Unsupervised Deep Learning for Microcopy Imaging Restoration

Nianyi Li

Assistant Professor

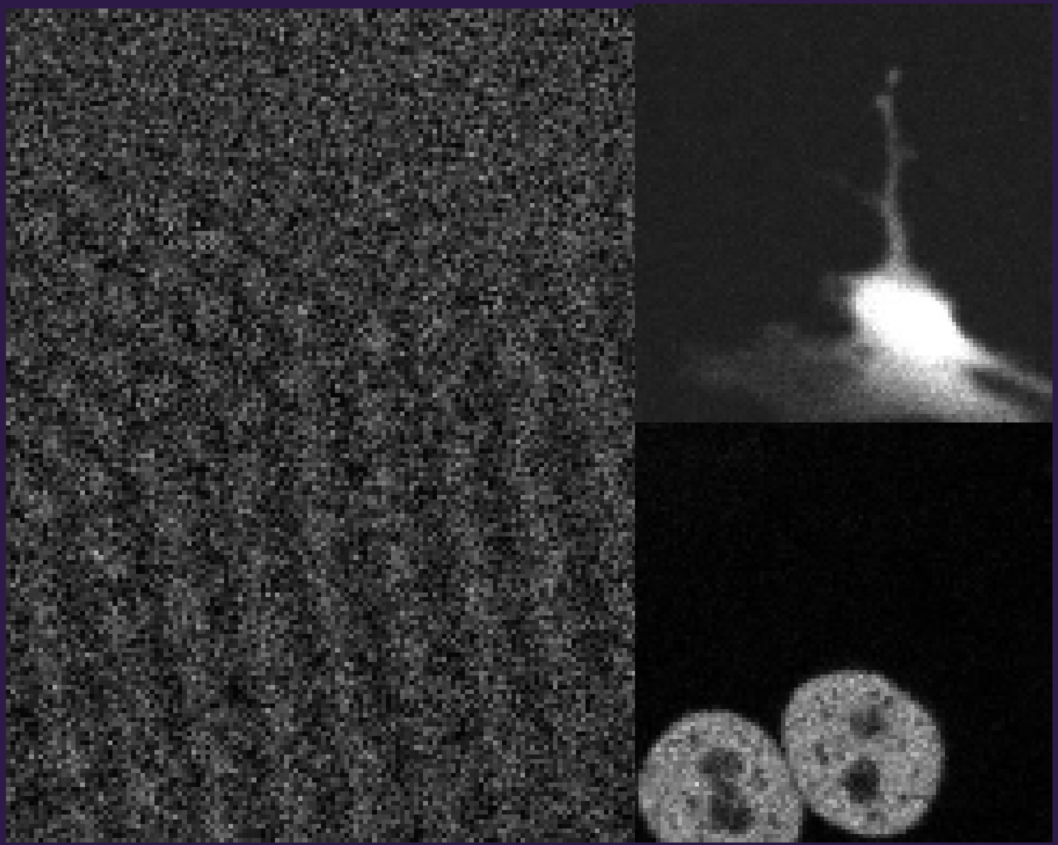
School of Computing

CLEMSON

Motivation

Challenges with Microscopy Imaging

- Low signal-to-noise ratio
- Different noise types
- Multi-modality
- Motion Blur

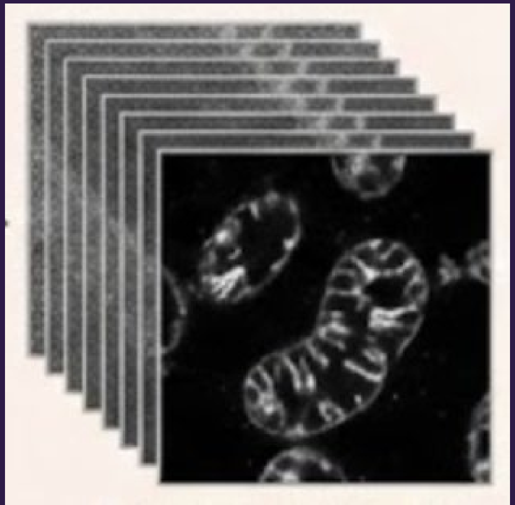


Motivation

Challenges with Microscopy Imaging

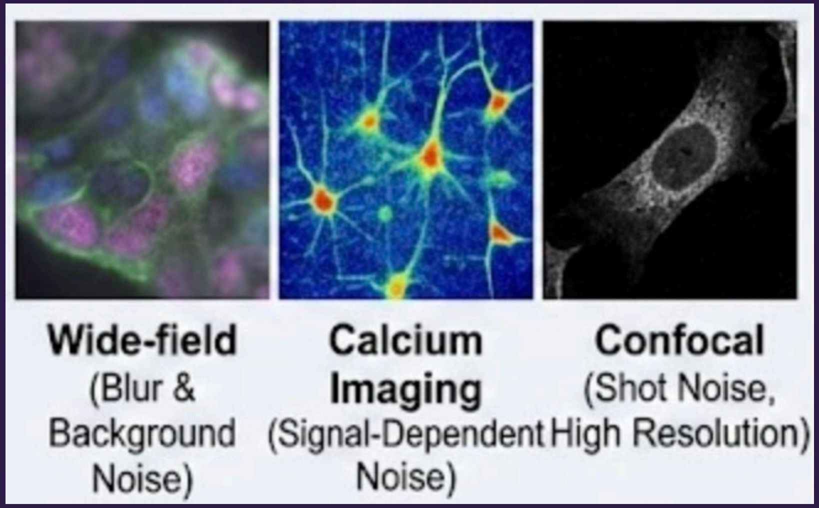
Small data problem:

Most denoising models require thousands of frames



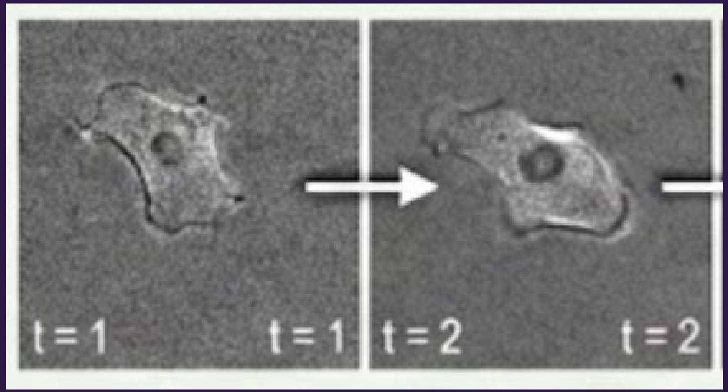
Modality diversity:

Noise varies across wide-field, calcium, and confocal imaging, limiting cross-modality generalization.



Temporal complexity:

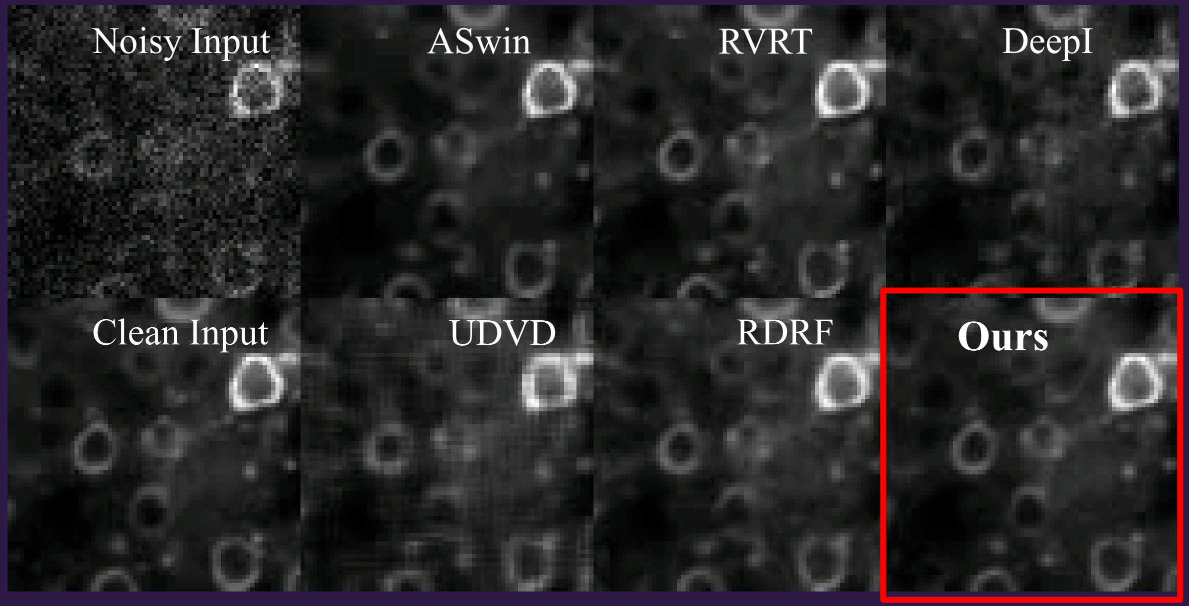
Balancing noise removal and motion preservation in complex or rapid motion scenarios.



Moving cells

Related Works

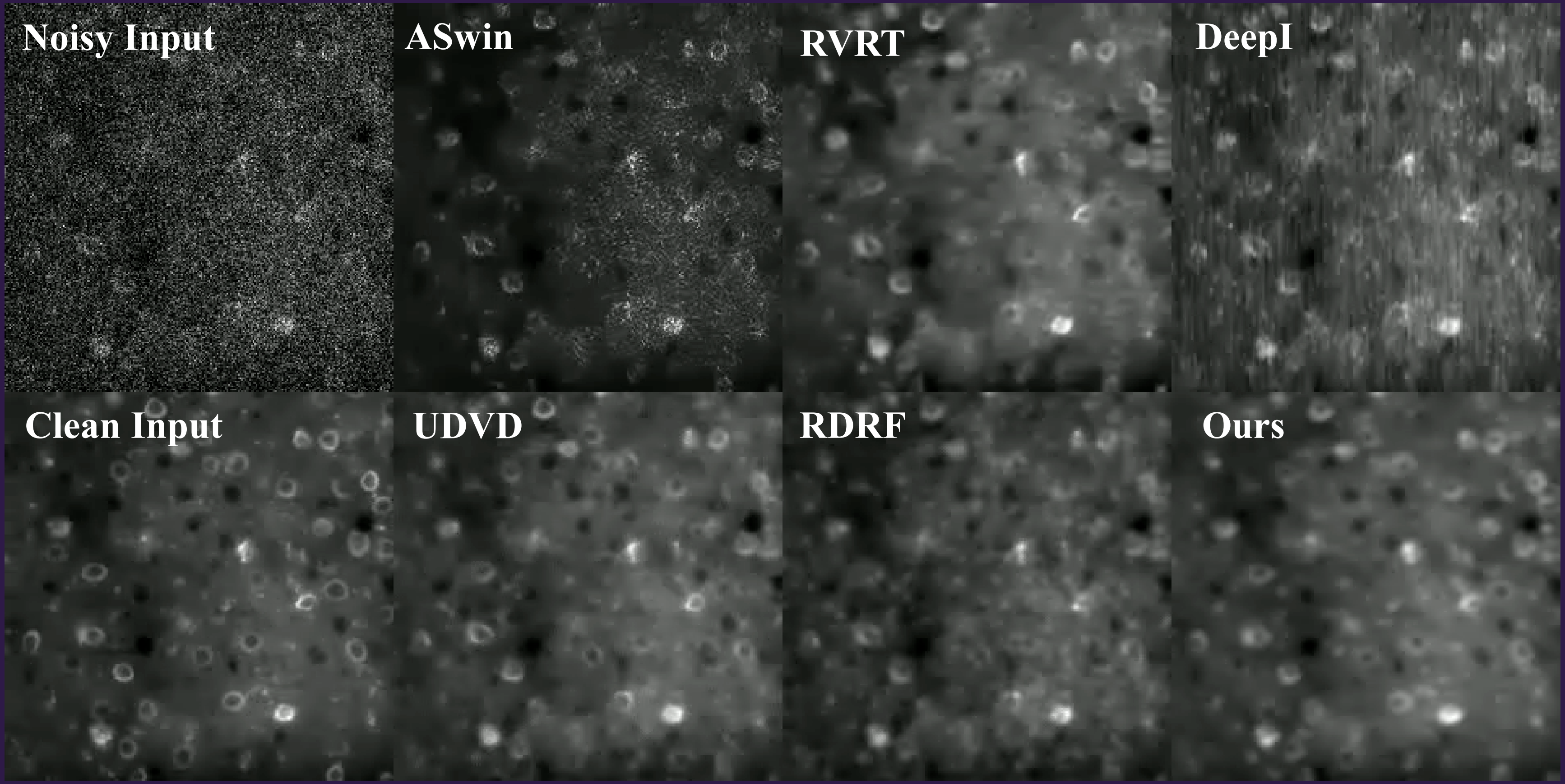
- Employs Blindspot mechanism to prevent identity noise mapping
- Extensively exploits neighboring pixels, which can lead to overfitting in microscopy images [1,2,3].
- Most existing denoising techniques are primarily designed for Gaussian noise [4,5,6].



[1] Lecoq et al., Removing independent noise in systems neuroscience data using deepinterpolation. Nature methods, 2021.
[2] Sheth et al., Unsupervised deep video denoising. ICCV, 2022
[3] Wang et al., Recurrent self-supervised video denoising with denser receptive field. In Proceedings of the 31st ACM ICM, 2023.
[4] Song et al., Neural anatomy and optical microscopy (naomi) simulation for evaluating calcium imaging methods. Journal of neuroscience methods 2
[5] Liang et al., Recurrent video restoration transformer with guided deformable attention. Neurips 2022
[6] Lindner et al., Lightweight video denoising using aggregated shifted window attention. ICCV, 2023.

Previous Solutions

Unsupervised/Self-Supervised Deep Learning solutions are not efficient and accurate



Outline

- Unsupervised Microscopy Video Denoising
- Unsupervised Denoising for Multi-Modal Microscopy Imaging for Small Data Scenarios

Unsupervised Microscopy Video Denoising

Our solution – Removing the Noise

- Calcium imaging is used for recording neural activity over a period
- Presence of noise can impact the detection of potential neural activities in the recordings
- Some learning-based approaches [1,2] for denoising require noisy/clean pair of images to train. However, ground truth clean images are usually unavailable in medical imaging
- We used an **unsupervised** learning-based approach trained on noisy data by learning the spatio-temporal relationship in the data

[1] HC. Burger, CJ. Schuler, and S. Harmeling (2012). Image denoising: Can plain neural networks compete with bm3d? CVPR
[2] Pascal V., Hugo L., Isabelle L., Y., and Pierre-Antoine M. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. JMLR

Pre-processed Denoised Frame

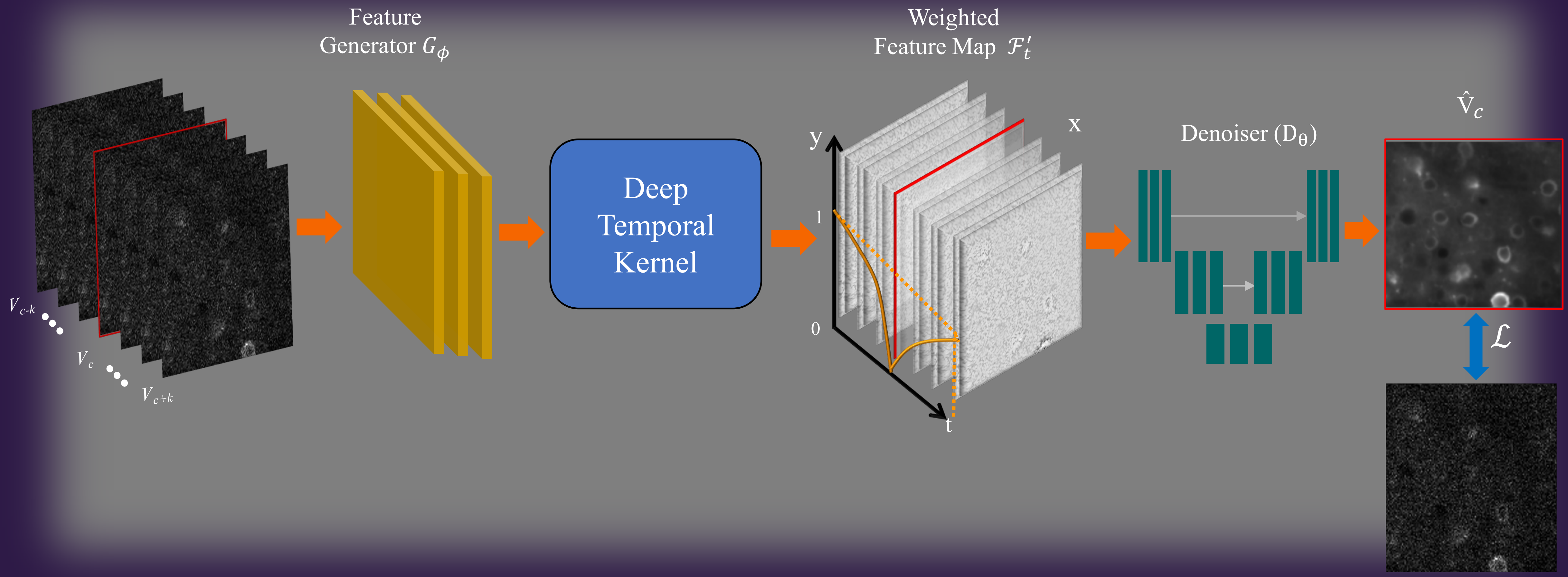
However, it will remove the subtle neuron signals and blur out the response area.

Our solution: Pre-process the input video to amplify the neuron response, then apply denoise algorithm

Pre-processed Denoised Frame

Our solution: Pre-process the input video to amplify the neuron response, then apply denoise algorithm

Our Method



Aiyetigbo et al. Unsupervised Microscopy Video Denoising. CVPRW 2024

Method

• DeepTemporal Filter (ω)

- Linear signal filter (ω) to adjust feature map weights based on temporal proximity to central frame

$$\omega(.) = \left[\left\{ \frac{k-i}{k} \right\}_{i=0}^{k-1}, \left\{ \frac{i}{k} \right\}_{i=1}^{k} \right]$$

$$\omega(.) = \{1, \dots, 0, \dots, 1\}$$

where

- $k = \lfloor N/2 \rfloor$ and N is the number of frames in the input
- $\{\omega_t | t = 1, \dots, N\}$

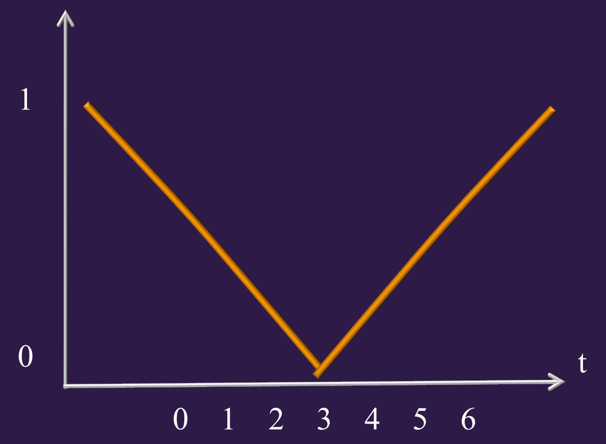


Fig 1. DeepTemporal filter shape

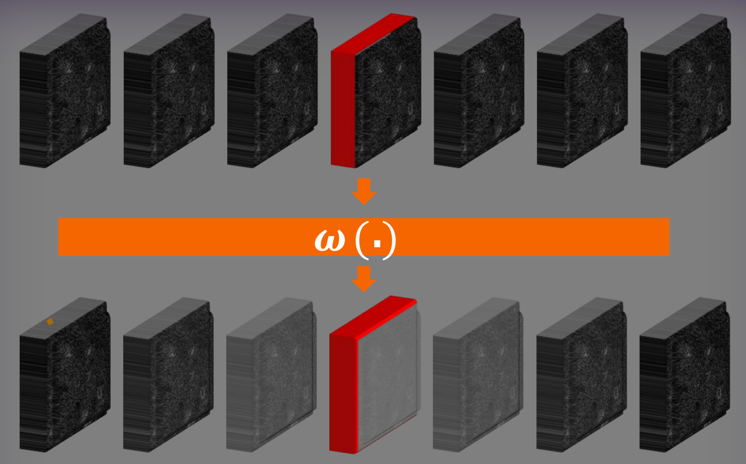


Fig 2. Illustration of our DeepTemporal filter

Method

- **Feature Generator**
  - Given a stack consecutive N input frame $\mathcal{V}_t \in \mathbb{R}^{N \times H \times W \times 1}$
  - Applies group convolution layers to extract features of each frame.

$$\mathcal{F}_t = \mathrm{G}_\phi(\mathcal{V}_t) \qquad \mathcal{F}_t \in \mathbb{R}^{N \times H \times W \times C}$$

- **DeepTemporal Filter**

$$\mathcal{F}_t' = \omega_t \odot \mathcal{F}_t \quad \forall t \in \{1, \dots, N\}$$

$$\{\mathcal{F}_t' | t = 1, \dots, c, \dots, N\}$$

$$\boldsymbol{\mathcal{L}} = \left\| \hat{\mathrm{V}}_{\boldsymbol{c}} - \boldsymbol{\mathcal{V}}_{\boldsymbol{c}} \right\|_{\mathbf{2}}$$

- **Denoiser**
  - UNet architecture. Comprises of 3 encoder layers and 2 decoder layers

$$\hat{\mathrm{v}}_c = \mathcal{D}_\theta([\mathcal{F}_1', \dots, \mathcal{F}_N'])$$

Experiment

Dataset

Simulated two-photon Calcium Imaging [1]

One-photon Calcium Imaging [2]

Fluorescence Microscopy [3]

Evaluation Metrics

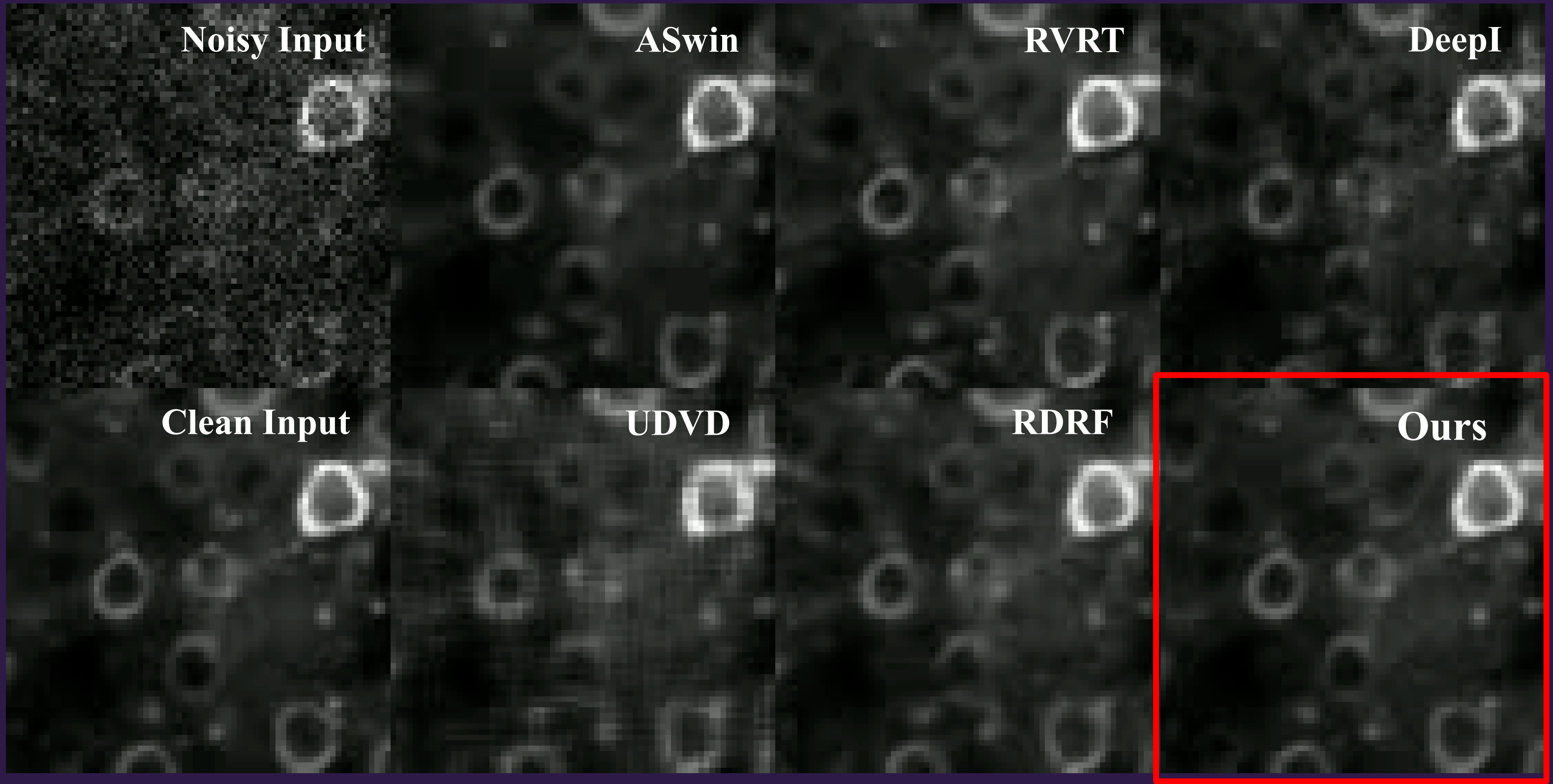
PSNR

SSIM

[1] Song et al., Neural anatomy and optical microscopy (naomi) simulation for evaluating calcium imaging methods. Journal of neuroscience methods 2021
[2] Ulman et al., An objective comparison of cell-tracking algorithms. Nature methods 2017.
[3] Kalivas Laboratory, The Medical University of South Carolina.

Results

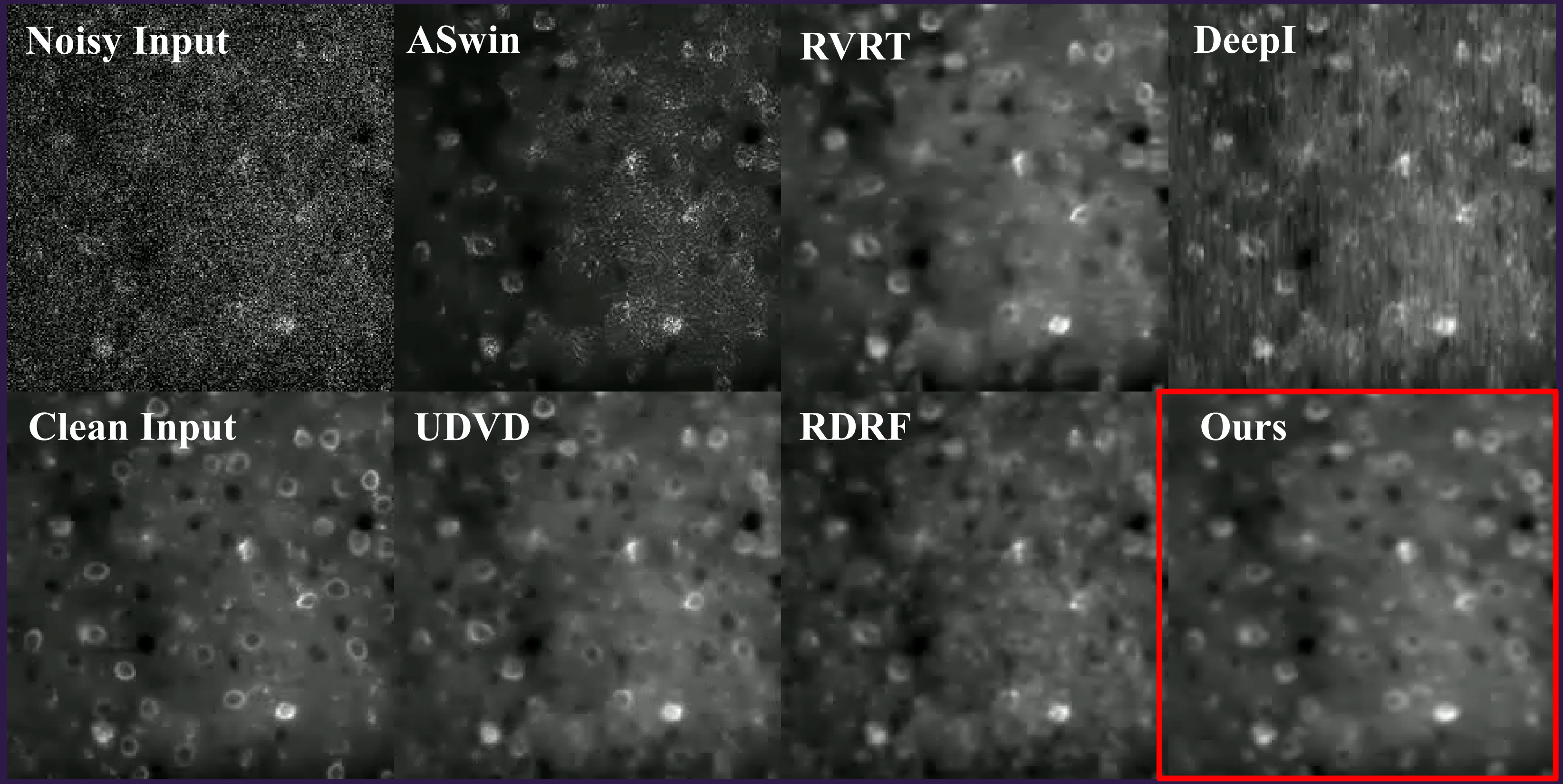
Simulated two-photon calcium imaging



[1] Lecoq et al., Removing independent noise in systems neuroscience data using deepinterpolation. Nature methods, 2021.
[2] Sheth et al., Unsupervised deep video denoising. ICCV, 2022
[3] Wang et al., Recurrent self-supervised video denoising with denser receptive field. In Proceedings of the 31st ACM ICM, 2023.
[4] Song et al., Neural anatomy and optical microscopy (naomi) simulation for evaluating calcium imaging methods. Journal of neuroscience methods 2
[5] Liang et al., Recurrent video restoration transformer with guided deformable attention. Neurips 2022
[6] Lindner et al., Lightweight video denoising using aggregated shifted window attention. ICCV, 2023.

Results

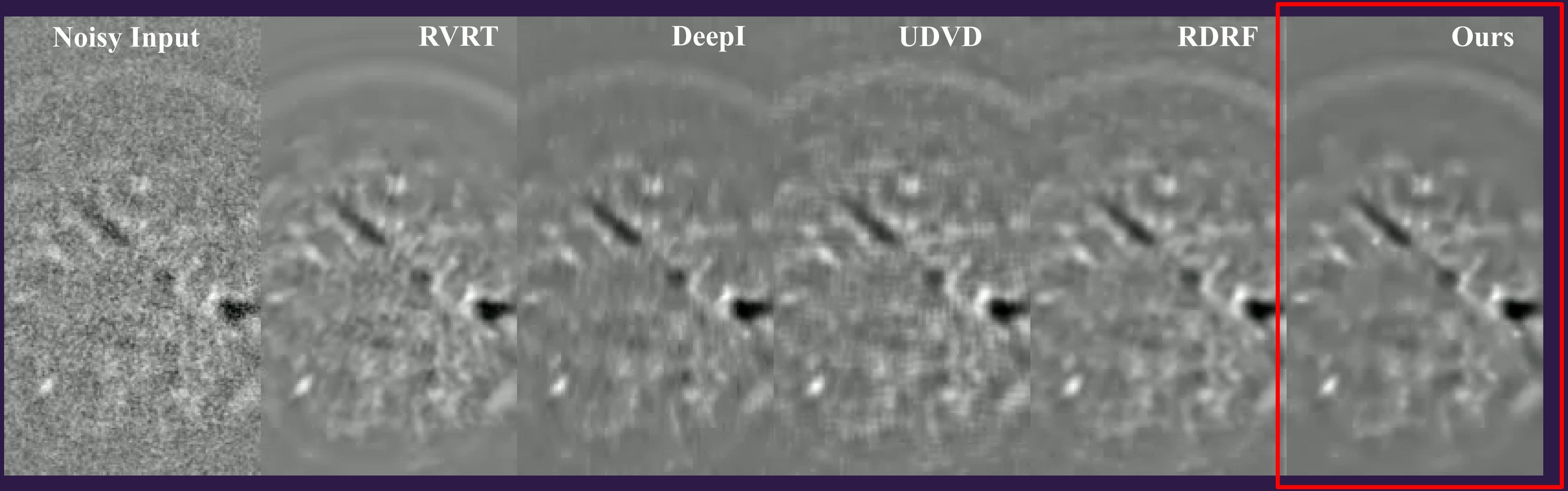
Simulated two-photon calcium imaging



Song et al., Neural anatomy and optical microscopy (naomi) simulation for evaluating calcium imaging methods. Journal of neuroscience methods 2021

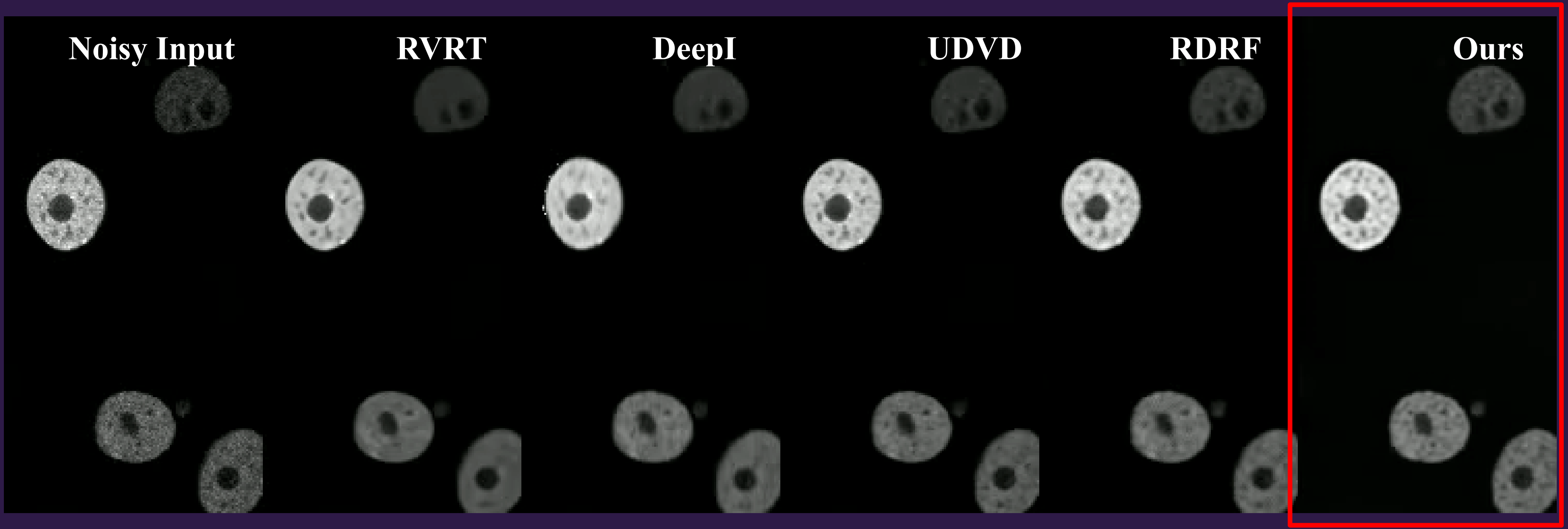
Results

One-photon calcium imaging



Results

Fluorescence Microscopy

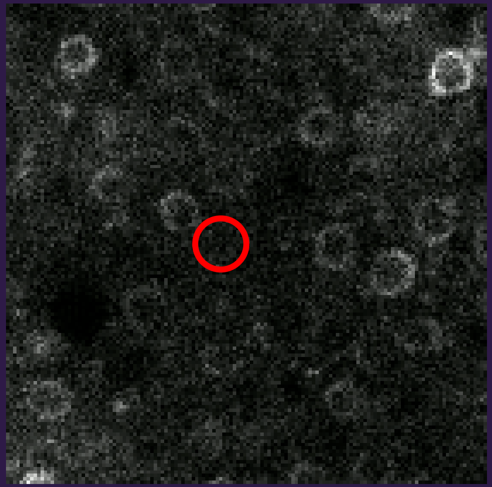


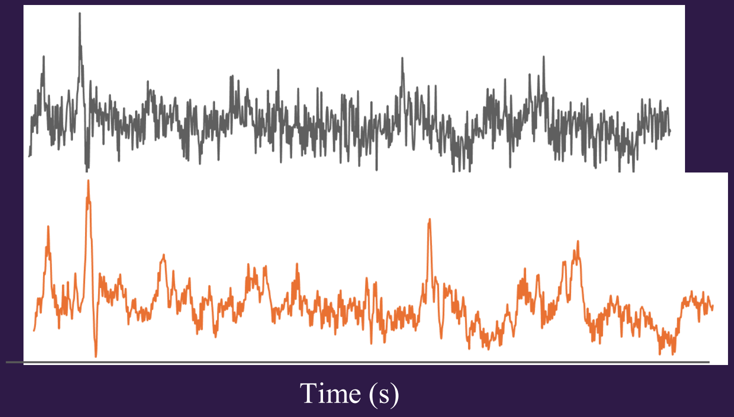
Results

- The result shows the average PSNR and SSIM values on the 2P imaging simulated with NAOMi[1]

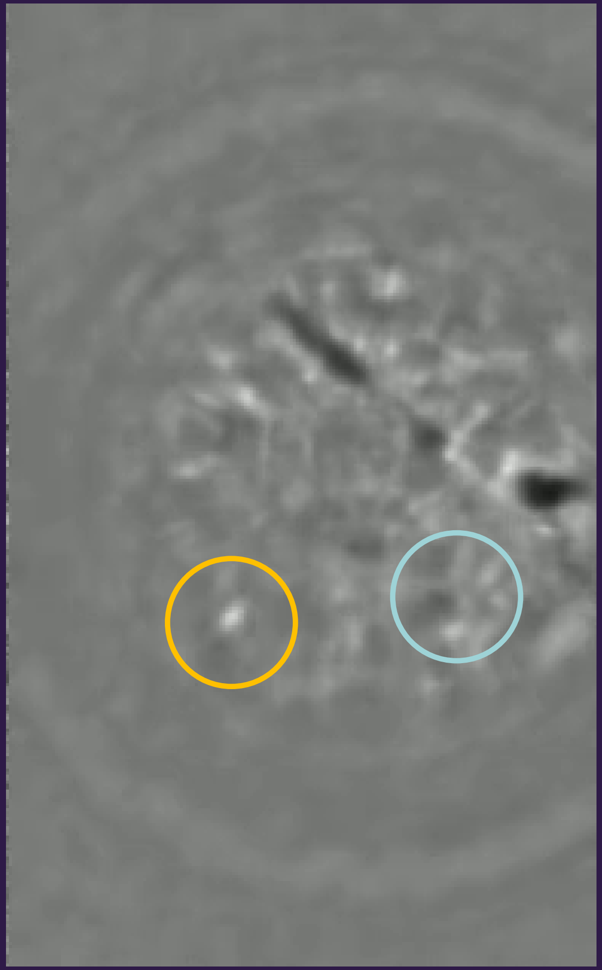
	RVRT	ASwin	DeepI	UDVD	RDRF	Ours
PSNR	<u>30.05</u>	25.94	28.64	25.75	28.31	**35.90**
SSIM	<u>0.92</u>	0.90	0.91	0.74	<u>0.92</u>	**0.95**

- Calcium traces of an example ROI from the two-photon calcium imaging



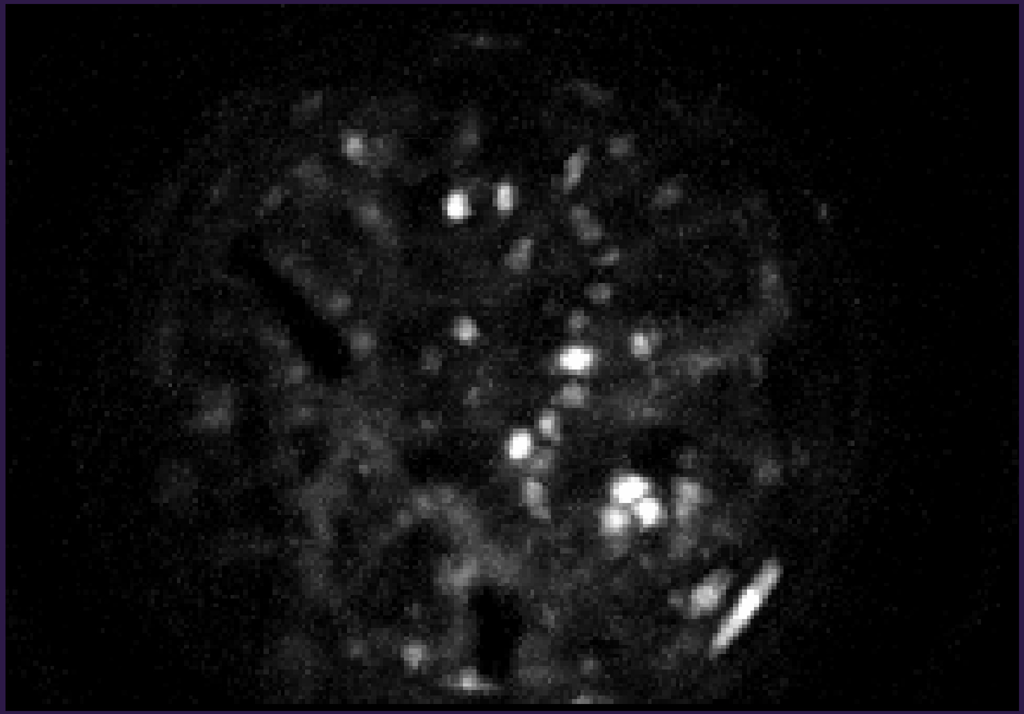


Results

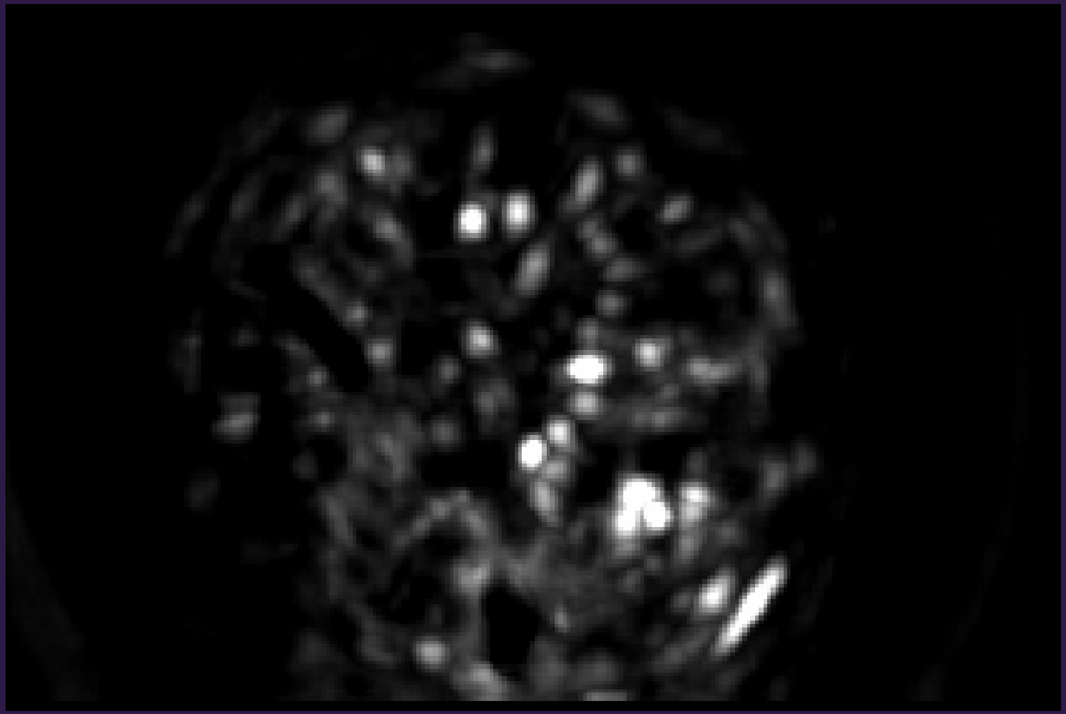


Effectively removed the noise while preserved the subtle neuron activity!

Maximum Projection

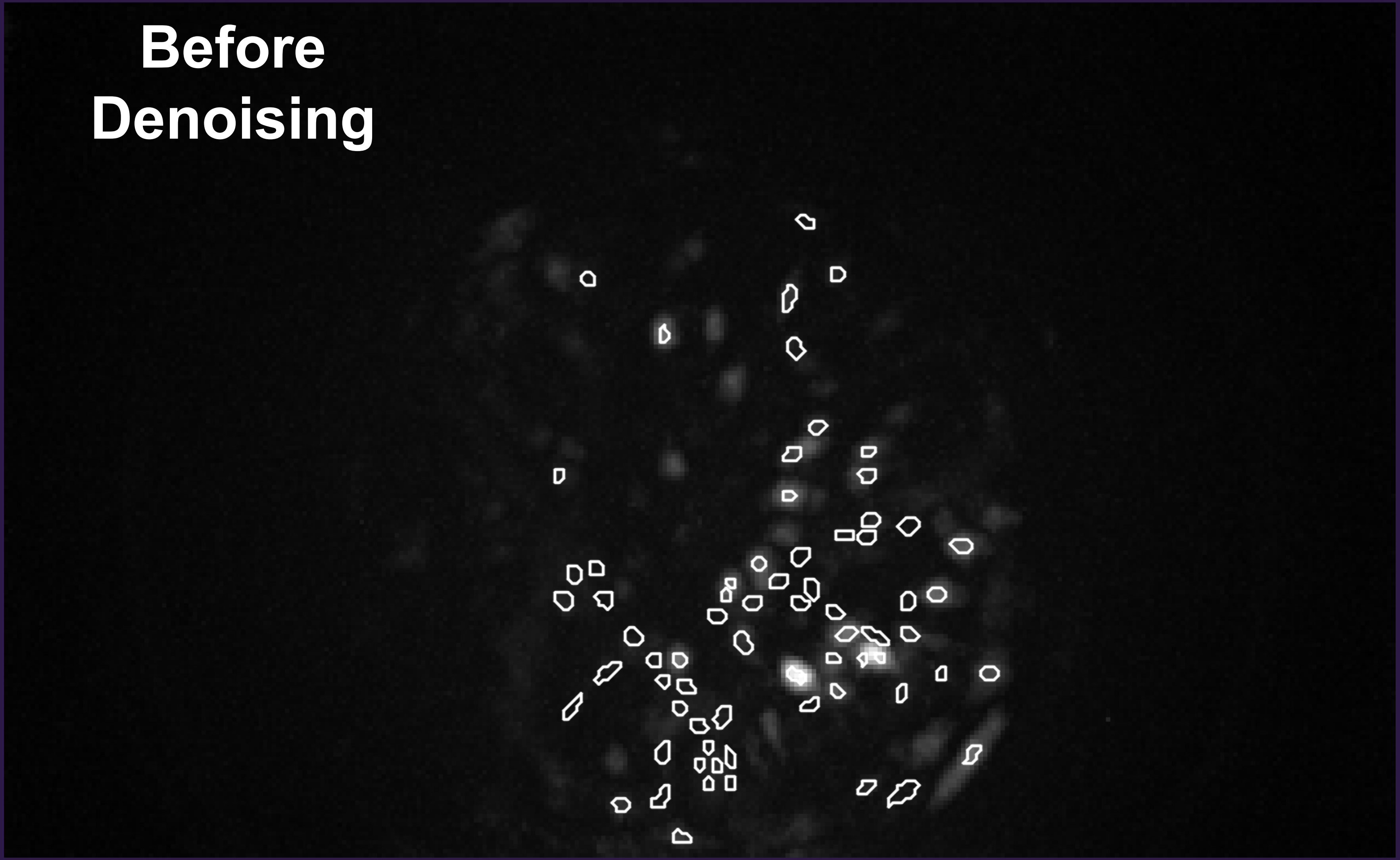


Original

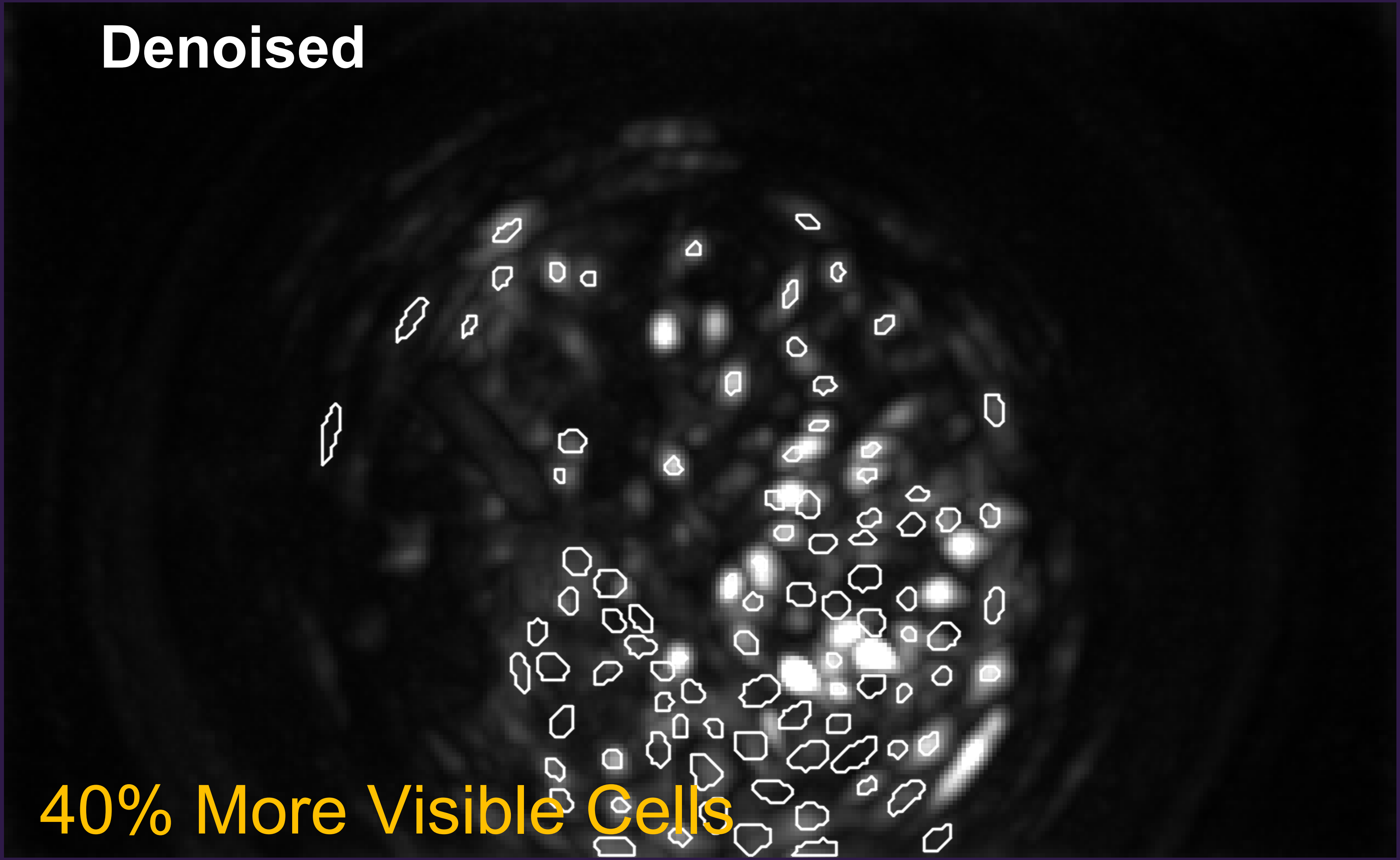


Ours

Results – Cell Tracking Rate



Results – Cell Tracking Rate

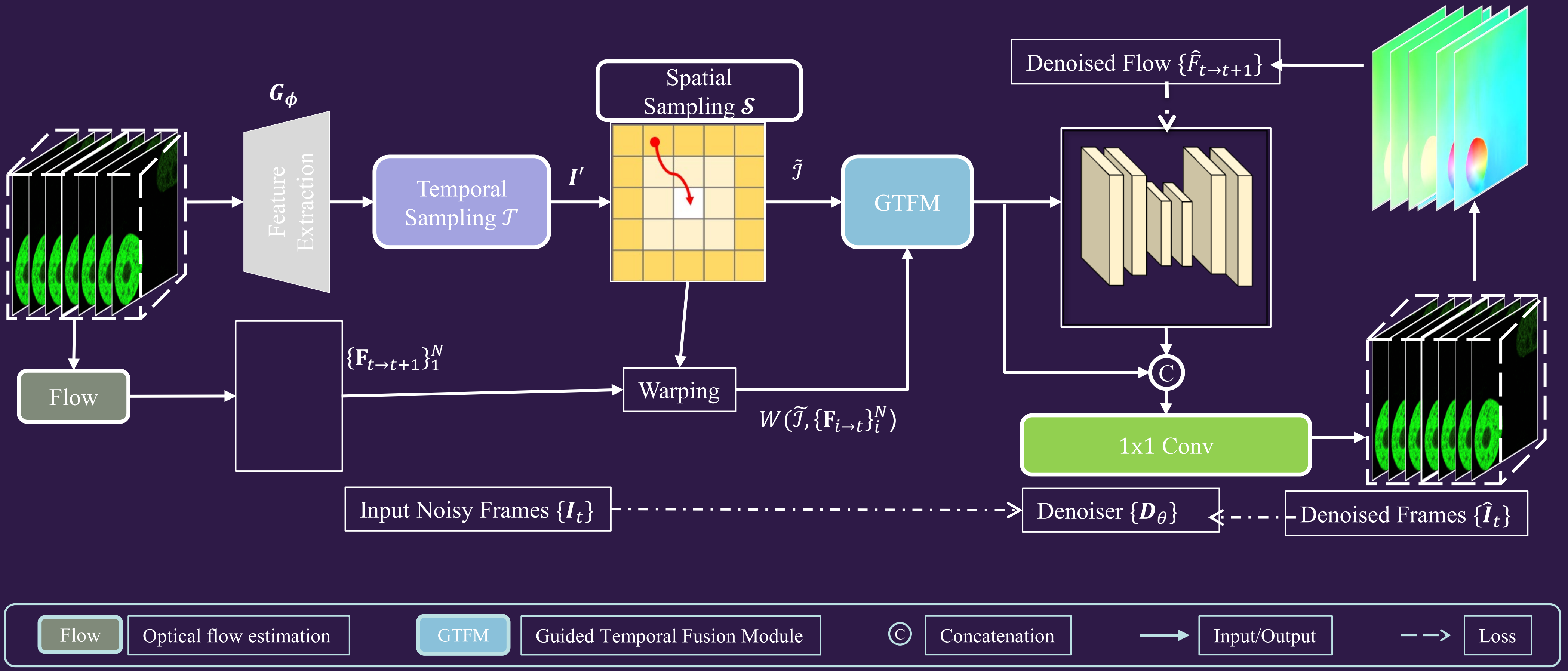


Conclusion

- We presented a novel unsupervised denoising method utilizing DeepTemporal Interpolation.
- Our approach demonstrates strong generalization across a variety of medical imaging and color datasets and shows robustness to different noise types and intensities.
- Limiting motion handling, due to the fixed temporal filter.

Unsupervised Denoising for Multi-Modal Microscopy Imaging for Small Data Scenarios

Method



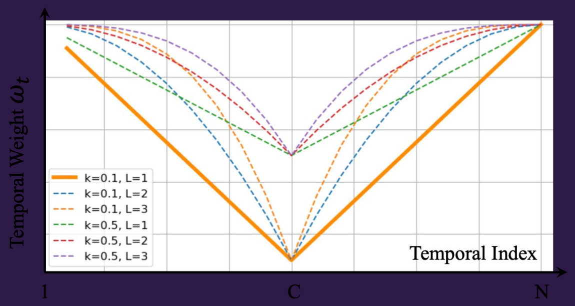
G_ϕ
Feature Extraction
Temporal Sampling $\mathcal{T}$
I'
Spatial Sampling $\mathcal{S}$
$\tilde{\mathcal{J}}$
GTFM
Denoised Flow $\{\hat{F}_{t\to t+1}\}$
Flow
$\{\mathbf{F}_{t\to t+1}\}_1^N$
Warping
$W(\tilde{\mathcal{J}}, \{\mathbf{F}_{i\to t}\}_i^N)$
C
1x1 Conv
Input Noisy Frames $\{I_t\}$
Denoiser $\{D_\theta\}$
Denoised Frames $\{\hat{I}_t\}$
Flow
Optical flow estimation
GTFM
Guided Temporal Fusion Module
C
Concatenation
Input/Output
Loss

Method

Weighted Temporal Sampling

1. Given the Magnitude $\bar{M}$ of optical flow, $\{F_{i\to t}\}_i^N$, we compute the initial minimum weight κ_0.

$$\kappa_0 = 0.2 \cdot \left(1 - e^{-\eta \times \bar{M}}\right) \qquad \kappa = \max\left(0, \kappa_0\left(1 - \frac{\text{epoch}}{\text{max\_epoch}}\right)\right)$$

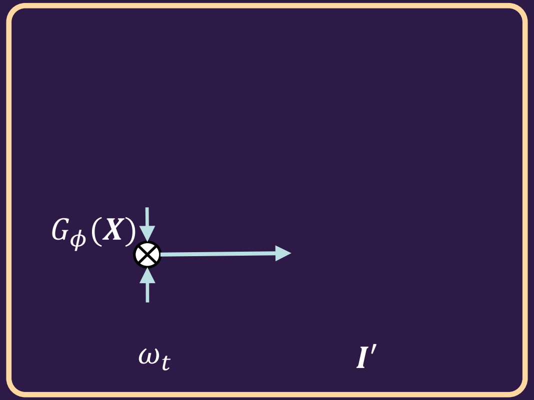


2. We compute the temporal weights $\{\omega_t | t = 1, \ldots, N\}$

$$\omega_t = \begin{cases} \kappa + (1-\kappa)\left(1 - \left(\frac{t}{C}\right)^L\right), & \text{if } t \le C \\ \kappa + (1-\kappa)\left(1 - \left(\frac{2C - t}{C}\right)^L\right), & \text{otherwise,} \end{cases}$$

3. We apply the temporal weights to the features

$$I'_t = \omega_t \odot G_\phi(X_t) \qquad \forall t \in 1, \ldots, N$$

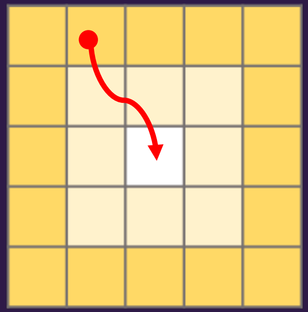


Method

Weighted Spatial Sampling

- **Replaces randomly selected pixels p with neighboring pixels q**

$$P(q|p) = \frac{e^{-\alpha \cdot d(p,q)}}{\sum_{q' \in \Omega_p} e^{-\alpha \cdot d(p,q')}}$$



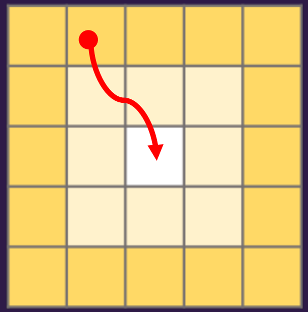
where:
$P(q|p)$ is the probability of choosing a neighboring pixel
$d(p,q)$ is the l_1 distance between pixels **p** and **q**

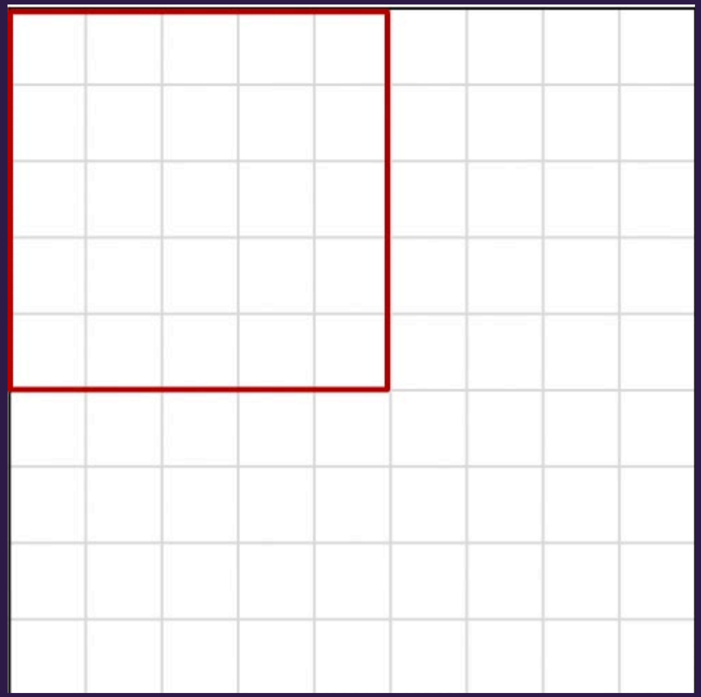
Method

Weighted Spatial Sampling

- **Replaces randomly selected pixels p with neighboring pixels q**

$$P(q|p) = \frac{e^{-\alpha \cdot d(p,q)}}{\sum_{q' \in \Omega_p} e^{-\alpha \cdot d(p,q')}}$$





where:
$P(q|p)$ is the probability of choosing a neighboring pixel
$d(p,q)$ is the l_1 distance between pixels **p** and **q**

Method

Guided Temporal Fusion

1. Given spatiotemporal sampled features, we first warp the neighboring features to the central frame and the mean warped $\bar{W}$.

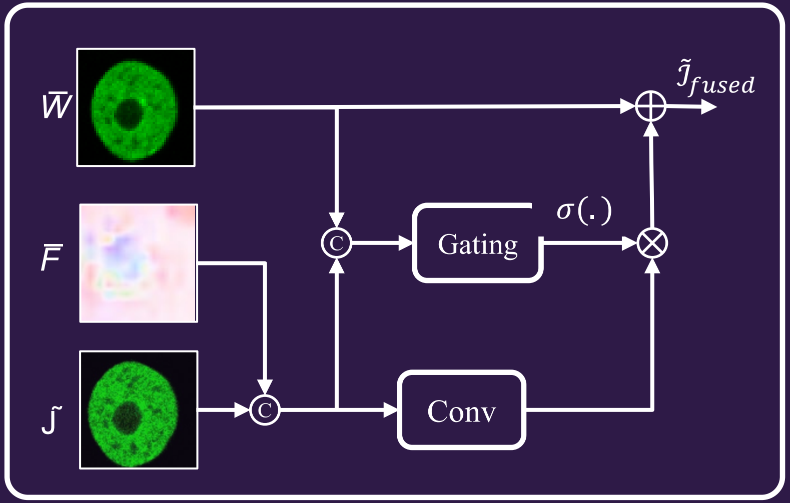
$$\bar{W} = \frac{1}{N\_1} \sum_{i \neq t}^{N} warp(\tilde{J}_i, \mathbf{F}_{i \to t})$$

2. We concatenate the mean flow $\bar{F}$, mean warped $\bar{W}$ and the central feature and pass to gating network to yield spatial attention map $\sigma(.)$

3. We concatenate the mean flow $\bar{F}$ and the central feature through convolution to get residual candidate.

4. Finally, we fused the temporal features using:

$$\tilde{J}_{fused} = \bar{W} + \sigma * \Delta\tilde{J}_t$$



Method

Recurrent Optimization

- To effectively guide the network during training, we employ two loss terms. One is the pixel-wise L1 loss between $\hat{I}_t$ and the reference frame I_t^{ref}, and the second loss term is a consistency loss based on the optical flow estimation of the denoised frames.

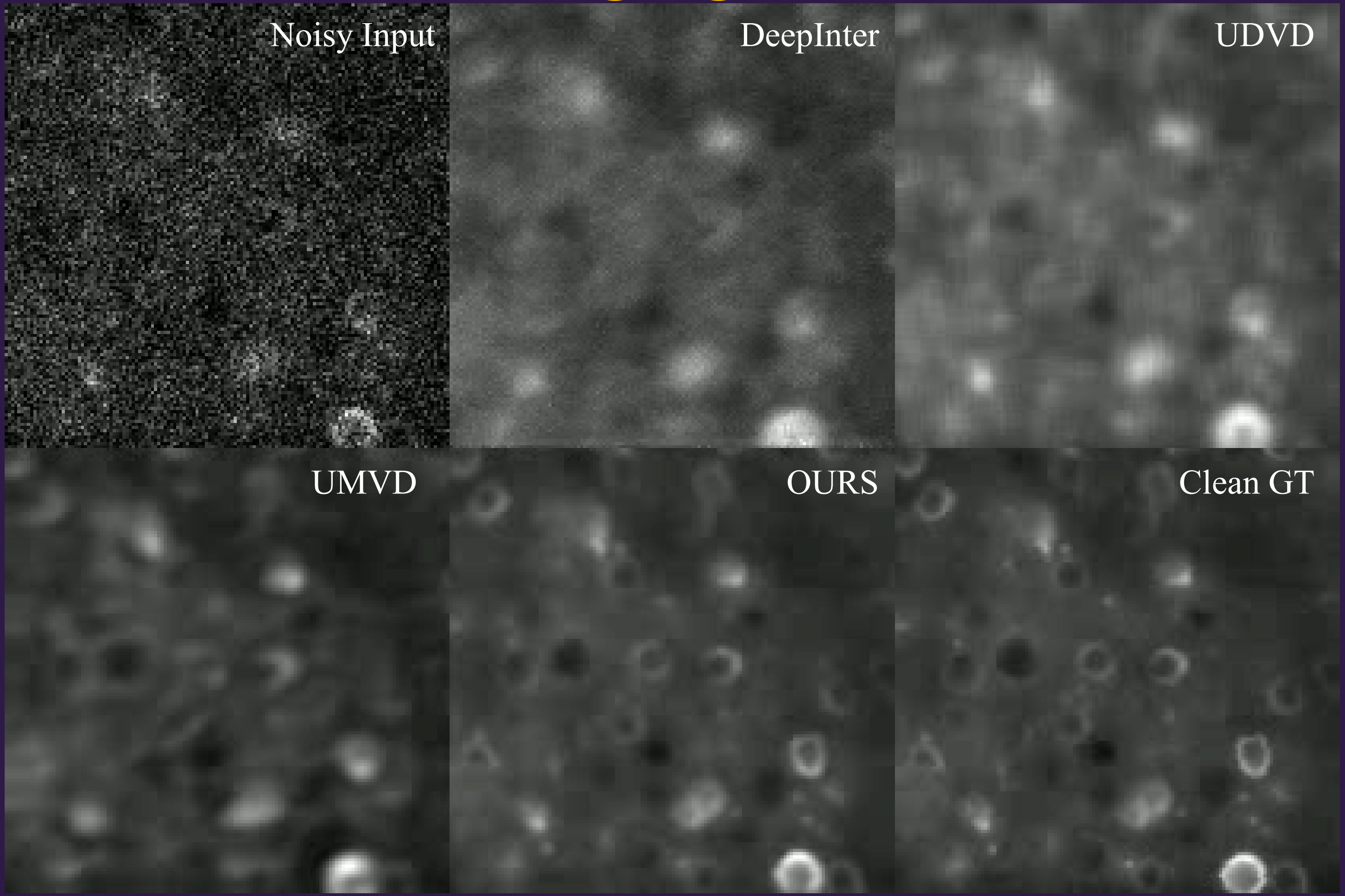
$$\mathcal{L}_1 = \sum_{p \in I_t} \left| \hat{I}_t(p) - I_t^{ref}(p) \right|$$

$$\mathcal{L}_2 = \frac{1}{N} \sum_{t=1}^{N-1} \left\| W(\hat{I}_{t+1}, \hat{F}_{t \to t+1}) - \hat{I}_t \right\|^2$$

$$\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2$$

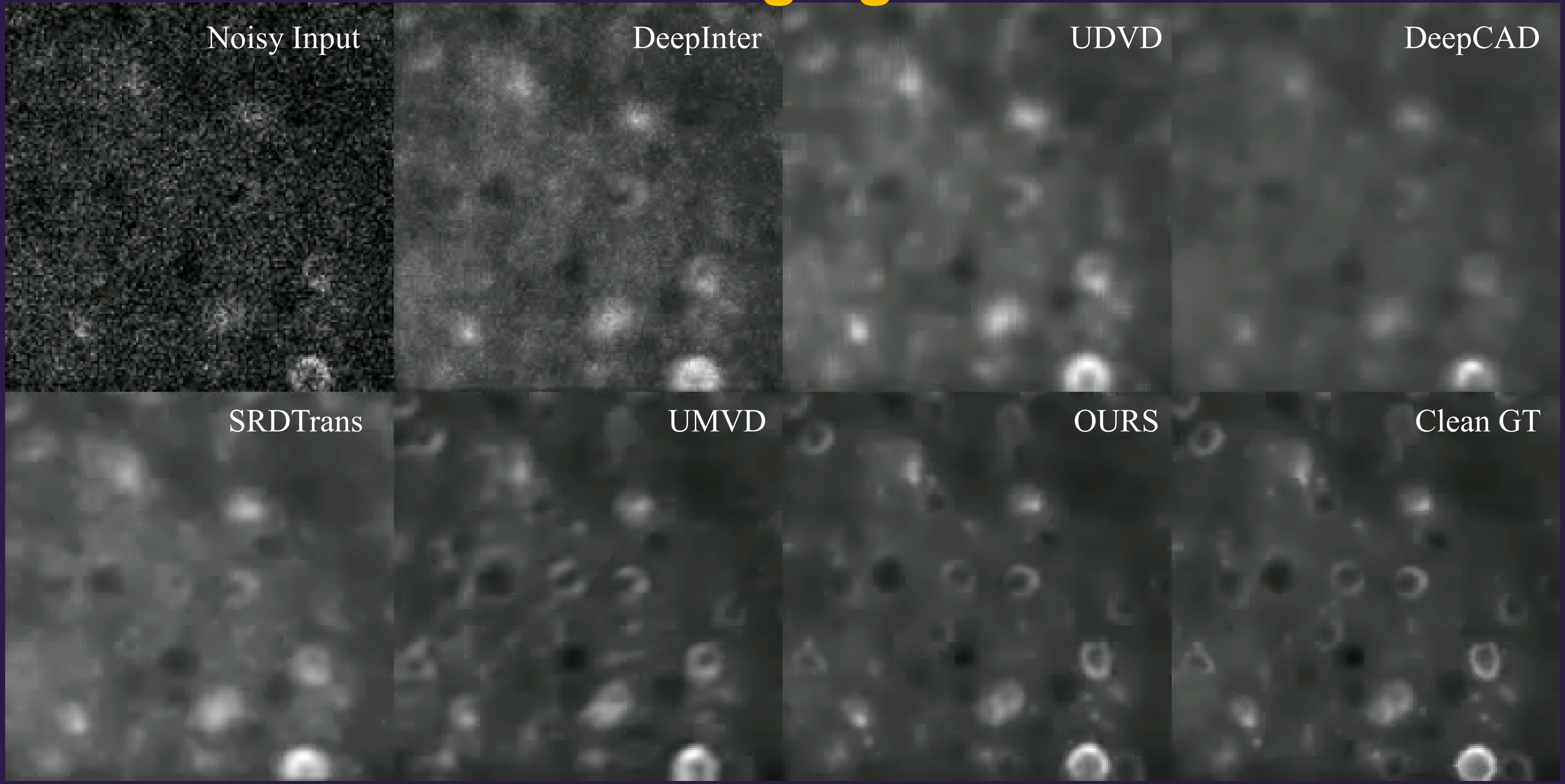
Result

Calcium Imaging – 10 Frames



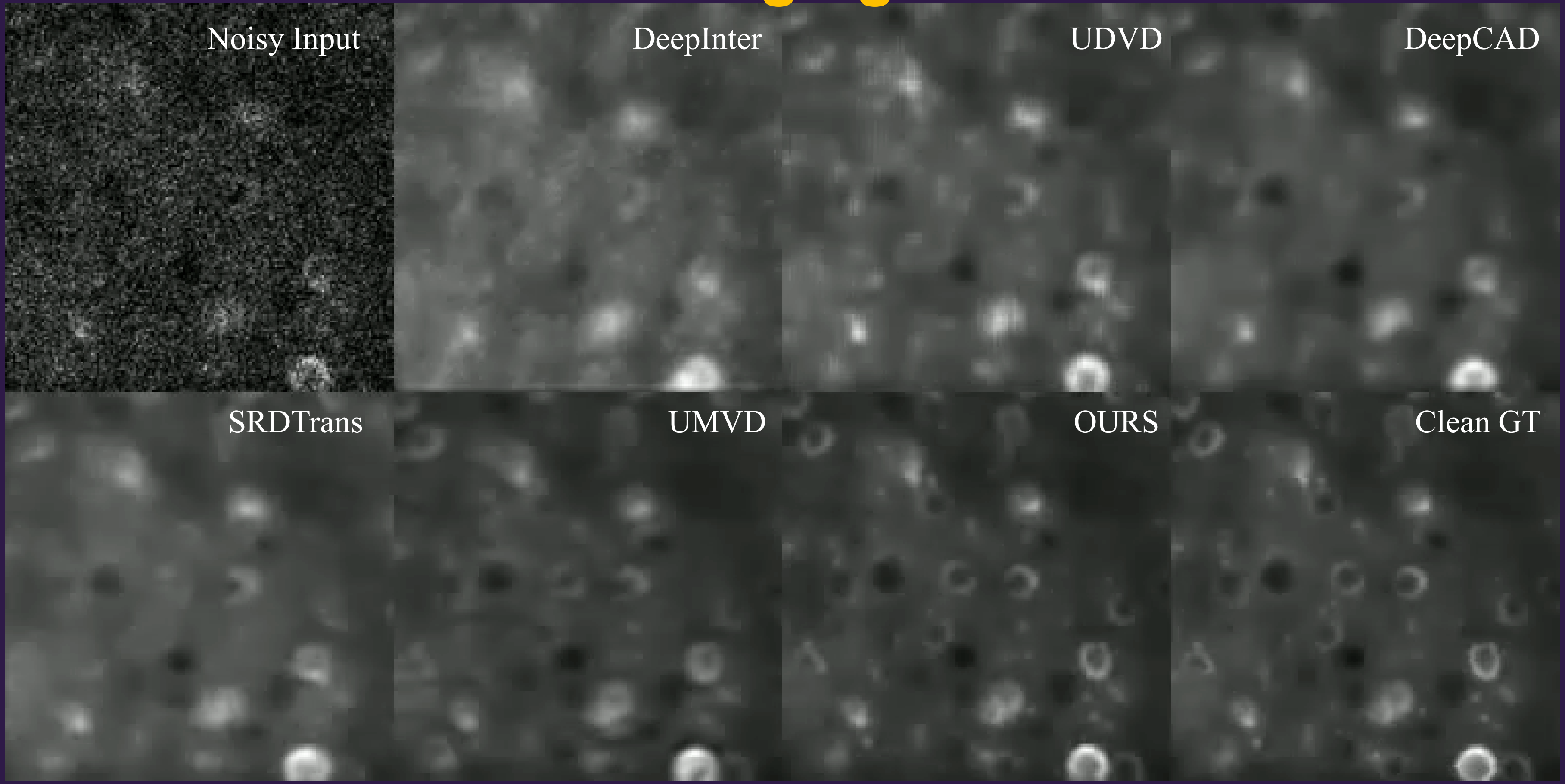
Result

Calcium Imaging – 20 Frames



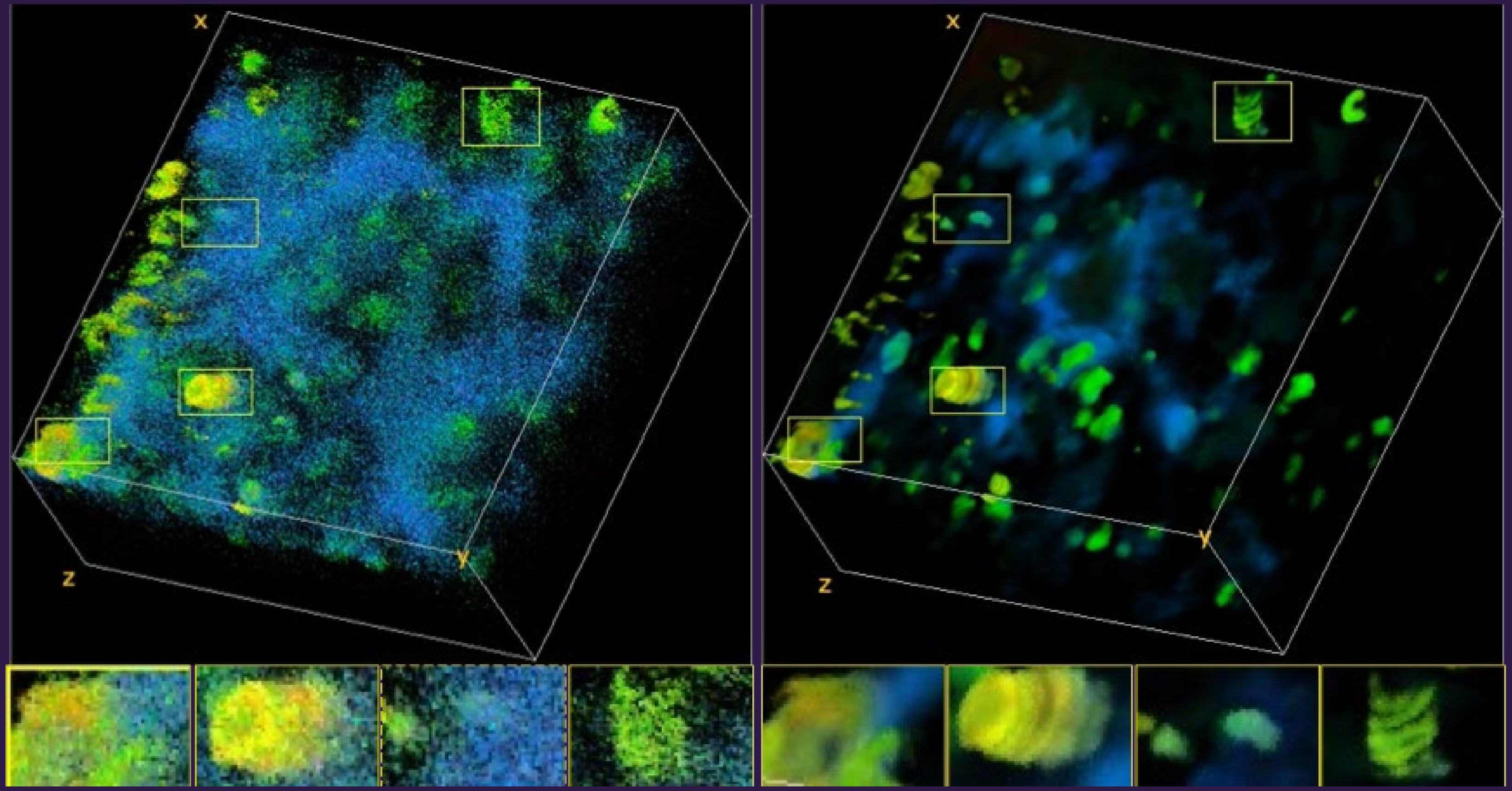
Result

Calcium Imaging – 50 Frames



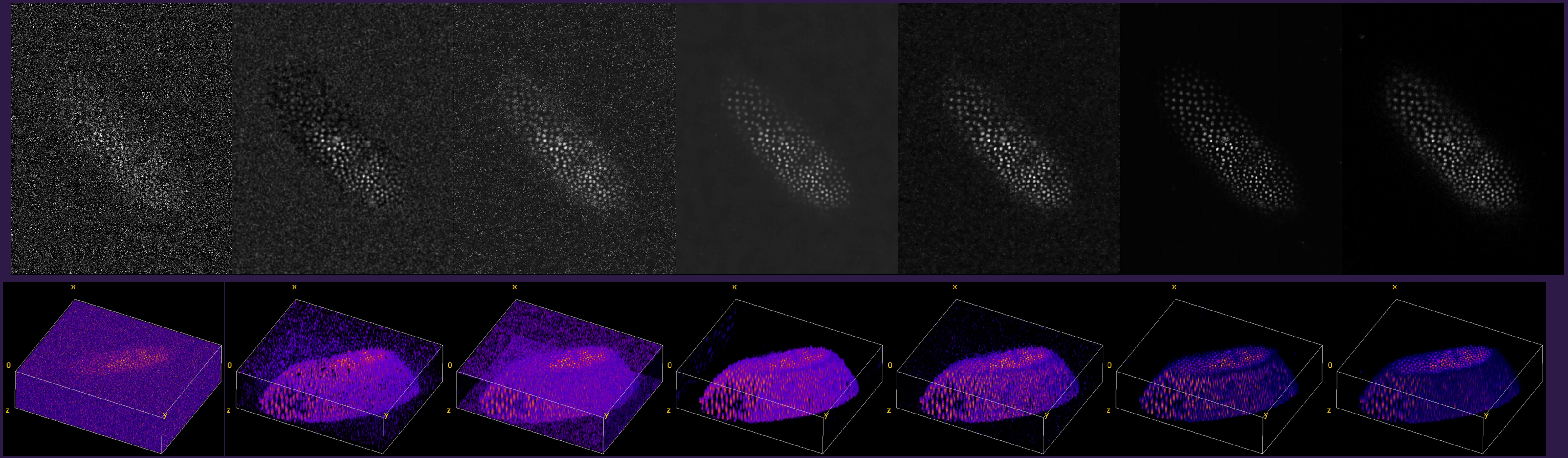
Result

Confocal Imaging



Results

Confocal Imaging



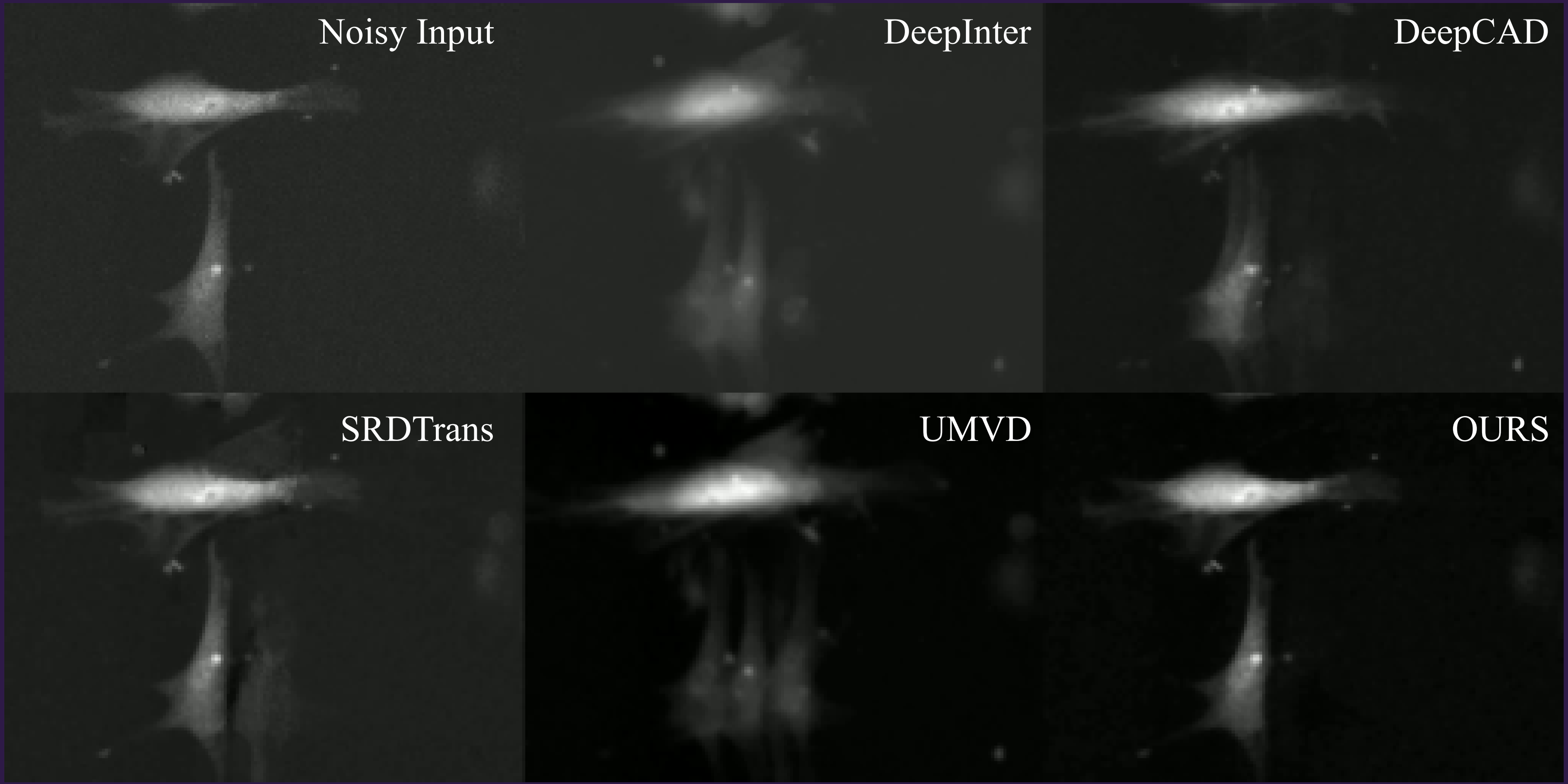
Noisy Input DeepCad DeepInterpolation UDVD UMVD SRDTrans Ours

Result

Fluorescence Microscopy

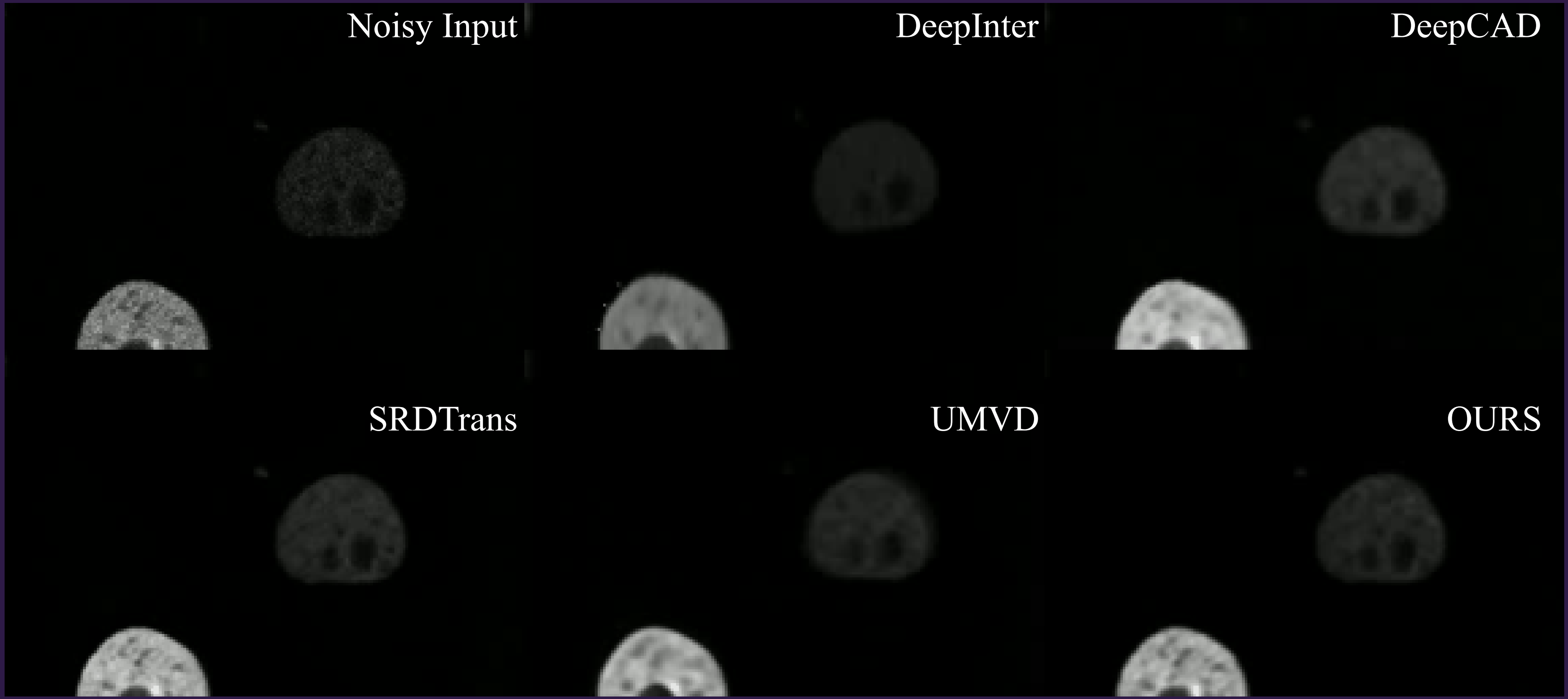
Result

Fluorescence Microscopy



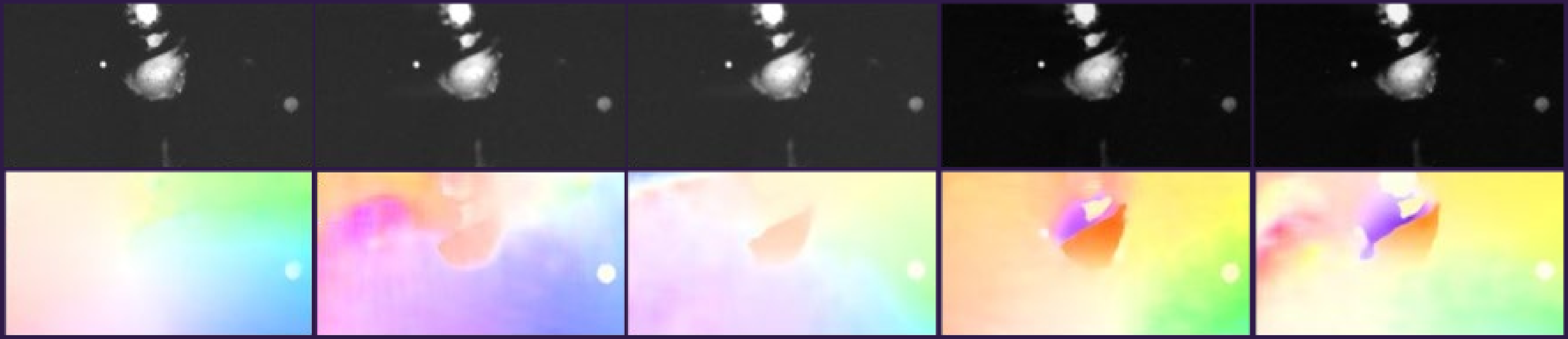
Result

Fluorescence Microscopy



Result

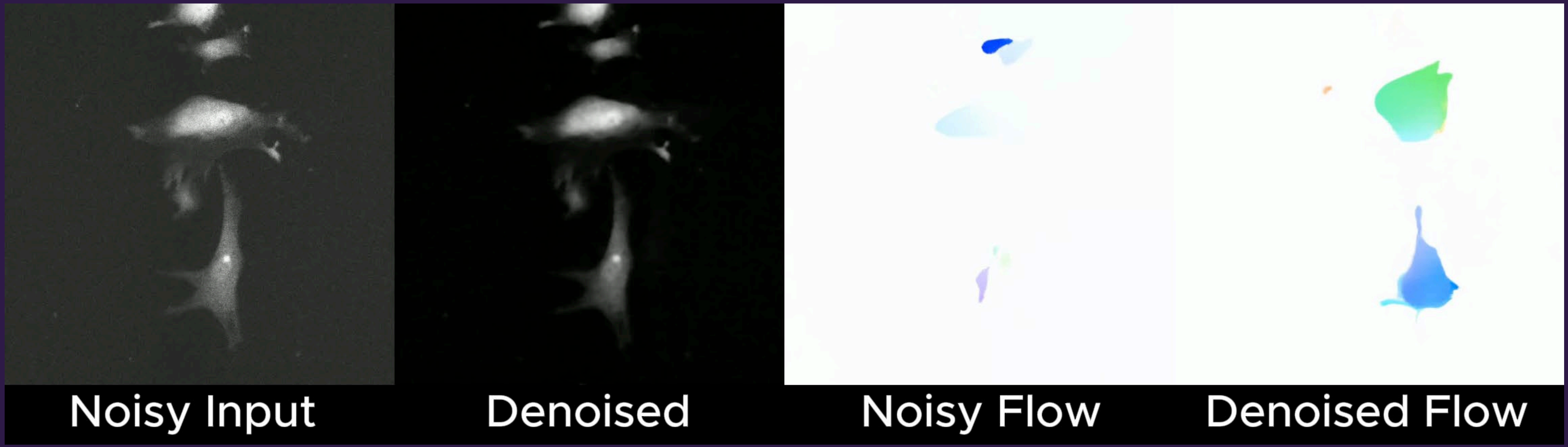
Refined Flow



Noisy Input | Epoch 1 | Epoch 2 | Epoch 4 | Denoised Output

Result

Optical Flow



Quantitative Results

- Quantitative comparison with SOTA video denoising techniques on simulated two-photon calcium imaging.

	DeepI		DeepCAD		UDVD		SRDTrans		UMVD		Ours	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
150 μm²	30.020	0.911	36.606	0.971	31.889	0.899	41.176	0.984	38.871	0.985	43.970	0.992
500 μm²	34.222	0.907	29.130	0.885	35.775	0.932	34.595	0.924	32.933	0.909	37.501	0.952
Average	32.121	0.909	32.868	0.928	33.832	0.916	<u>37.886</u>	<u>0.954</u>	35.902	0.947	**40.736**	**0.972**

Conclusion

- We presented a novel unsupervised denoising method which dynamically adapts to observed motion in videos by utilizing weighted SpatioTemporal sampling.
- Our approach demonstrates strong generalization across a variety of medical imaging datasets.
- To improve scalability to high-resolution or real-time settings, future work will explore lightweight optical flow models and adaptive sampling strategies.

Thanks to all Collaborators and Sponsors!

Mary Aiyetigbo
Ph.D student,
Clemson University

Feng Luo
Professor
Clemson University

Peter Kalivas
Professor
MUSC

Tong Ye
Associate Professor
Clemson University

NSF

DEVCOM
GROUND VEHICLE
SYSTEMS CENTER

Wanqi Yuan
Ph.D Student
Clemson University

Reda Chalhoub
Medical Doctor
MUSC

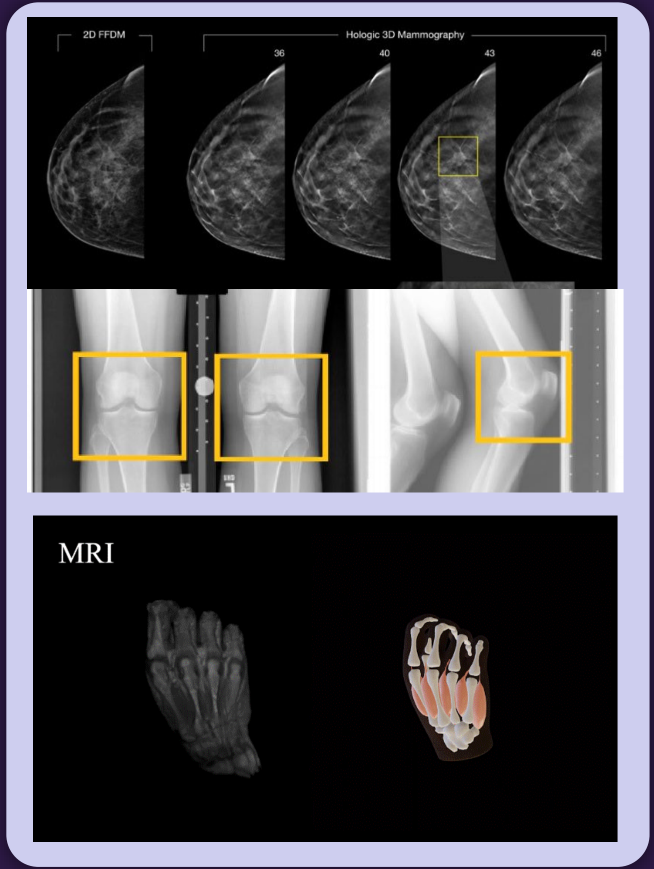
Ethan Anderson
Graduate Student
Clemson University

Alexander Korte
Undergraduate
Clemson University

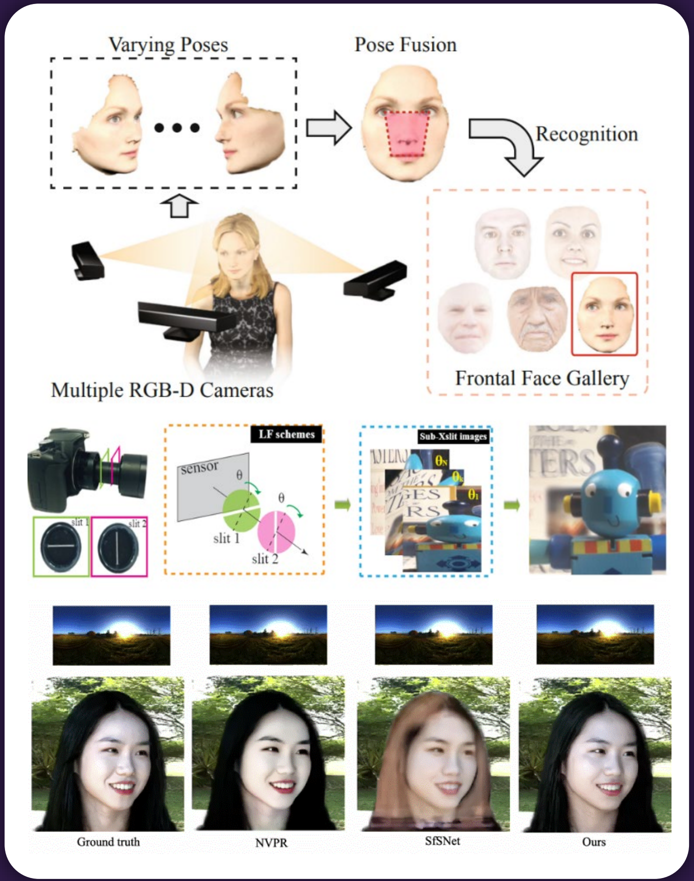
SC EPSCoR

Advanced Imaging Laboratory

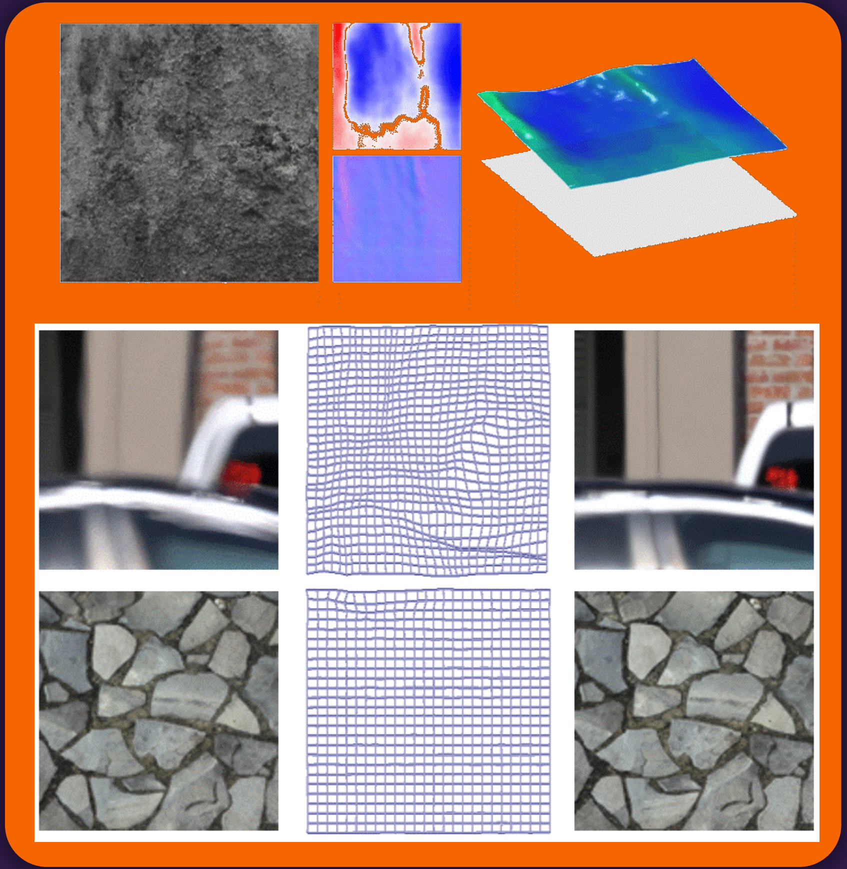
Medical Image Processing



Computer Vision



Physics-informed Machine Learning



Email: nianyil@clemson.edu

THANK YOU

Project Website

https://nianyil.people.clemson.edu/