center for systems biology dresden

Max Planck Institute of Molecular Cell Biology

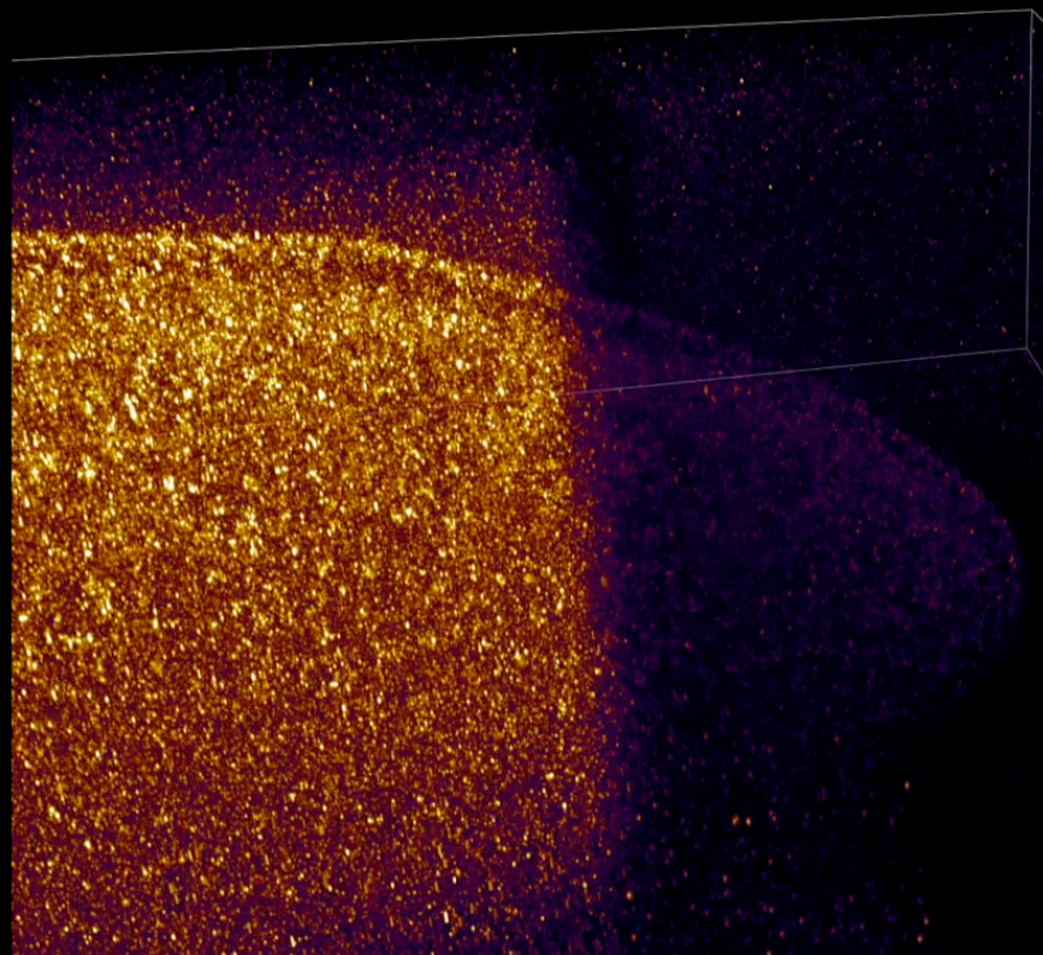
Google Research

Interpretable Unsupervised Diversity Denoising and Artefact Removal

Mangal Prakash, Mauricio Delbracio, Peyman Milanfar, Florian Jug

ICLR

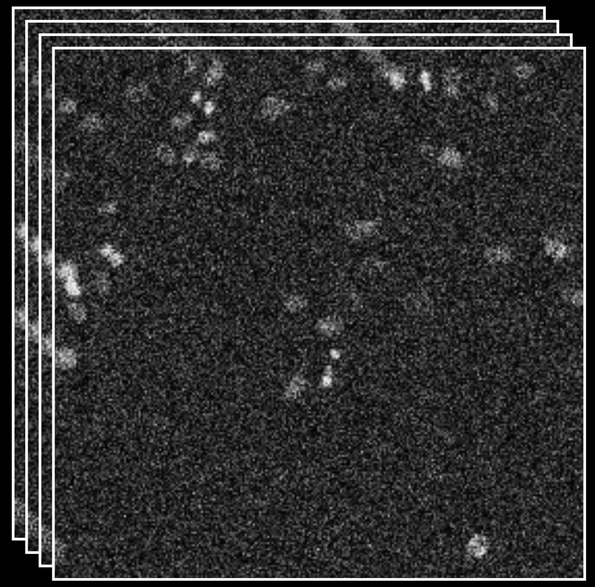
The Denoising Task



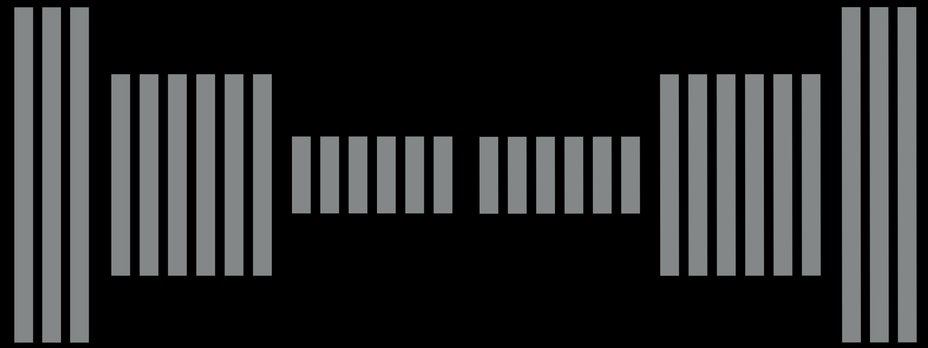
Weigert *et al.*, 2018

Traditional supervised denoisers: Training

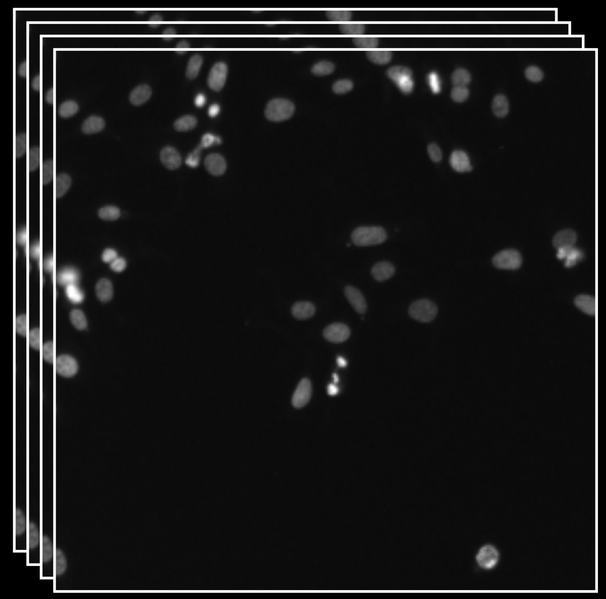
Noisy Input



Convolutional Neural Network



Target

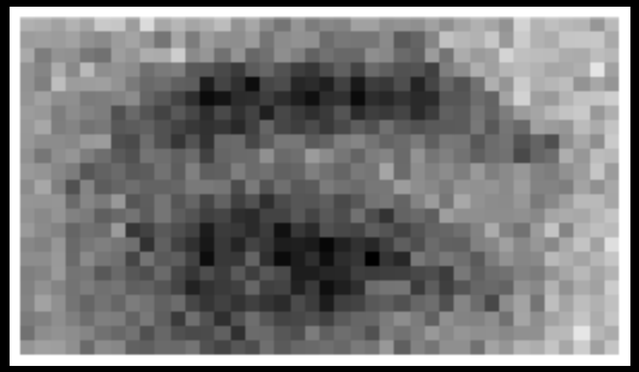


Pairs of High Quality/Low Quality are needed!

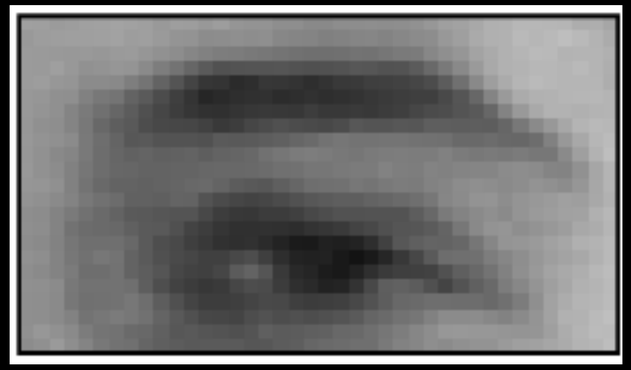
Zhang et al. 2017, Weigert *et al.*, 2018, Chen et al. 2018, Delbracio et al., 2020

Noisy images are ambiguous

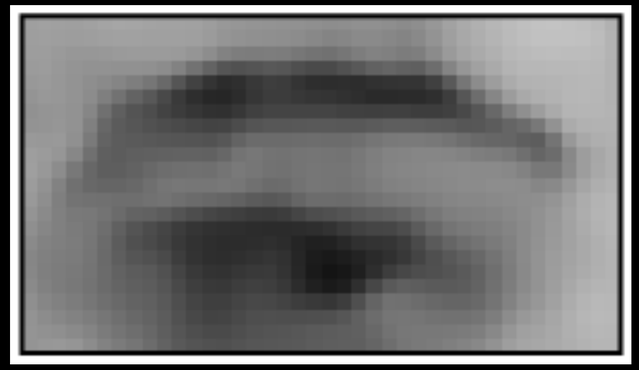
INPUT



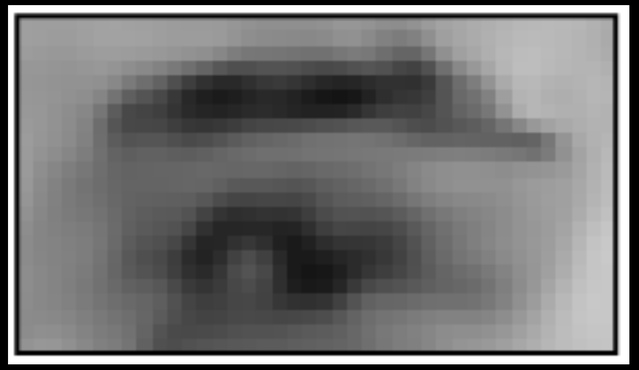
Potential solution 1



Potential solution 2

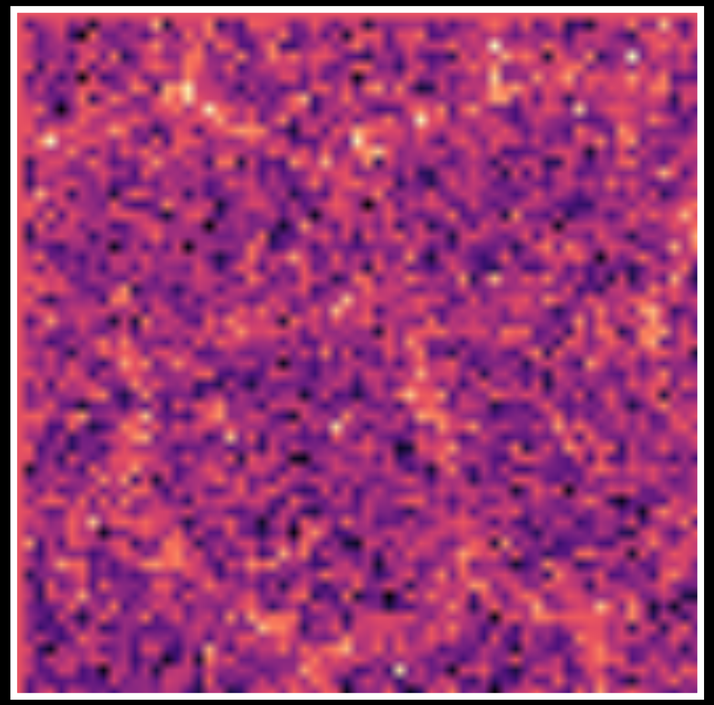


Potential solution 3

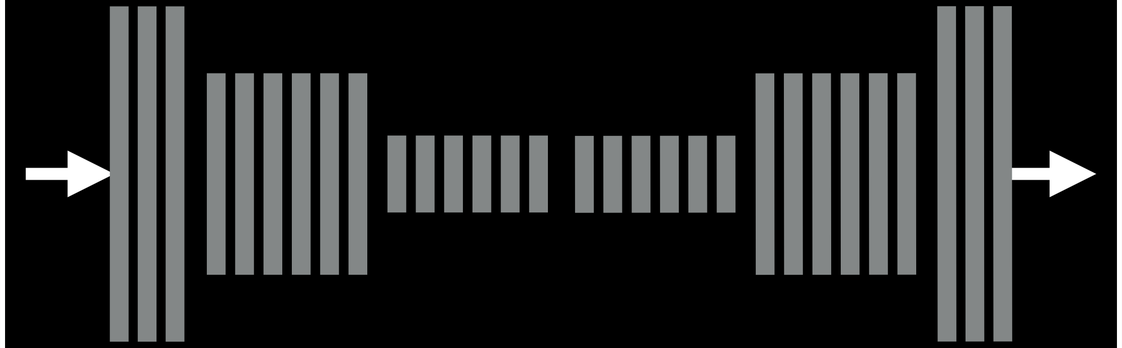


DL based denoising

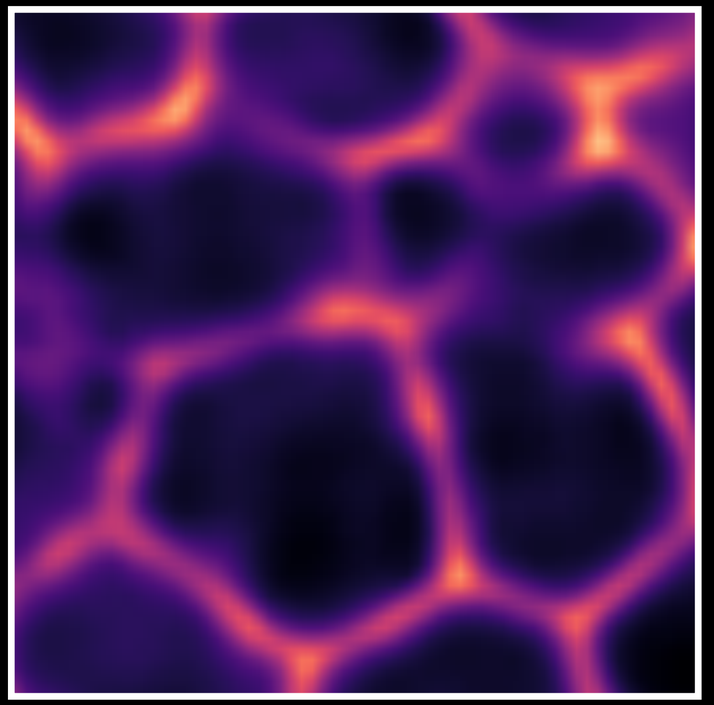
Noisy input (x)



Existing supervised/
unsupervised methods



Denoised output ($\hat{s} \approx s$)



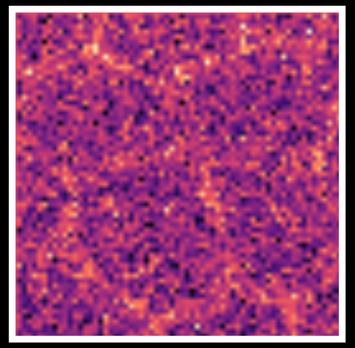
Most methods predict single denoised solution!

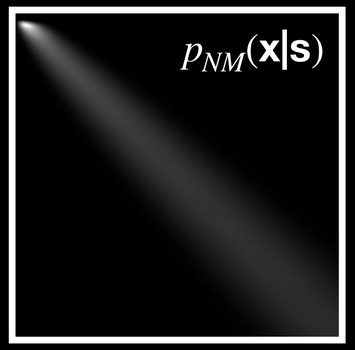
Unsupervised methods have weaker performance than supervised ones.

Unsupervised methods limited to pixel-wise noise (don’t handle spatially correlated/structured noise).

Prior work: DivNoising

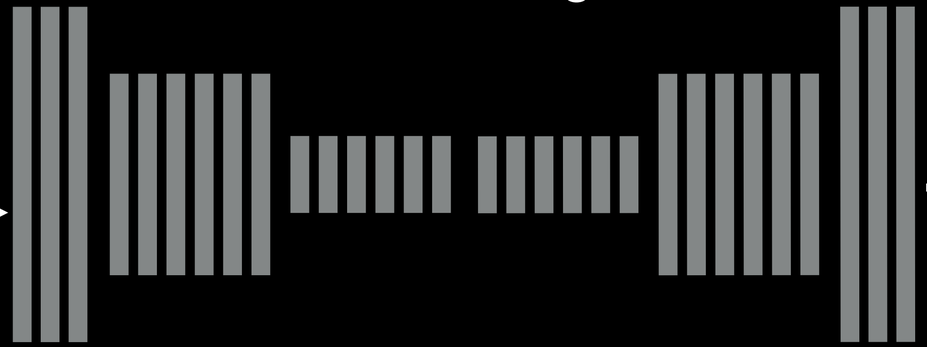
Noisy input (x)





Can be bootstrapped/
learned from noisy data

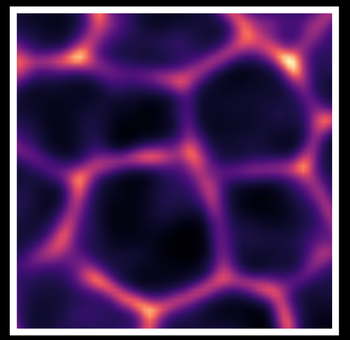
DivNoising

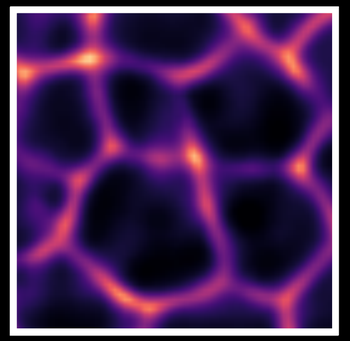


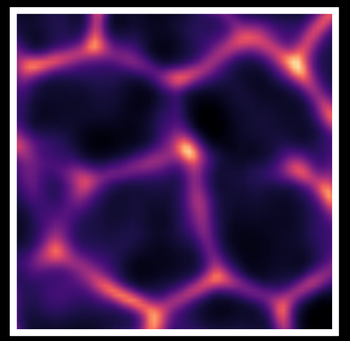
Needs a noise model for training

Diverse solutions

$s^k \sim p(s \mid x)$

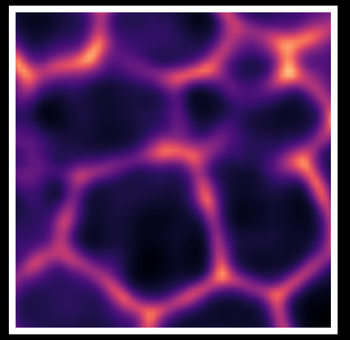




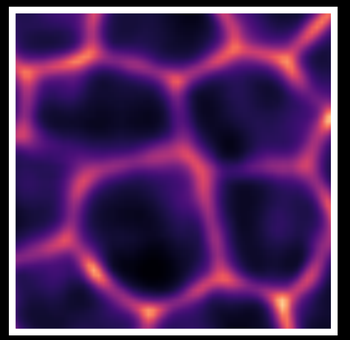


INFERENCE

MMSE



MAP

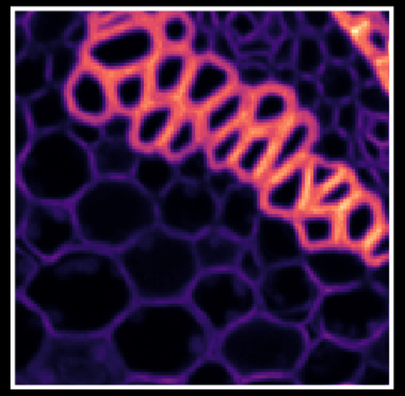


Prakash et al., ICLR'21

Limitations of DivNoising

Input (Pixel Noise)

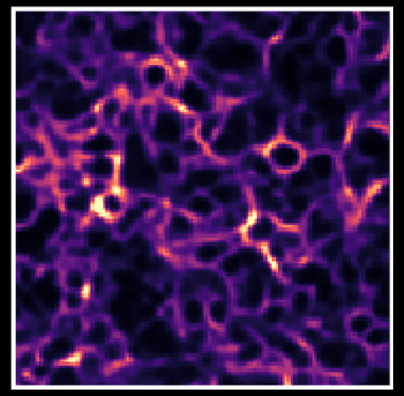
Real Image



DN Denoised

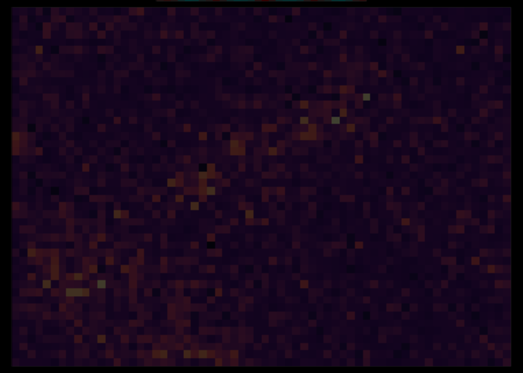
DivNoising does not work well for complex data

DN Generated



DivNoising generative model is not so accurate

Input (Structured Noise)



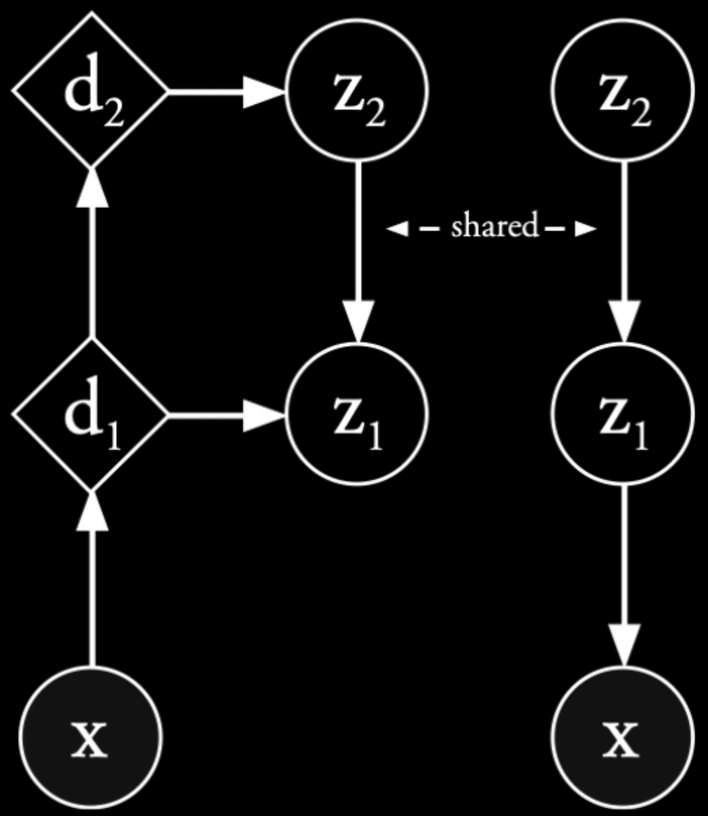
DN Denoised

DivNoising does not deal with structured noises

Hypotheses:

- Not expressive enough architecture.
- Does not capture long range interactions

Hierarchical VAEs are expressive for image generation but expensive to train



Model	Trainable Params
BIVA	103 million
NVAE	268 million
VDVAE	115 million

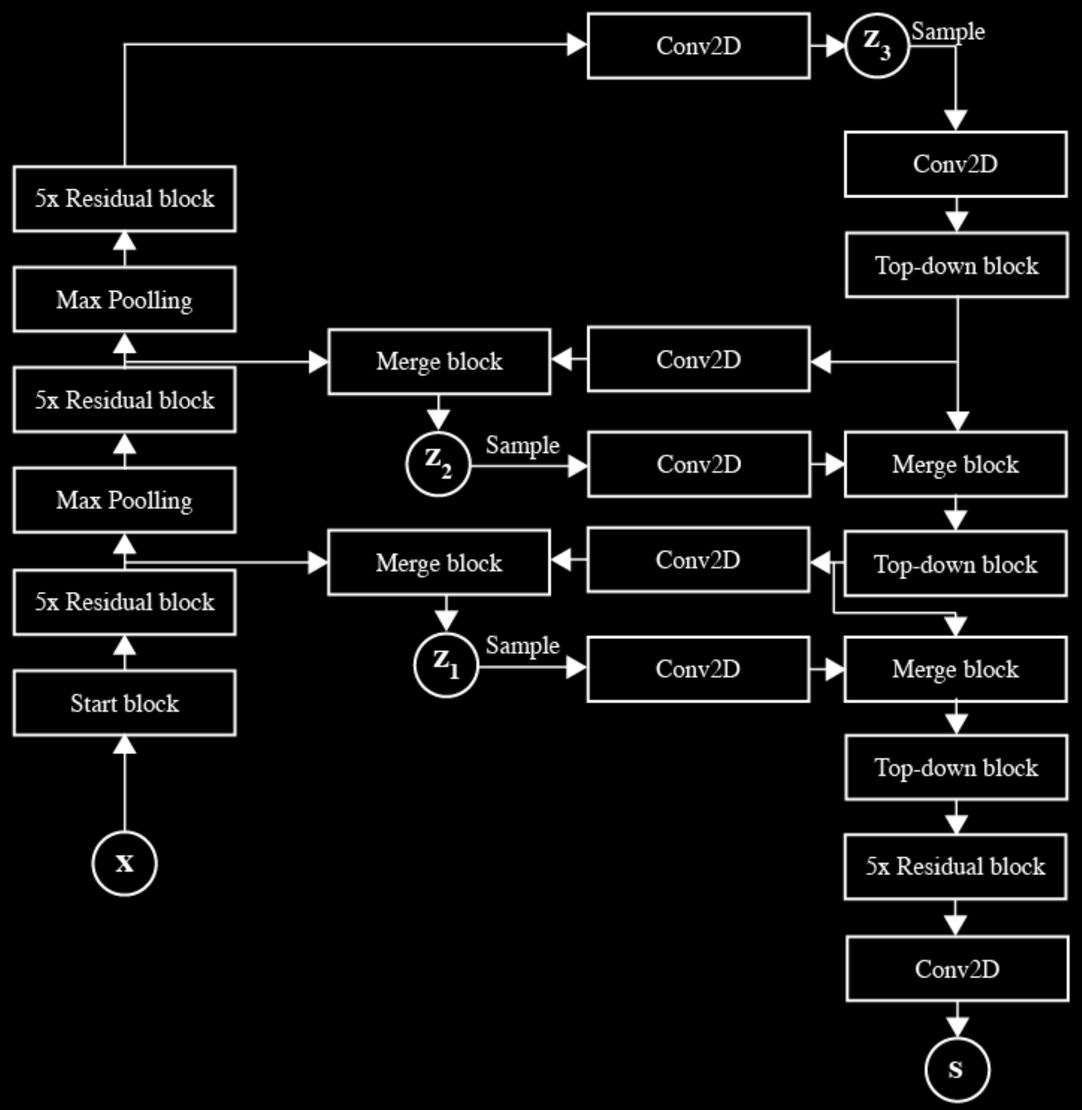
Training takes days to weeks on multiple high performance GPUs

- HVAEs have never been applied to image restoration
- Computationally very expensive

Do more expressive HVAE architectures improve unsupervised image restoration while being computationally cheap?

Sonderby et al. 2016, Maaløe et al., 2019, Vahdat et al., 2020, Child et al., 2020

Hierarchical DivNoising is expressive and computationally efficient



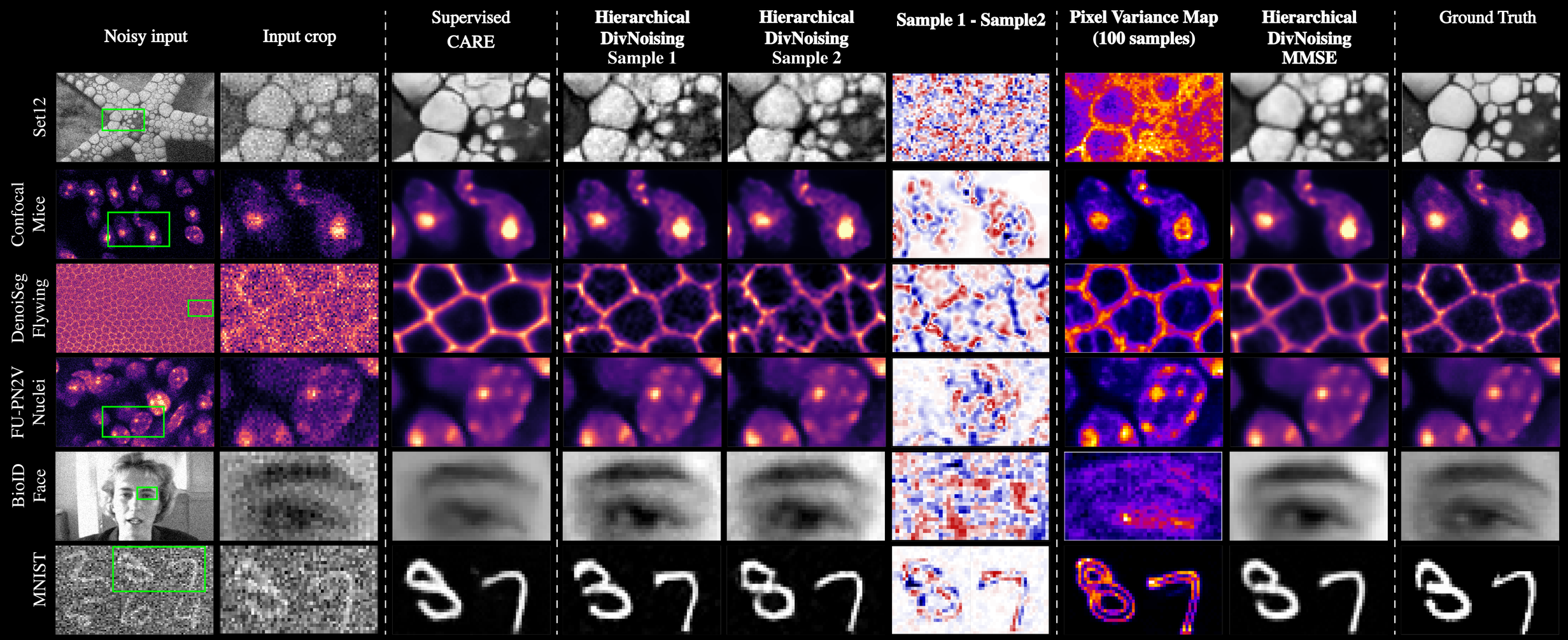
Ablation of other units
(6 latent layers, 5 res. blocks/layer)

Gating blocks	Batch norm.	Gen. skips	Free bits	PSNR
×	✓	✓	✓	28.58
✓	×	✓	✓	28.37
✓	✓	×	✓	28.61
✓	✓	✓	×	28.66
✓	✓	✓	✓	**28.82**

- Training time ~1 day on single 6 GB Tesla P100 GPU
- Number of trainable parameters 7.301 million

- Careful architecture design keeps the network small but expressive
- Suitable for practical applications

HDN Generates Diverse Plausible Denoised Results

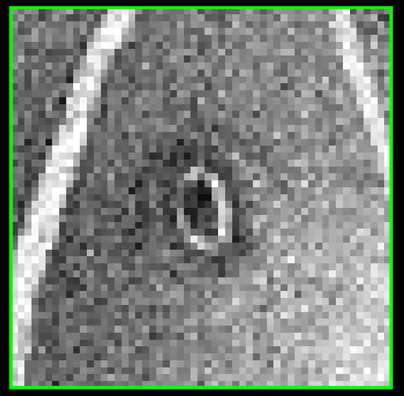


HDN is SOTA for Unsupervised Pixel Noise Removal

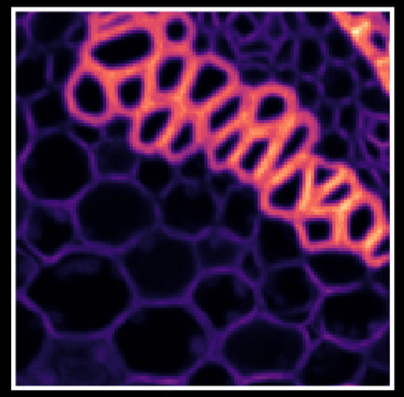
	Non DL	Single-image DL		Multi-image DL			Diversity DL			Supervised	
Dataset	**BM3D**	**DIP**	**S2S**	**N2V**	**N2Same**	**PN2V**	**HVAE**	**DN**	**HDN**	**N2N**	**CARE**
Convallaria	35.45	-	-	35.73	36.46	36.47	34.11	36.90	**37.39**	36.85	36.71
Confocal Mice	37.95	-	-	37.56	37.96	38.17	31.62	37.82	**38.28**	38.19	38.39
2 Photon Mice	33.81	-	-	33.42	33.58	33.67	25.96	33.61	**34.00**	34.33	34.35
Mouse Actin	33.64	-	-	33.39	32.55	33.86	27.24	33.99	**34.12**	34.60	34.20
Mouse Nuclei	36.20	-	-	35.84	36.20	36.35	32.62	36.26	**36.87**	37.33	36.58
Flywing	23.45	24.67	-	24.79	22.81	24.85	25.33	25.02	**25.59**	25.67	25.79
BSD68	28.56	27.96	28.61	27.70	27.95	28.46	27.20	27.42	**28.82**	28.86	29.07
Set12	29.94	28.60	29.51	28.92	29.35	29.61	27.81	28.24	**29.95**	30.04	30.36
BioID Faces	33.91	-	-	32.34	34.05	33.76	31.65	33.12	**34.59**	35.04	35.06
CelebA HQ	33.28	-	-	30.80	31.82	33.01	31.45	31.41	**33.54**	33.39	33.57
Kanji	20.45	-	-	19.95	20.28	19.40	20.44	19.47	**20.72**	20.56	20.64
MNIST	15.82	-	-	19.04	18.79	13.87	20.52	19.06	**20.87**	20.29	20.43

Limitations of DivNoising

Input (Pixel Noise)



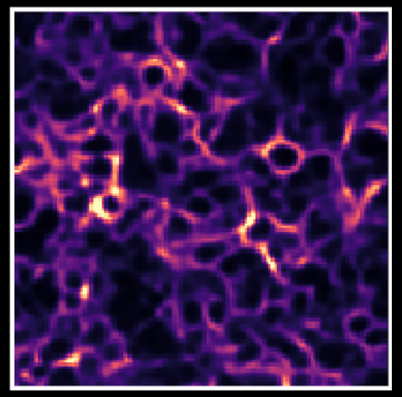
Real Image



DN Denoised

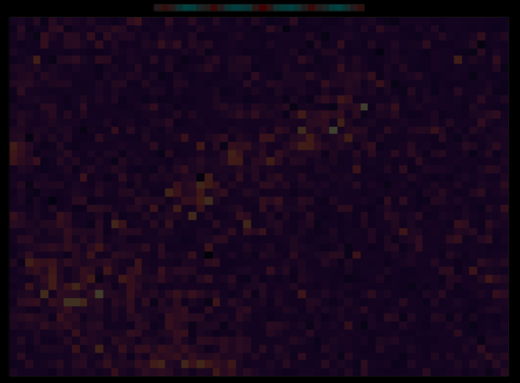
DivNoising does not work well for complex data

DN Generated



DivNoising generative model is not so accurate

Input (Structured Noise)



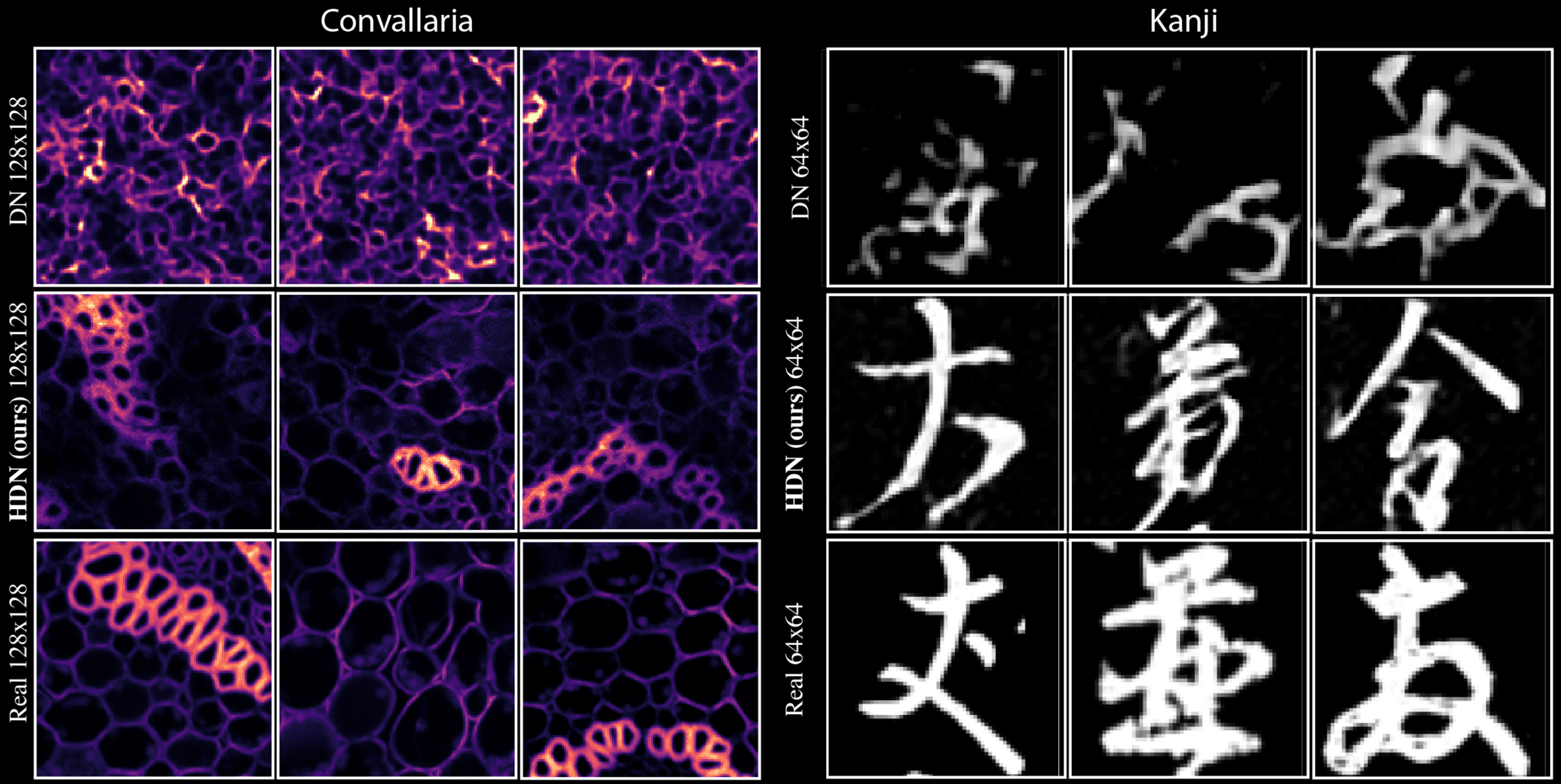
DN Denoised

DivNoising does not deal with structured noises

Hypotheses:

- Architecture is not expressive enough.
- Does not capture long range interactions

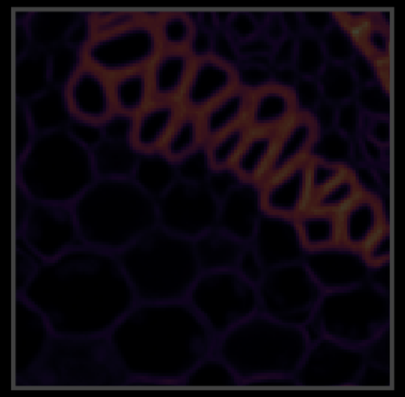
Hierarchical DivNoising generative model is much more accurate



Limitations of DivNoising

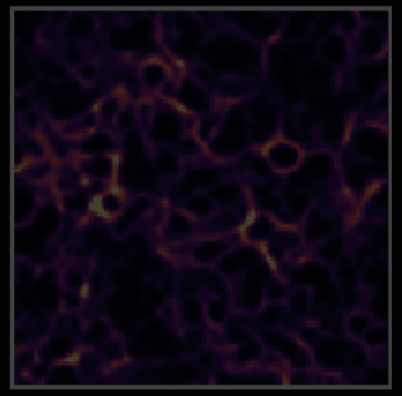
Input (Pixel Noise)

Real Image



DN Denoised

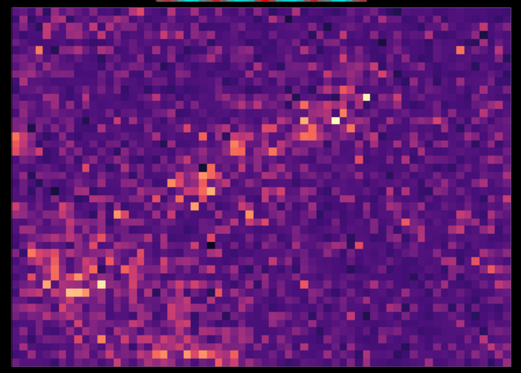
DN Generated



DivNoising does not work well for complex data

DivNoising generative model is not so accurate

Input (Structured Noise)

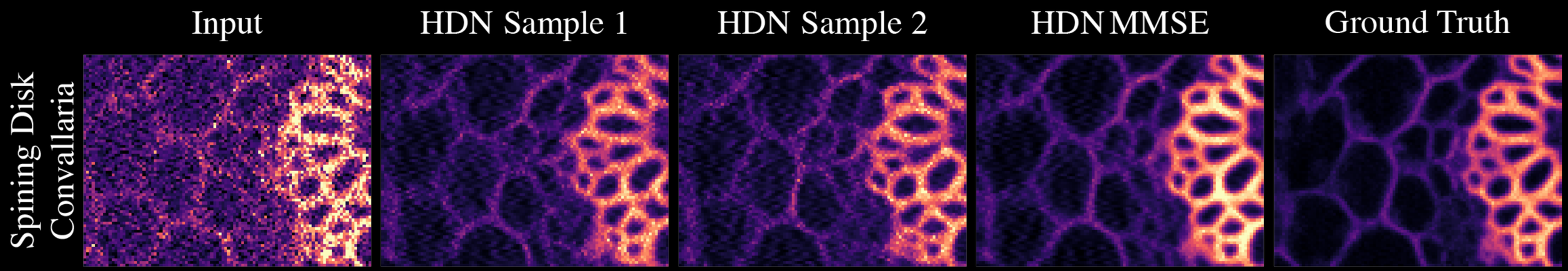


DN Denoised

DivNoising does not deal with structured noises

Easy to obtain a pixel noise model though

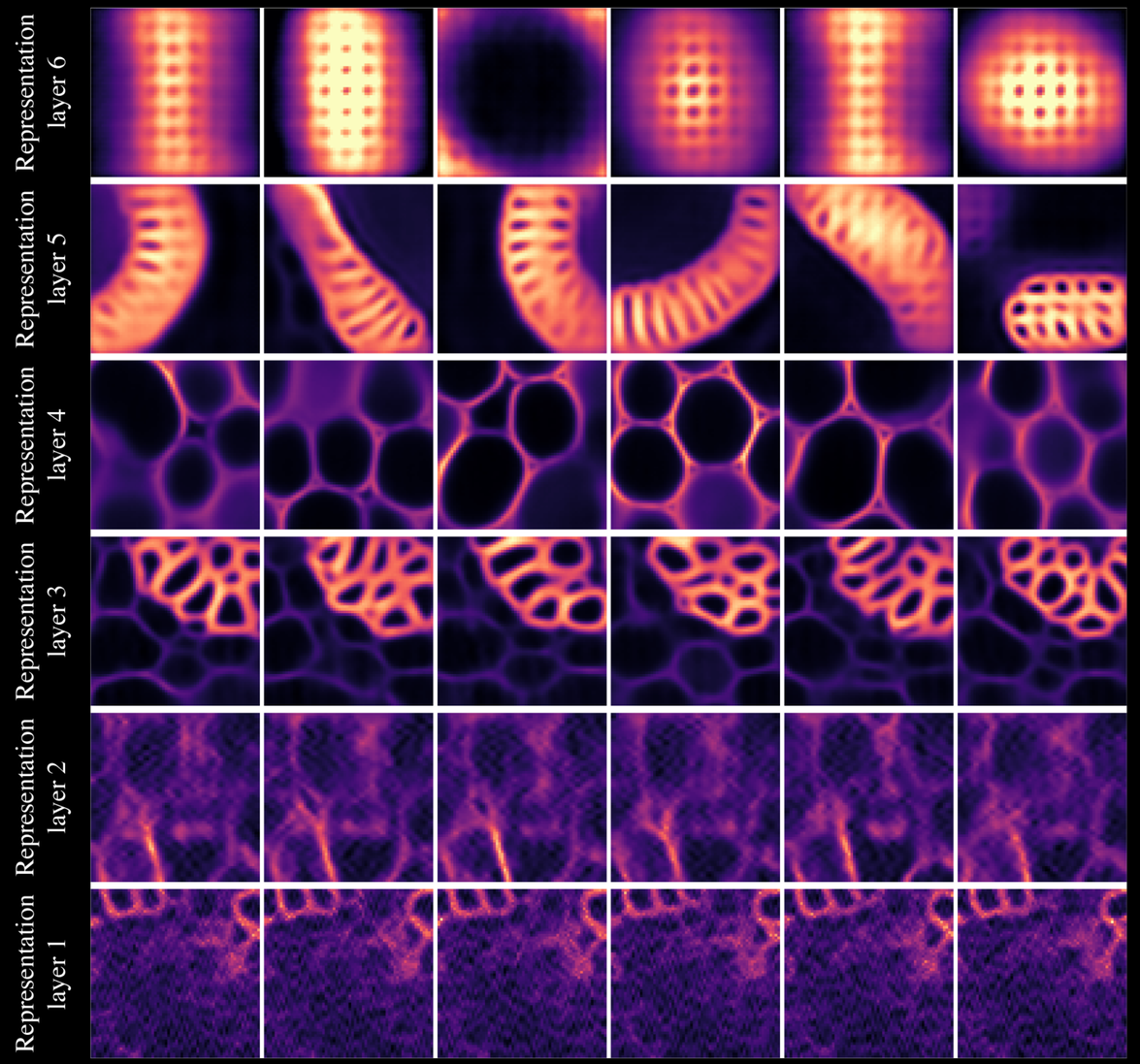
HDN is expressive enough to capture structured noise



HDN trained only with pixel noise model
(has no information about structured noise)

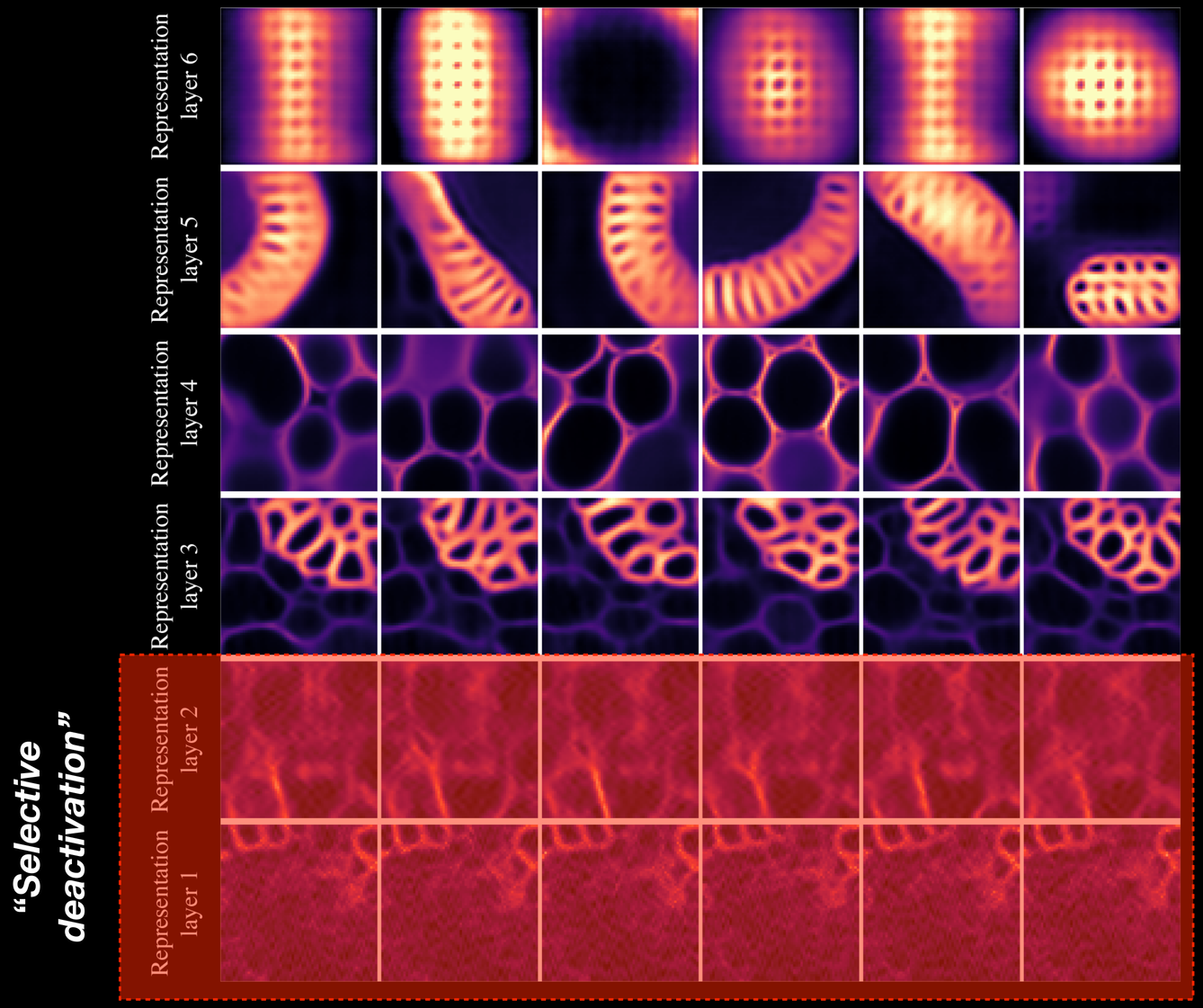
Not surprising structured noises are not removed

HDN is interpretable

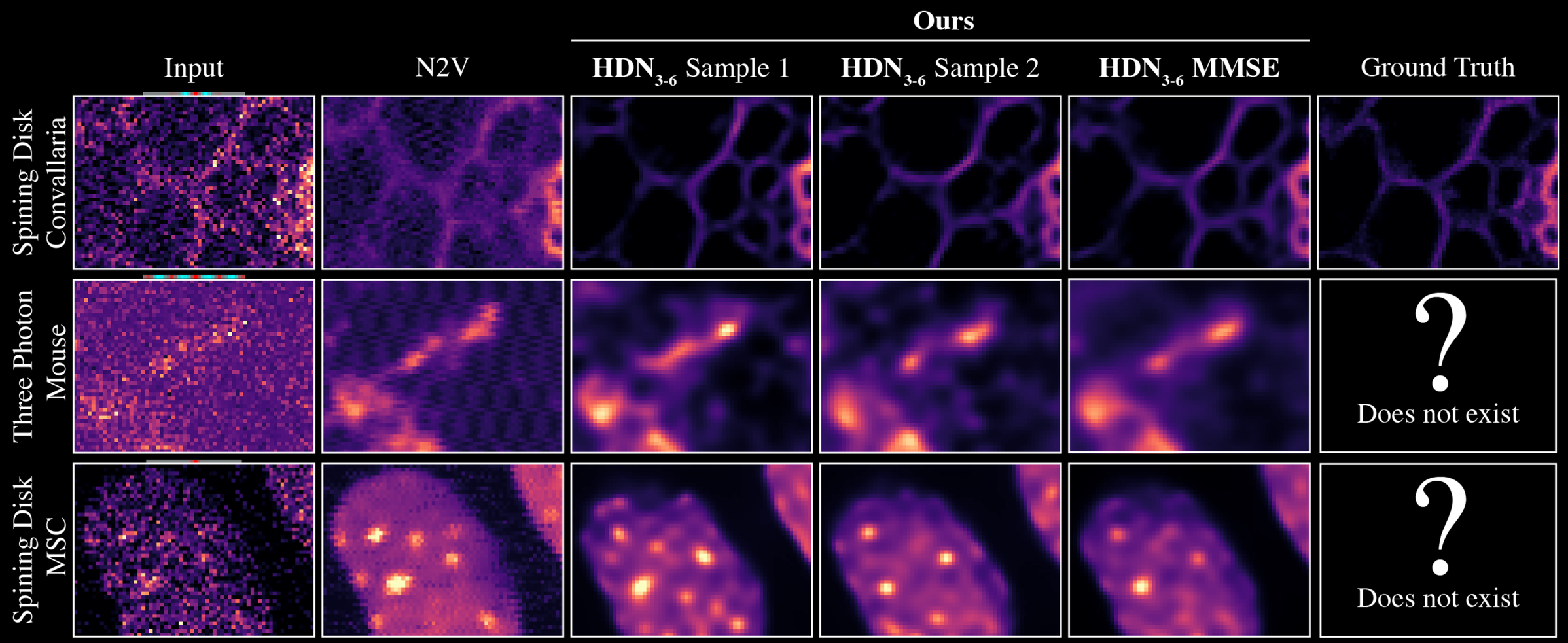


Structured line artefacts are only captured in layers 1 and 2

Modified HDN Architecture Removes Structured Noise



HDN is SOTA for Unsupervised Structured Noise Removal



	Unsupervised						Supervised
	N2V	PN2V	Struct N2V	DN	HDN (Ours)	HDN3-6 (Ours)	CARE
Struct.Convallria	29.33	29.43	30.02	31.09	29.96	**31.36**	31.56

Take home messages

HDN is:

- unsupervised
- generates plausible diverse interpretations
- expressive while being computationally cheap
- interpretable

center for systems biology dresden

CBG
Max Planck Institute of Molecular Cell Biology and Genetics

Google Research

HUMAN TECHNOPOLE

Thank You!

GitHub

Paper

@Mangal_Prakash_