EUROPEAN CONFERENCE ON COMPUTER VISION

MILANO
2024

# TTT-MIM: Test-Time Training with Masked Image Modelling for Denoising Distribution Shifts

Youssef Mansour, Xuyang Zhong, Serdar Caglar, Reinhard Heckel

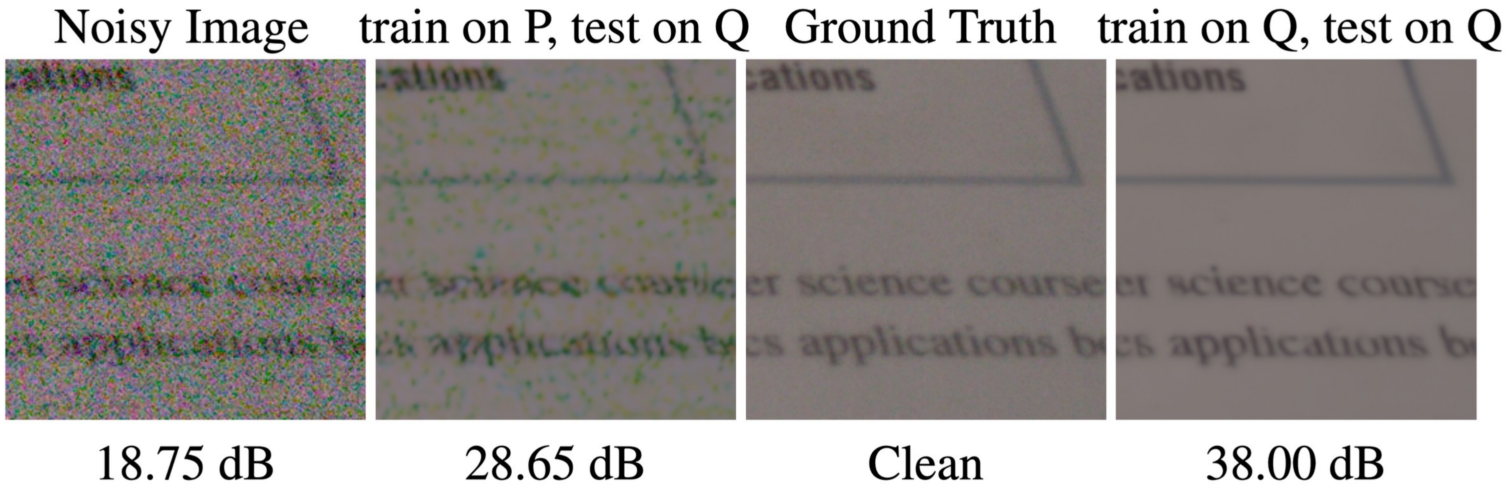
10:30 am
Thursday Oct. 3
Poster number 20

Munich Center for Machine Learning

Technical University of Munich

# Overview

Distribution shift: mismatch between training and test sets

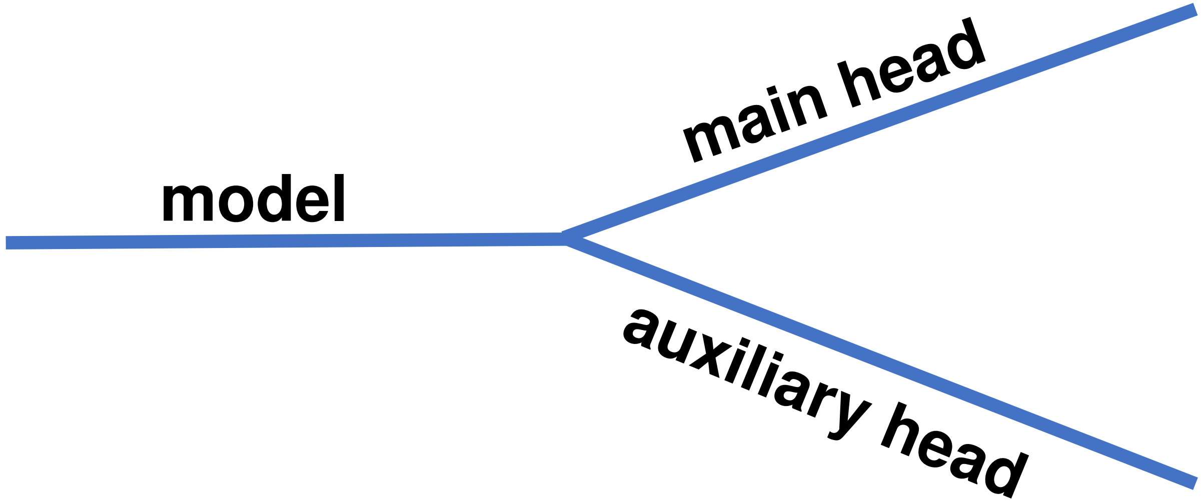


**P: training distribution (cheap) Q: test distribution**

TTT: adapt weights of a trained model to new test instant

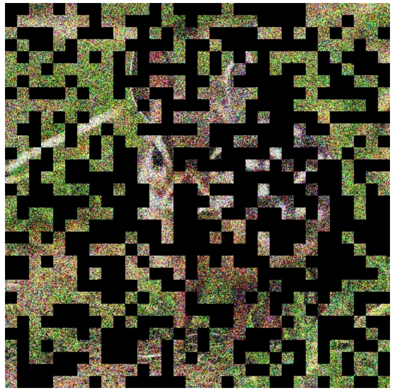
Goal: apply TTT for denoising to a single test image blindly

# TTT framework



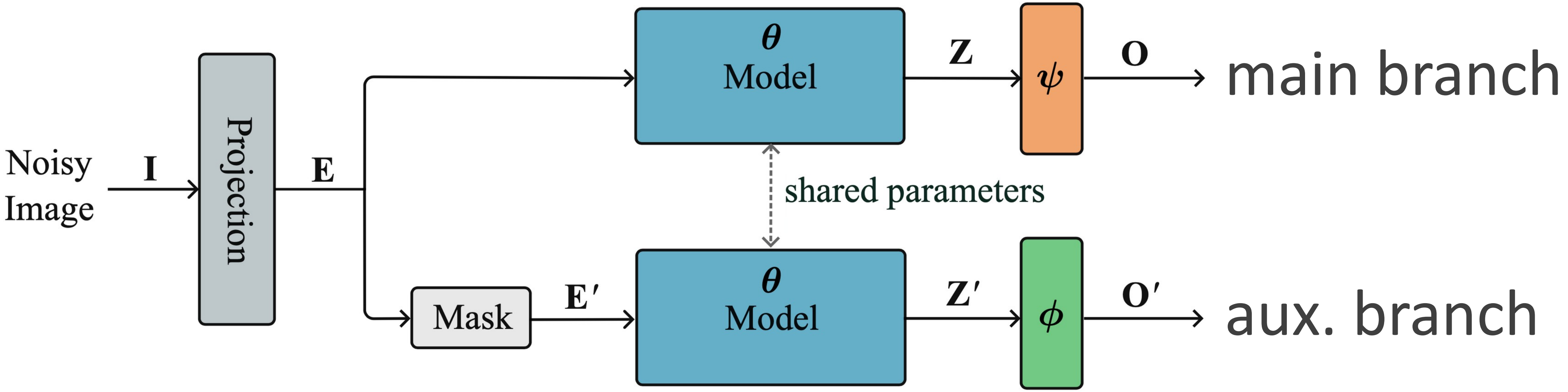
- Main head (denoising): regular supervised loss
- Auxiliary head (meaningful representations): self-supervised loss

**Masked Autoencoders**



**masked noisy img**

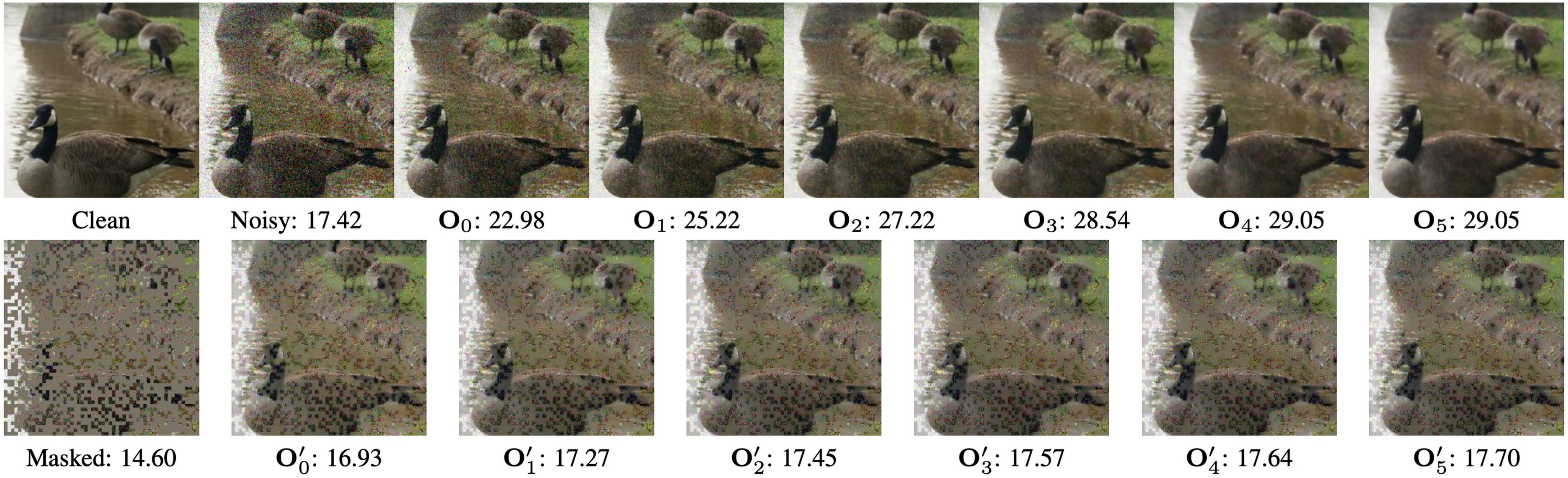
**noisy img**



$$\mathcal{L}_{\text{main}}(\boldsymbol{\theta}, \psi) + \mathcal{L}_{\text{aux.}}(\boldsymbol{\theta}, \phi)$$

Train loss $$||\mathbf{O}_{\boldsymbol{\theta},\psi} - \mathbf{X}||_2^2 + ||\mathbf{O}'_{\boldsymbol{\theta},\phi} - \mathbf{I}||_2^2$$

Test loss $$||\mathbf{O}_{\boldsymbol{\theta},\psi} - \mathbf{O}'||_2^2 + ||\mathbf{O}'_{\boldsymbol{\theta},\phi} - \mathbf{I}||_2^2$$ 8 it. only



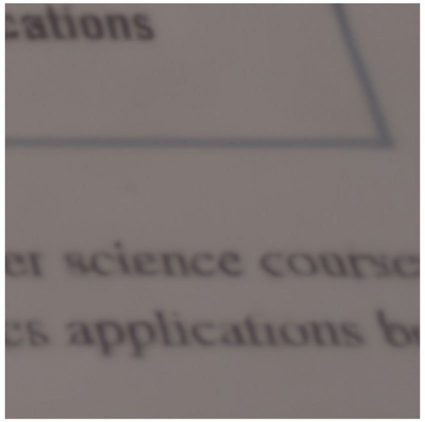
Clean Noisy: 17.42 $\mathbf{O}_0$: 22.98 $\mathbf{O}_1$: 25.22 $\mathbf{O}_2$: 27.22 $\mathbf{O}_3$: 28.54 $\mathbf{O}_4$: 29.05 $\mathbf{O}_5$: 29.05

Masked: 14.60 $\mathbf{O}'_0$: 16.93 $\mathbf{O}'_1$: 17.27 $\mathbf{O}'_2$: 17.45 $\mathbf{O}'_3$: 17.57 $\mathbf{O}'_4$: 17.64 $\mathbf{O}'_5$: 17.70

| | Input | Target | Static target | Iterations |
|---|---|---|---|---|
| DIP | random noise | noisy image | yes | 3000-5000 |
| Ours | noisy image | aux. head’s output | no | 8 |

# Experiments

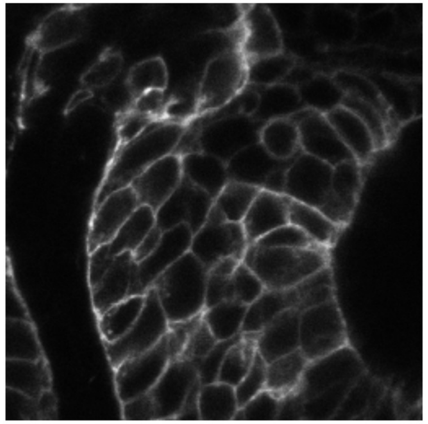
## Train on P: ImageNet images with fixed Gaussian noise (variance 0.005)

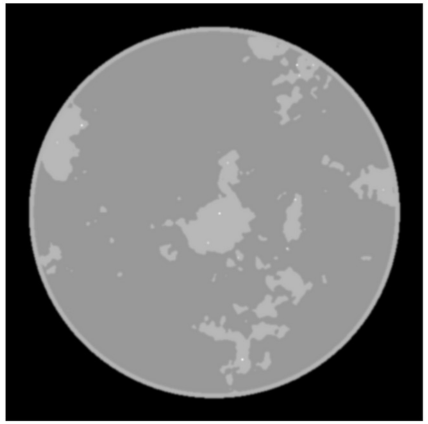


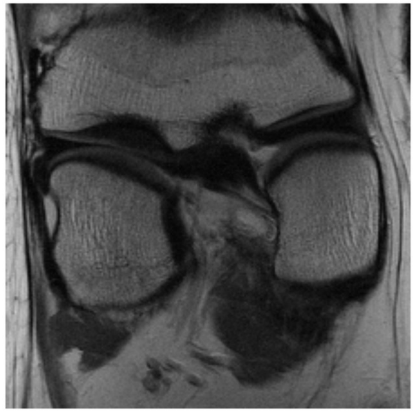
cations

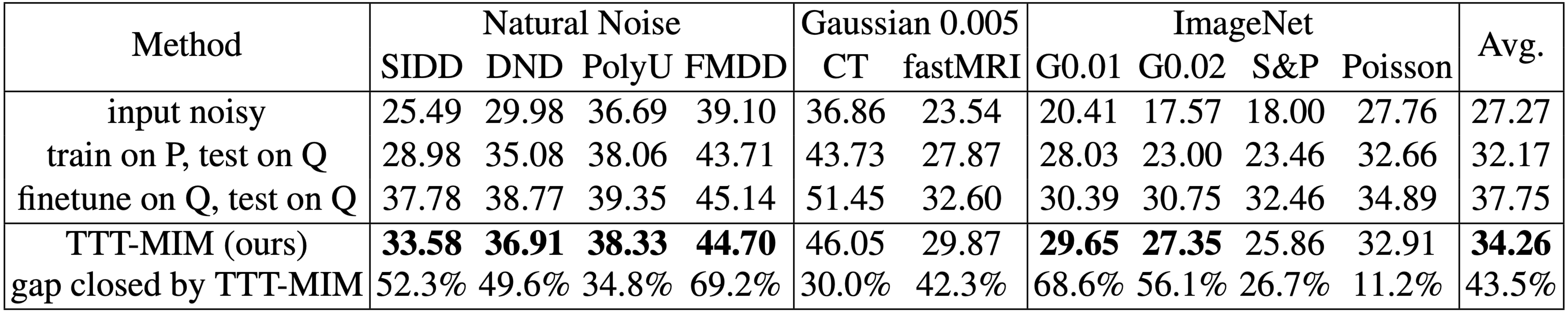
er science course

cs applications b



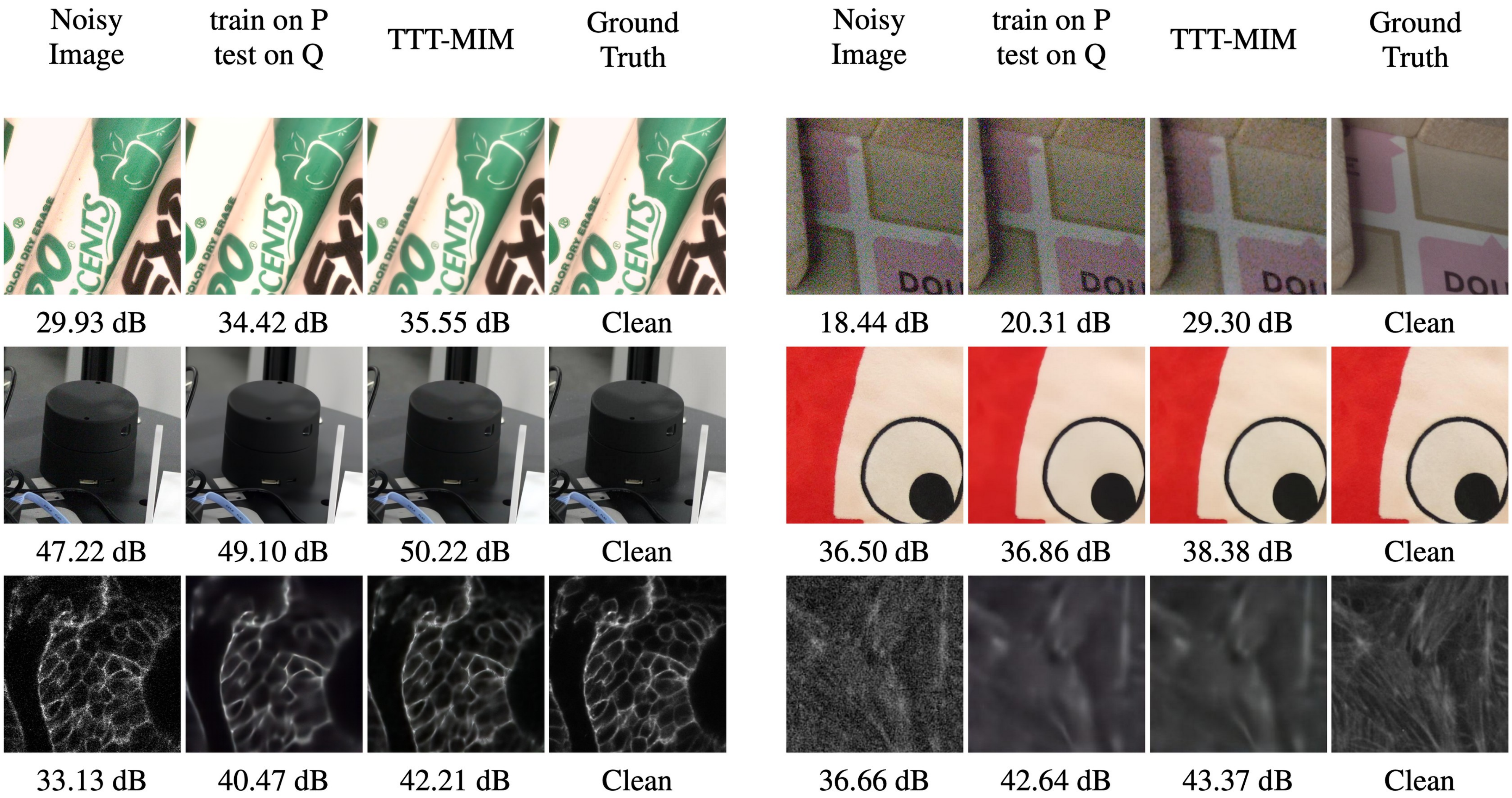




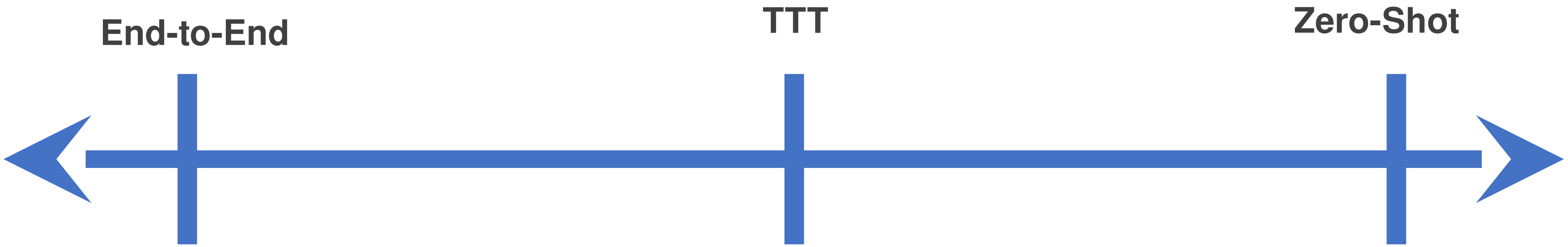


| Method | Natural Noise | | | | Gaussian 0.005 | | ImageNet | | | | Avg. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | SIDD | DND | PolyU | FMDD | CT | fastMRI | G0.01 | G0.02 | S&P | Poisson | |
| input noisy | 25.49 | 29.98 | 36.69 | 39.10 | 36.86 | 23.54 | 20.41 | 17.57 | 18.00 | 27.76 | 27.27 |
| train on P, test on Q | 28.98 | 35.08 | 38.06 | 43.71 | 43.73 | 27.87 | 28.03 | 23.00 | 23.46 | 32.66 | 32.17 |
| finetune on Q, test on Q | 37.78 | 38.77 | 39.35 | 45.14 | 51.45 | 32.60 | 30.39 | 30.75 | 32.46 | 34.89 | 37.75 |
| TTT-MIM (ours) | **33.58** | **36.91** | **38.33** | **44.70** | 46.05 | 29.87 | **29.65** | **27.35** | 25.86 | 32.91 | **34.26** |
| gap closed by TTT-MIM | 52.3% | 49.6% | 34.8% | 69.2% | 30.0% | 42.3% | 68.6% | 56.1% | 26.7% | 11.2% | 43.5% |

| | DIP | S2S | ZS-N2N | Ours |
|---|---|---|---|---|
| Iterations | 5000 | 150k | 3000 | 8 |
| Denoising time | 7 mins | 1.2 hrs | 30 secs | <1 sec |


Noisy Image
train on P test on Q
TTT-MIM
Ground Truth
Noisy Image
train on P test on Q
TTT-MIM
Ground Truth
29.93 dB
34.42 dB
35.55 dB
Clean
18.44 dB
20.31 dB
29.30 dB
Clean
47.22 dB
49.10 dB
50.22 dB
Clean
36.50 dB
36.86 dB
38.38 dB
Clean
33.13 dB
40.47 dB
42.21 dB
Clean
36.66 dB
42.64 dB
43.37 dB
Clean

# Conclusion



- Fast (milli seconds)
- Training set dependent

- Slow (minutes-hours)
- Training set independent