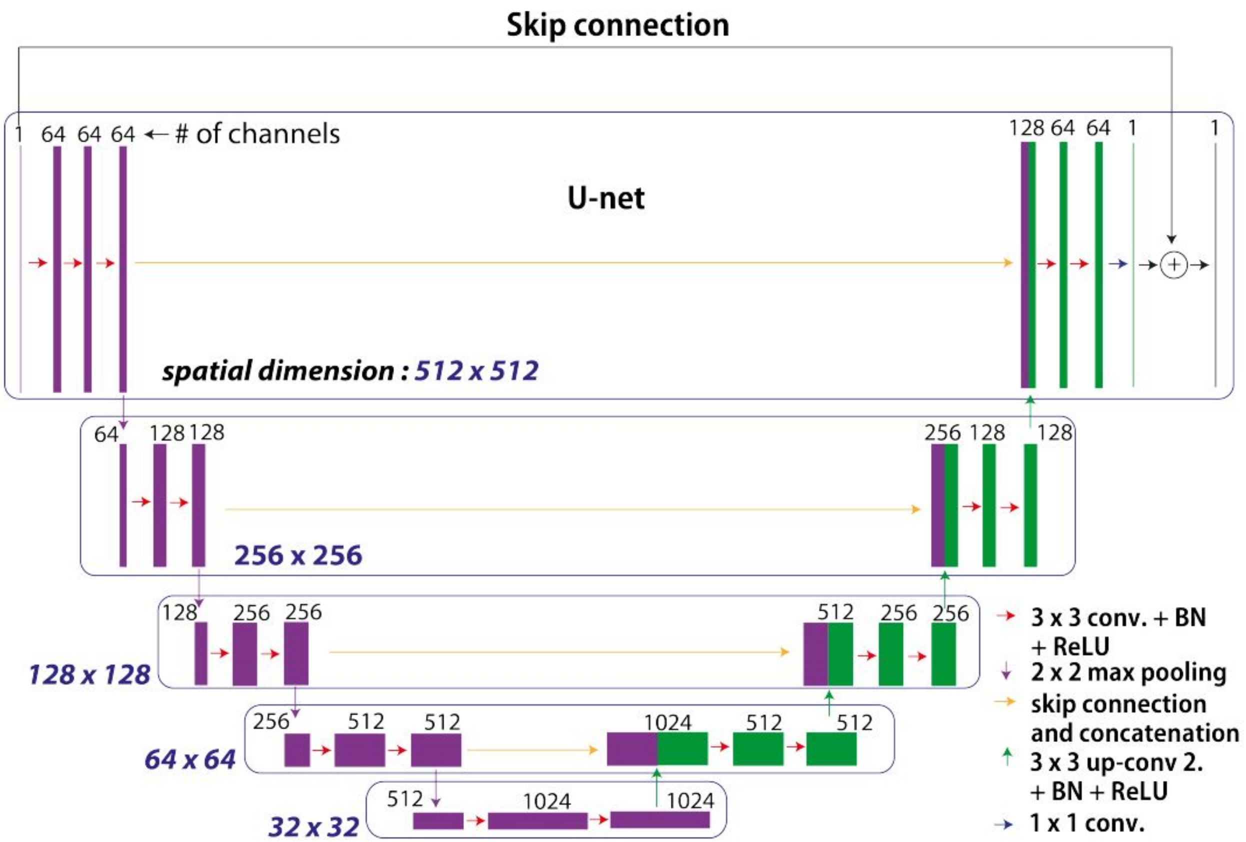
(f) Primal-Dual, non-linear
PSNR 43.91 dB, SSIM 0.969, 670 ms

State-of-the-art: FBPConvNet



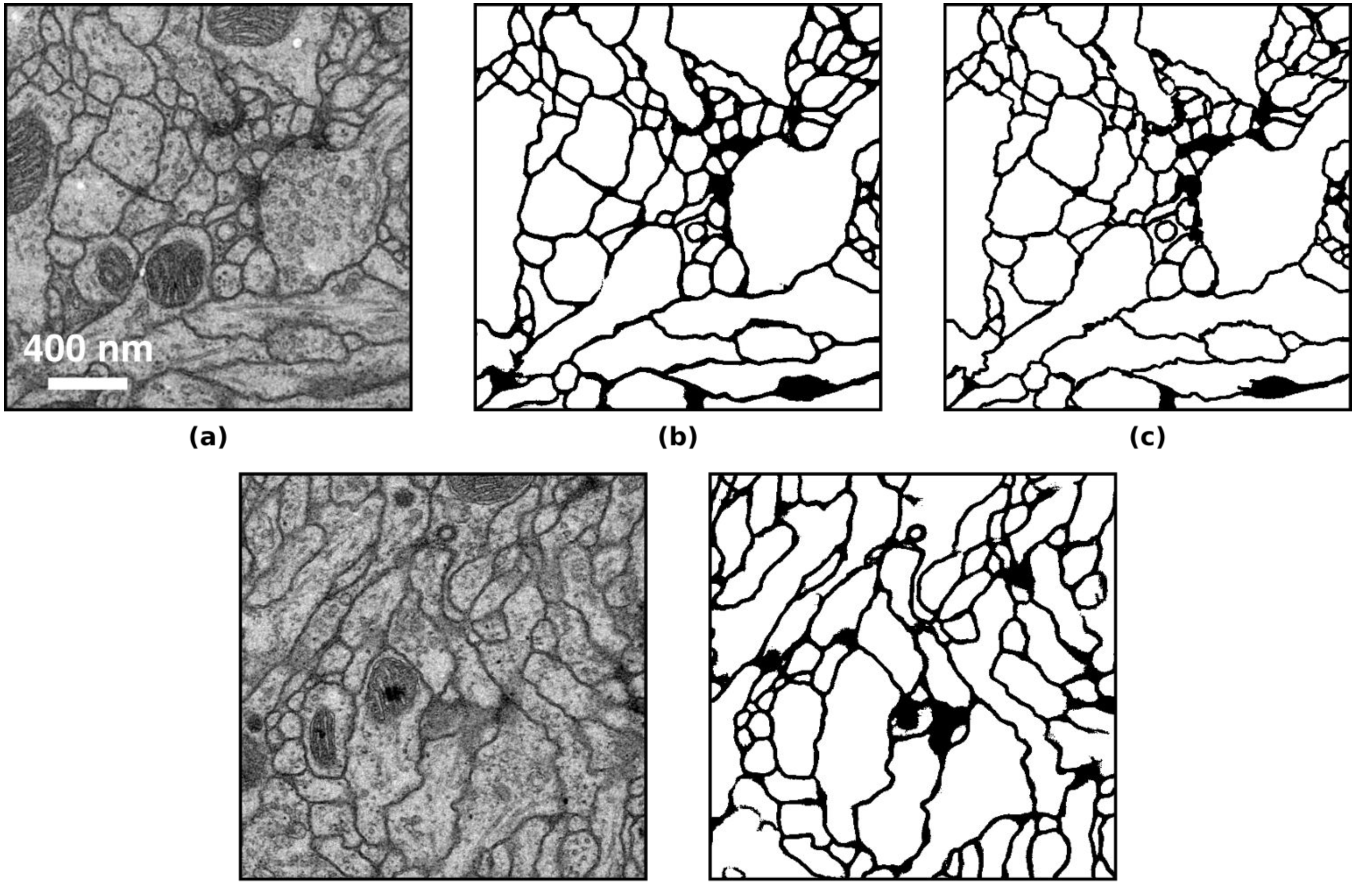
(Jin et al 2017)

State-of-the-art: FBPConvNet



(Jin et al 2017)

Results - ssTEM segmentation



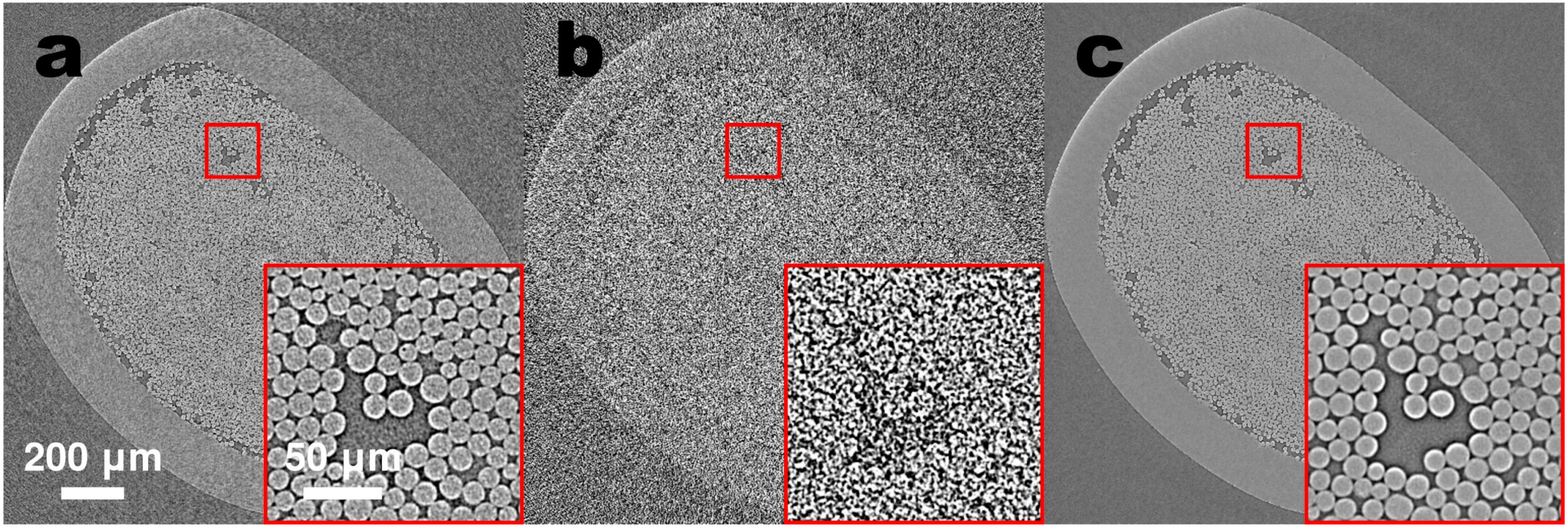
(a) (b) (c)

(d) (e)

- 512x512 pixels, only 30 images for training

Arganda-Carreras, Ignacio, et al. "Crowdsourcing the creation of image segmentation algorithms for connectomics." Frontiers in neuroanatomy 9 (2015): 142.

Results - low-dose tomography



- 2560x2560 tomography images of fiber composite
  - Left: 1024 projections, middle/right: 128 projections
- 500 slices for training, 100 for validation, from top
  - Result for slice from bottom

Advantages

	FBPConvNet	MS-D Net
Parameters	~ 31 million	~ 46 thousand
Intermediate images	822	100
Max. theoretical size	~ 1024x1024	~ 4096x4096

Advantages

	FBPConvNet	MS-D Net
Parameters	~ 31 million	~ 46 thousand
Intermediate images		
Max. theoretical size