DZNE [1]
DeepMI Lab
Martinos Center [2]
HARVARD MEDICAL SCHOOL [3]

Do "denoisers" actually denoise?

Aligning denoiser evaluation with the real world

Sebastian Rassmann[1], David Kügler[1], Sascha Brunheim[1], Philipp Ehses[1], Martin Reuter[1,2,3]

ECCV 2026 | MALMÖ | SEPT 8–12
MedFM-Bench workshop / ECCV main track paper

Noise in MRI

- MRI is inherently noisy, **"clean" MRI does not exist [1]**

⇒ How to evaluate **removal of real, physical MRI noise?**

⇒ What is the ground truth?

Denoiser evaluation flaws

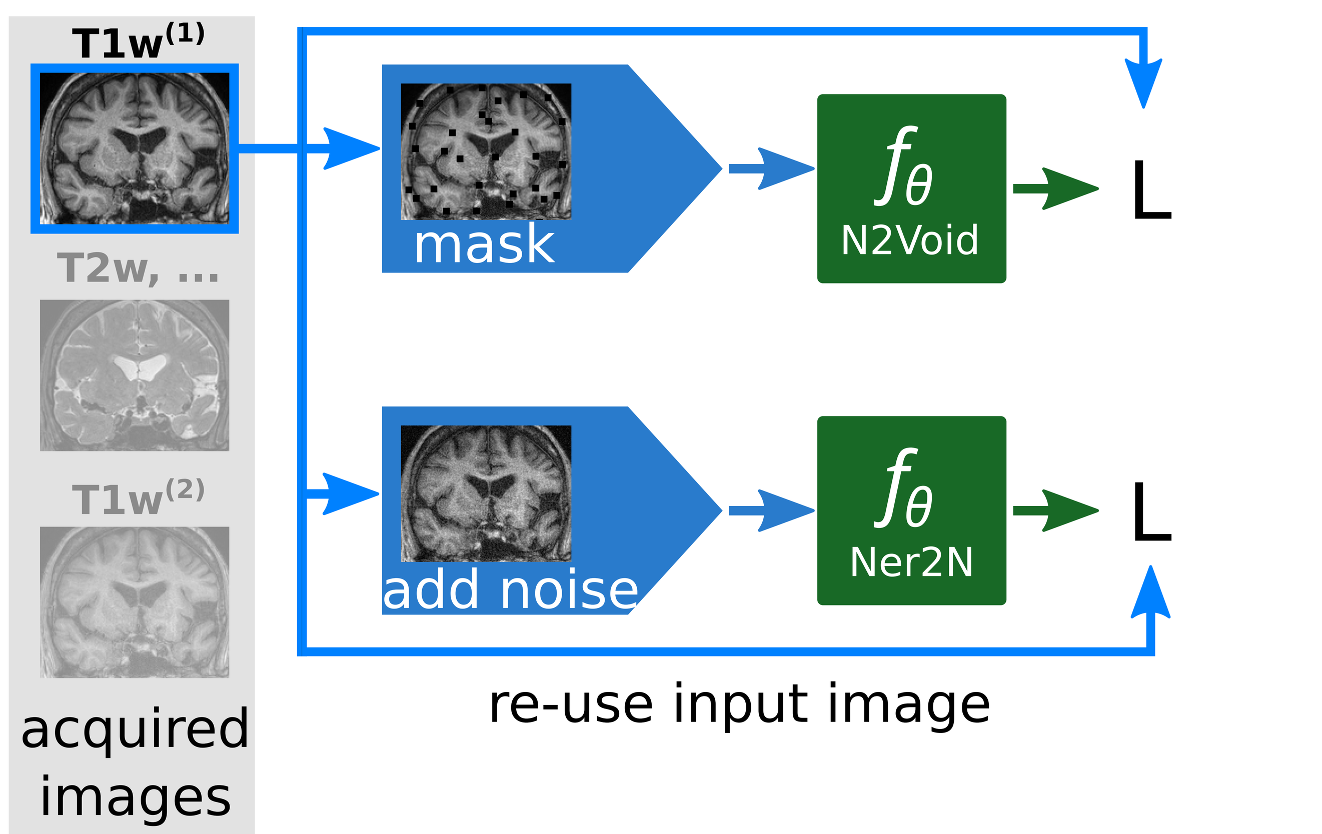
- Common practice in SOTA [e.g. 2-4]:

  Add synthetic noise → evaluate against original image

  - Rewards sustaining acquisition noise
  - Evaluates only removal of added synthetic noise

⇒ Misalignment with evaluating real-world performance

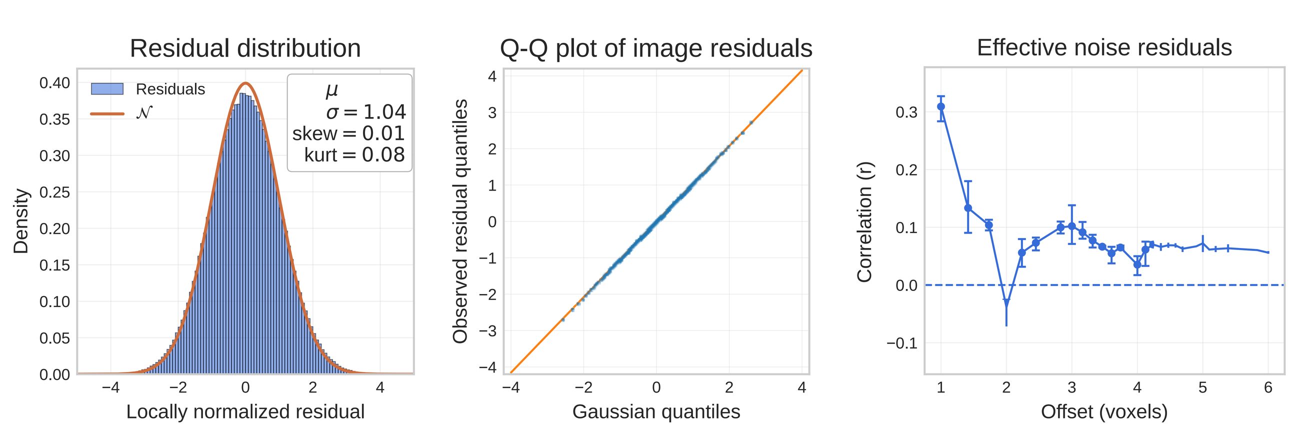
Existing MRI denoisers



- Train with surrogate objective (blindspot [5,6] / noising [7])

  ⇒ Tradeoff: **noise concealment vs. domain shift**

Real-world MRI noise



✔ MRI noise is (approximately) Gaussian

✗ MRI noise is not white!

- autocorrelation due to CS, motion & interpolation

⇒ evaluation with AWGN oversimplifies

Benchmarking denoisers in the real world

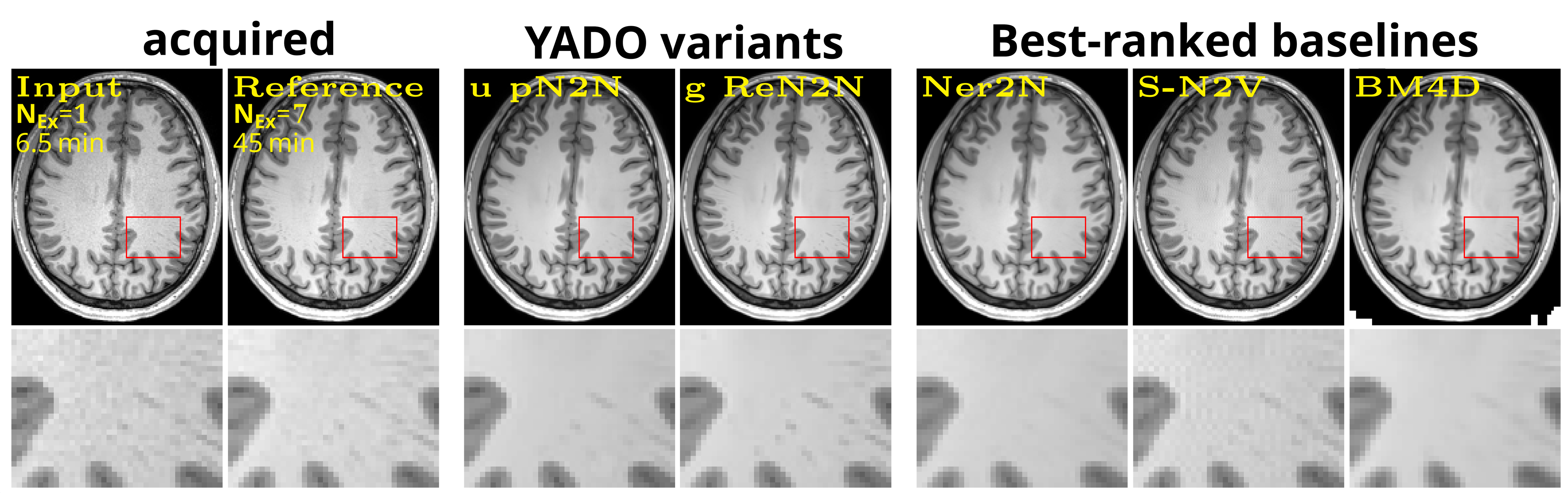
		OASIS-3 trained (ca. 1.0 mm)							HCP-A trained (0.8 mm)				all
		SSIM (%)						CNR	SSIM (%)			CNR	rank
Denoiser	Venue	OAS3	CHDI	Kirby	HCP-E	-YA	RS	PVS	HCP-E	-YA	RS	PVS	all
u pN2N	proposed	89.39	94.25	95.69	89.65	94.06	93.83	**2.63**	87.05	85.85	90.44	2.15	4.55
u ReN2N	proposed	89.41	94.33	**95.99**	90.60	94.50	94.37	**2.74**	88.25	86.66	91.23	**2.32**	2.09
g pN2N	proposed	**90.03**	94.11	(94.63)	89.96	94.11	**94.31**	**2.85**	86.87	86.05	89.93	2.25	4.17
g ReN2N	proposed	**89.77**	**94.59**	(95.63)	**90.91**	**94.59**	**94.59**	**2.85**	**88.74**	**86.98**	**91.69**	**2.29**	**1.46**
Ner2N [7]	CVPR '20	88.59	93.53	95.58	89.47	93.91	93.42	2.28	84.63	82.65	88.45	2.10	9.26
N2Score [8]	NeurIPS '21	89.07	93.77	95.21	87.89	93.61	92.75	1.06	79.97	65.71	90.24	1.71	12.24
N2Void [5,6]	CVPR '19	88.60	93.78	95.72	88.19	90.86	92.63	2.24	86.99	85.14	90.11	2.20	8.58
S-N2Void [9]	ISBI '20	88.69	93.67	95.59	88.43	91.38	92.62	2.17	87.42	85.89	90.56	2.22	7.81
Neigh2N. [10]	CVPR '21	87.94	93.03	94.59	89.08	93.61	92.68	2.31	87.36	85.40	89.63	1.83	9.75
RED-WGAN [2]	MedIA '19	86.35	91.38	93.42	87.21	89.48	90.49	2.18	84.70	83.53	88.34	1.88	15.35
CNN3D [2]	MedIA '19	87.79	93.00	95.11	88.86	92.62	92.62	2.25	86.28	84.51	89.31	1.95	11.62
DDM^2 [11]	ICLR '23	51.05	43.04	39.29	35.15	39.49	48.27	0.16	23.64	30.25	39.82	0.02	22.00
DDM^2 Reg. [11,1]	ICLR '23	87.28	91.14	92.73	84.10	89.80	91.28	1.76	72.24	75.28	84.41	1.01	17.70
BME-X [3]	Nat. BE '25	83.10	87.81	86.64	82.87	88.38	88.05	1.55	82.71	82.17	88.14	1.56	18.68
N2Contr [4]	IPMI '23	87.32	93.12	95.34	87.98	90.46	92.16	2.42	85.51	83.39	88.97	2.12	11.82
YODA [1]	TMI '26	88.29	87.95	(70.80)	83.48	86.87	89.18	2.17	71.19	80.11	84.17	1.69	17.30
DIP [12]	CVPR '18	87.87	93.27	95.31	89.04	92.23	92.88	2.33	86.77	84.86	89.58	1.97	9.84
Pix2Pix [13]	TPAMI '25	85.76	90.81	92.16	87.20	91.39	90.23	1.41	86.20	83.87	87.94	1.21	16.46
ANTs NLM [14]	JMRI '10	87.89	93.11	94.97	88.95	93.21	92.62	2.36	87.22	85.42	89.73	1.91	9.49
BM4D [15]	TIP '12	88.28	93.40	95.33	89.37	93.71	92.93	2.13	87.81	86.09	90.27	1.83	7.96
BME-X as-is [3]	Nat. BE '25	52.72	55.39	55.48	46.47	68.41	53.13	1.53	35.33	68.85	64.45	1.53	20.52
Identity	*proposed*	86.91	92.66	95.03	87.09	89.50	91.60	2.33	84.61	82.68	88.47	2.10	14.35

bold: best (p<0.05) — gray: no improvement over Identity (NOP, p≥0.05)

- **17 methods across 11 test conditions** (+ T2w/FLAIR)

  ⇒ **7 / 17 rank worse than the Identity** (NOP)

  ⇒ traditional BM4D denoiser [15] (2012!) ranks second



Fix MRI denoiser evaluation

- Model acquiring MRI X_t as:

$$X_t \;=\; \mathbb{E}\left[X\right] + \mathbf{n}_t \;=:\; X' + \mathbf{n}_t\,,$$

X': clean image
$\mathbf{n} \sim \mathcal{D}_{\text{Noise}}$
$\mathbb{E}\left[\mathbf{n}\right] := 0$

- Denoising aims to find $f(\bullet)$ s.t.:

$$\text{err}_f \;:=\; f(X_{\text{in}}) - X' \;\longrightarrow\; \min$$

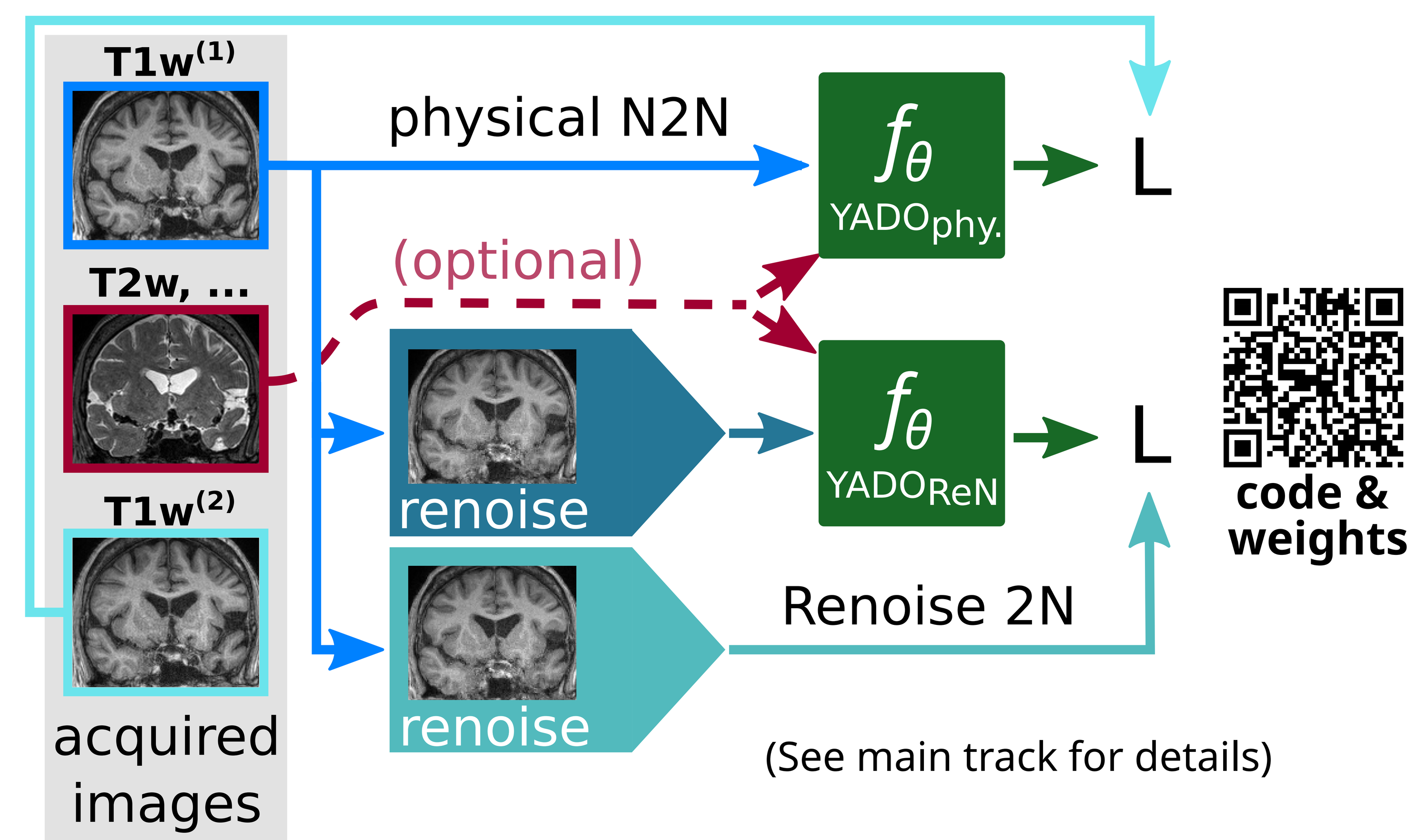
- Take (co-reg.) repeats $X_{\text{in}}, X_{\text{ref}}$ (s.t. $\mathbf{n}_{\text{in}} \perp \mathbf{n}_{\text{ref}}$):

$$\text{MSE}\;\left(f(X_{\text{in}}), X_{\text{ref}}\right) \;=\; \mathbb{E}\left[\left\|f(X_{\text{in}}) - (X' + \mathbf{n}_{\text{ref}})\right\|_2^2\right]$$

$$= \underbrace{\mathbb{E}\left[\left\|\text{err}_f\right\|_2^2\right]}_{\text{true error}} + \underbrace{\mathbb{E}\left[\left\|\mathbf{n}_{\text{ref}}\right\|_2^2\right]}_{\substack{\text{aleatoric error}\\ \text{const. } \forall f}} - \underbrace{2\,\mathbb{E}\left[\text{err}_f^{\mathsf{T}}\,\mathbf{n}_{\text{ref}}\right]}_{\substack{\text{correlation}\\ =0 \text{ iff } \mathbf{n}_{\text{in}} \perp \mathbf{n}_{\text{ref}}}}$$

⇒ Repeated acquisition X_{ref} is a proxy for the clean image X'

Proposed: YADO



- **Learn from physical noise ⇒ near-ideal training [7]**

Take-aways

- Denoisers are meant to remove **real acquisition noise**

  ⇒ Evaluation must be on independent noise instances

- **Aligned & unbiased benchmarks** are key to develop **denoisers that actually remove physical, real-world noise**

- Potential self-supervised Foundation Model tasks:
  - Denoising (& translation [1]): large datasets [16]
  - ReN2N as data augmentation ⇒ resilience to noise!

References: [1] Rassmann et al., "Regression is all you need for Medical Image Translation", IEEE TMI '26; [2] Ran et al., "Denoising of 3D MRI Using a Residual Encoder-Decoder WGAN", MedIA '19; [3] Sun et al., "A Foundation Model for Enhancing MRI and Downstream Segmentation, Registration and Diagnostic Tasks", Nat Biomed Eng '25; [4] Wagner et al., "Noise2Contrast: Multi-Contrast Fusion Enables Self-Supervised Tomographic Image Denoising", IPMI '23; [5] Krull et al., "Noise2Void: Learning Denoising from Single Noisy Images", CVPR '19; [6] Batson et al., "Noise2Self: Blind Denoising by Self-Supervision", ICML '19; [7] Moran et al., "Noisier2Noise: Learning to Denoise from Unpaired Noisy Data", CVPR '20; [8] Kim et al., "Noise2Score: Tweedie's Approach to Self-Supervised Image Denoising", NeurIPS '21; [9] Broaddus et al., "Removing Structured Noise with Self-Supervised Blind-Spot Networks", ISBI '20; [10] Huang et al., "Neighbor2Neighbor: Self-Supervised Denoising from Single Noisy Images", CVPR '21; [11] Xiang et al., "DDM2: Self-Supervised Diffusion MRI Denoising with Generative Diffusion Models", ICLR '23; [12] Ulyanov et al., "Deep Image Prior", CVPR '18; [13] Ma et al., "Pixel2Pixel: A Pixelwise Approach for Zero-Shot Single Image Denoising", IEEE TPAMI '25; [14] Manjón et al., "Adaptive Non-Local Means Denoising of MRI with Spatially Varying Noise Levels", JMRI '10; [15] Maggioni et al., "Nonlocal Transform-Domain Filter for Volumetric Data Denoising and Reconstruction", IEEE TIP '12; [16] Cerri et al., "A large-scale heterogeneous 3D MR brain imaging dataset for self-supervised learning" Sci Data '26