Evaluating Established Denoising Methods for Voltage Imaging

Comparison of SUPPORT, DeepCAD-RT, and PMD when applied to voltage imaging data

Jiayi Wang[1]
Supervisor(s): Nergis Tömen[1]**, Alejandro Castañeda Garcia**[1]
[1]EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 22, 2025

Name of the student: Jiayi Wang
Final project course: CSE3000 Research Project
Thesis committee: Nergis Tömen, Alejandro Castañeda Garcia, Chirag Raman

An electronic version of this thesis is available at http://repository.tudelft.nl/.

Abstract

Voltage imaging using genetically encoded voltage indicators (GEVIs) enables high-speed, population-scale monitoring of neural activity, but it suffers from significant noise due to low photon yield and high frame rates. Effective denoising is essential to recover meaningful signals from such data. In this study, we present a comparative evaluation of three state-of-the-art denoising methods, SUPPORT, DeepCAD-RT, and PMD, on both synthetic and real voltage imaging datasets. Our analysis considers spatial and temporal signal quality, as well as computational efficiency. We find that each method has distinct advantages, and the most suitable choice depends on the specific requirements of the imaging application. SUPPORT is well-suited for tasks requiring spatial detail, PMD offers strong temporal stability and speed, and DeepCAD-RT provides an efficient, balanced alternative. These insights aim to support researchers in selecting and refining denoising tools for real-world neuroscience workflows.

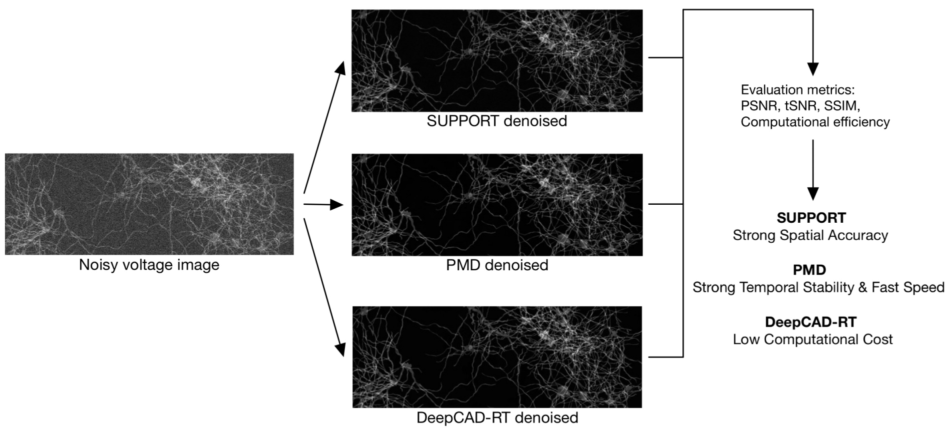


Figure 1: Overview of the project. The pipeline begins with noisy voltage imaging inputs and applies each method to assess performance in terms of spatial accuracy (PSNR, SSIM), temporal stability (tSNR), and computational efficiency (inference speed and memory usage). Key findings highlight SUPPORT's strength in spatial reconstruction, PMD's advantage in temporal denoising and runtime, and DeepCAD-RT's computational efficiency.

1 Introduction

Voltage imaging with genetically encoded voltage indicators (GEVIs) has become an important tool in modern neuroscience. It provides a powerful optical method to monitor rapid changes in membrane potential. This enables high-speed, population-scale monitoring of the electrical activity of neurons with millisecond precision [1]. Despite its significance, a central limitation of this technique is that it produces very noisy results. This is due to the inherently low photon yield (light level) of GEVIs, especially under the high temporal resolution (frame rate), which is required to capture action potentials. The results are often in a poor signal-to-noise ratio (SNR), making it difficult to distinguish real neural activity from background noise. Consequently, effective denoising has become a critical step in the voltage imaging analysis pipeline.

Traditional denoising techniques, such as Gaussian smoothing or wavelet filtering, often fall short in this context. These methods tend to introduce temporal distortions or spatial blurring that can obscure fast neural signals and fine morphological features [2]. To address this, more advanced approaches specifically tailored for voltage imaging data have been introduced. These approaches use either model-based statistical methods or data-driven neural architectures. Notable among these are SUPPORT [3], DeepCAD-RT [4], and Penalized Matrix Decomposition (PMD) [5], each offering distinct trade-offs between denoising quality, temporal performance, and computational efficiency.

While these methods have demonstrated improved reconstruction performance, prior evaluations have largely centered on a limited set of image-quality metrics, and often on isolated datasets [4, 5, 6]. Less attention has been given to systematic comparisons across both synthetic and real imaging conditions. In particular, few studies evaluate computational efficiency and performance under varying noise levels. Comprehensive assessments of the trade-offs between spatial accuracy, temporal fidelity, and practical runtime efficiency are yet to be explored. This gap limits our ability to make informed choices about which denoising strategies are best suited for specific experimental needs or resource constraints.

To address this gap, we pose the central research question of this study: **How do three state-of-the-art denoising methods, SUPPORT, DeepCAD-RT, and PMD, compare in terms of spatial performance, temporal stability, and computational efficiency when applied to voltage imaging data with varying levels of noise?**

This study aims to fill that gap by conducting a comparative evaluation of three leading denoising approaches, SUPPORT, DeepCAD-RT, and PMD, on both synthetic and real voltage imaging datasets. Our evaluation framework is designed to examine each method across multiple axes: spatial reconstruction fidelity, temporal signal stability, and computational efficiency. By standardizing the evaluation conditions and applying consistent metrics, we seek to characterize each method's strengths and limitations and to provide a practical guide for researchers selecting denoising tools in voltage imaging workflows.

The remainder of this paper is organized as follows. Section 2 reviews background literature on voltage imaging and specialized denoising methods. Section 3 introduces the three methods evaluated in this study. Section 4 outlines the experimental design, datasets, and evaluation metrics. Section 5 presents the comparative results. Section 6 discusses responsible research considerations. Section 7 explores the broader implications of our findings and potential limitations. Section 8 concludes with a summary and suggestions for future work.

2 Background

Voltage imaging is a powerful optical technique that enables the direct visualization of membrane potential dynamics in neurons with high temporal resolution. By using GEVIs or synthetic voltage-sensitive dyes, it provides a means to monitor electrical activity across populations of neurons in animals, complementing traditional electrophysiological approaches [7]. The method is particularly valuable for studying fast neuronal events, such as action potentials and subthreshold fluctuations, which are critical for understanding brain function at the cellular and network levels. However, voltage imaging faces significant challenges: the fluorescent signals are typically weak and rapidly varying, making them highly susceptible to noise from photon shot noise, motion artifacts, and background fluorescence.

The presence of noise in voltage imaging data poses significant obstacles to accurate analysis and interpretation. High levels of noise can obscure rapid voltage transients, such as action potentials, leading to false negatives or missed spikes. In some cases, noise may even mimic real signals, resulting in false positives and misinterpretation of neuronal activity. Moreover, noise degrades signal-to-noise ratio (SNR), making it difficult to detect subtle subthreshold events that are essential for understanding synaptic integration and network dynamics. Temporal noise can also distort the timing of events, which is particularly problematic for studying precise spike timing and synchrony across neurons. Spatially, noise complicates the identification of active regions, especially when imaging densely packed neural tissue. As a result, without effective denoising, voltage imaging data may lead to inaccurate conclusions about neuronal behavior and network function. These difficulties necessitate robust denoising algorithms to recover reliable voltage traces without distorting the underlying physiological signals, making denoising a crucial preprocessing step in voltage imaging analysis.

Matrix factorization-based denoising algorithms, such as Principal Component Analysis (PCA) [8], Non-negative Matrix Factorization (NMF), and Penalized Matrix Decomposition (PMD) [5], have been widely used to extract structured signals from noisy voltage imaging data. These methods decompose the data matrix, typically representing spatial and temporal components, into a low-rank approximation, isolating dominant signal patterns while suppressing unstructured noise. They are computationally efficient, interpretable, and well-suited for separating overlapping neural signals in population recordings [9]. However, their effectiveness relies heavily on the assumption that neural activity lies in a low-dimensional subspace, which may not hold in the presence of complex or sparse firing patterns. Additionally, these methods often struggle with preserving fast, transient events like action potentials, as such signals may be relegated to higher-rank components or interpreted as noise. Matrix factorization also typically assumes linear signal structure and lacks the flexibility to model nonlinear dynamics or spatial inhomogeneities, limiting its performance in challenging in vivo imaging conditions [6].

Deep learning-based denoising methods have appeared as a powerful alternative to traditional model-driven approaches, offering greater flexibility and expressiveness in capturing the complex spatiotemporal structure of voltage imaging data. Unlike matrix factorization techniques, deep neural networks can learn nonlinear and high-dimensional signal representations, making them well-suited for the intricate dynamics and variability of neuronal activity. However, supervised deep learning approaches typically require paired noisy and clean data for training, an impractical requirement in voltage imaging, where high-fidelity ground truth is unavailable [10]. To overcome this limitation, researchers have developed self-supervised denoising strategies that train models directly on noisy data by leveraging statistical properties of noise. Notable examples include Noise2Noise [11], Noise2Void [12], and Noise2Self [13], which have shown impressive results in fluorescence microscopy [14]. Building on these ideas, several methods have been tailored specifically for imaging time-series data. For instance, DeepCAD-RT [4] extends the DeepCAD [15] framework by reconstructing masked frames from local temporal windows, while DeepInterpolation [16] uses a Noise2Noise-inspired training setup to interpolate entire frames from neighboring ones. DeepSeMi [17] introduces an asymmetric prediction strategy using transformed patches, and DeepVID [18] combines temporal and spatial masking to exploit redundancy across both domains. SUPPORT [3] further enhances spatiotemporal modeling by predicting pixel values using full contextual information across time and space. These methods have demonstrated substantial improvements in signal-to-noise ratio and recovery of fine-scale neuronal signals, yet challenges remain in balancing denoising performance with computational efficiency and ensuring generalization across diverse imaging conditions.

An important and often overlooked consideration in denoising voltage imaging data is the trade-off between spatial and temporal performance. High spatial performance refers to a method's ability to accurately preserve fine anatomical details and sharp structural boundaries. High temporal performance implies maintaining the fidelity of fast-changing signals such as action potentials or subthreshold voltage fluctuations over time [3]. Denoising methods that aggressively suppress noise may improve visual clarity and enhance spatial signal-to-noise ratio (e.g., through spatial smoothing or filtering), but this often comes at the cost of blurring or distorting rapid temporal events. Conversely, methods optimized for temporal fidelity may retain noisy fluctuations to avoid corrupting true dynamic signals, potentially preserving timing accuracy but introducing spatial artifacts or noise residuals [19]. Achieving a balance between these two aspects is particularly challenging in voltage imaging, where both fine spatial structures and millisecond-scale voltage dynamics are critical for accurate interpretation [20]. Different denoising approaches tend to prioritize one domain over the other, either explicitly through their architecture or implicitly through their loss functions and training objectives [3, 4, 6, 18]. As such, understanding this trade-off is crucial when selecting or designing denoising algorithms for specific neuroscience applications, where the relative importance of spatial precision versus temporal accuracy may vary depending on the experimental goal.

Given the critical role of denoising in recovering meaningful signals from noisy voltage imaging data, there is a practical need to systematically evaluate the performance of established denoising methods under realistic conditions. While traditional matrix factorization techniques such as PMD [5] offer interpretable and computationally efficient solutions, they may struggle with preserving fast, transient features characteristic of voltage signals. In contrast, recent advances in self-supervised deep learning, exemplified by methods such as DeepCAD-RT [4] and SUPPORT [3], promise to better capture the nonlinear and dynamic properties of neuronal activity without requiring clean training targets. These methods represent two major classes of denoising approaches, model-based (PMD) and learning-based (DeepCAD-RT, SUPPORT), each with distinct assumptions and operational regimes. By analyzing and comparing these methods on a common voltage imaging dataset, this study aims to clarify their relative strengths and limitations.

3 Overview of Selected Denoising Methods

To evaluate the strengths and weaknesses of current voltage imaging denoising techniques, we focus on three representative methods: SUPPORT [3], DeepCAD-RT [4], and PMD [5]. These methods were chosen because they cover both deep learning-based and traditional approaches, and they represent some of the most widely used or promising tools for denoising fast, noisy imaging data. SUPPORT [3] has already become a benchmark in the field: it outperforms other classic frame-based denoisers, and has seen rapid adoption with public code and citations [3]. DeepCAD-RT [4] is widely used in high-speed fluorescence microscopy, especially calcium imaging, because of its real-time denoising capability, cross-platform software support, and demonstrated improvements in practical experiments [4]. Finally, PMD [5] remains a strong traditional baseline; its low-rank factorization has stood the test of time in multiple imaging domains due to its robustness and speed, making it a common comparison point in literature [3, 6]. In contrast, newer or less specialized alternatives often fall short in voltage imaging. Methods like DeepInterpolation [16], DeepVID [18], or basic wavelet-based filters [2] have been consistently outperformed by the three selected approaches [3, 6]. These alternatives are either not tailored for the rapid dynamics of voltage signals or make strong assumptions about noise, which limits their effectiveness.

PMD [5] is a model-driven technique that assumes the observed data can be approximated as a combination of a few dominant spatial and temporal patterns. It decomposes the data into a low-rank matrix that captures structured signals and a sparse residual that accounts for noise or outliers. PMD is appealing due to its simplicity, speed, and interpretability. However, it operates under linear assumptions and may miss subtle or complex features-especially rapid transients like action potentials, that do not align well with the low-rank structure.

In contrast, DeepCAD-RT [4] is a self-supervised deep learning method that leverages convolutional neural networks to denoise calcium or voltage imaging data. It works by reconstructing masked frames from neighboring frames in small temporal windows. This enables the model to learn both spatial and temporal patterns without requiring clean ground truth data. DeepCAD-RT is especially effective at enhancing the signal-to-noise ratio and preserving smooth activity patterns but may struggle with extremely sparse or highly localized events if temporal context is too limited.

SUPPORT [3] represents a more recent and advanced approach. It also follows a self-supervised learning paradigm but makes full use of the spatiotemporal context to predict each pixel. By integrating information from both nearby pixels and nearby frames, SUPPORT achieves high-fidelity denoising, even in data with low signal-to-noise ratio or fast, complex dynamics. Its design is well-suited to the challenges of voltage imaging, where both spatial and temporal continuity are essential to recover meaningful signals.

These three methods offer fundamentally different perspectives on how to approach denoising: PMD is deterministic and model-based, DeepCAD-RT uses local learning from deep neural networks, and SUPPORT integrates global spatiotemporal learning. Comparing them side-by-side is valuable for understanding how different design choices affect denoising performance, particularly in the context of preserving fast voltage dynamics. This comparison can help researchers choose the appropriate method for their specific needs and inform the development of more effective hybrid or next-generation algorithms.

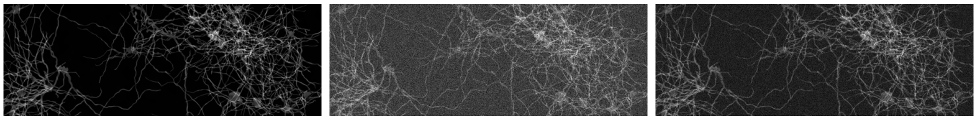
4 Experimental Setup

4.1 Datasets

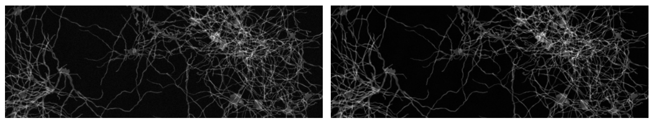
To evaluate and compare the denoising performance of the mentioned methods, we selected both synthetic and real voltage imaging datasets. This dual approach allows us to assess each method under controlled, ground-truth-aware conditions as well as in realistic, biologically complex settings.

4.1.1 Synthetic dataset

For the synthetic dataset, we use data generated by Optosynth [21], a simulation framework that produces realistic voltage imaging movies with access to ground truth signals. Optosynth combines single-neuron morphologies and electrophysiological recordings from the Allen Brain Atlas to simulate fluorescence responses to action potentials. For each frame, it computes propagation delays across neuron structures, maps voltage to fluorescence using reporter-specific dynamics, and adds background fluorescence. Optical blur is applied via a point spread function, followed by the addition of pixel-wise Poisson-Gaussian noise to simulate realistic imaging conditions.



(a) Example frame without noise (ground truth) (b) Example frame with SNR level 1 (c) Example frame with SNR level 2



(d) Example frame with SNR level 3 (e) Example frame with SNR level 4

Figure 2: Example frames of Optosynth dataset [21]

In our experiments, we use a composite dataset consisting of five distinct synthetic datasets, each capturing different neuron morphologies and activity patterns. Each dataset includes four different SNR levels to represent a broad spectrum of imaging quality. We refer to them as SNR levels 1 to 4, with level 1 corresponding to the lowest SNR and level 4 to the highest. Example frames are shown in figure 2 Every dataset variant contains 7,000 frames, resulting in a large and diverse benchmark that challenges denoising methods under varying noise intensities and structural complexity.

This structured setup provides several key advantages. First, the inclusion of multiple SNR levels allows for evaluating denoising models across different noise regimes, which is critical for generalizability. Second, access to ground truth voltage traces for every pixel enables precise quantitative evaluation of denoising performance, which is impossible in real experimental data. Finally, the diversity across datasets ensures that models are exposed to a wide variety of spatial and temporal dynamics, helping prevent overfitting to specific neuron morphologies or firing patterns.

Due to limitations in time and computational resources, we used pre-generated data released by the authors of Optosynth, which still retains the framework's high fidelity and variability, making it ideal for benchmarking purposes.

4.1.2 Real dataset

For the real dataset, we use recordings made publicly available by a previous study [22]. These recordings capture population-level neural activity in vivo, acquired using GEVIs and high-speed optical imaging. The full dataset comprises 22 recordings, each containing on average 10,000 frames, providing a rich and diverse collection of real-world voltage imaging data. However, due to computational constraints, we limit our analysis to a representative subset of 3 recordings selected to span different imaging conditions and neural dynamics.

Real datasets introduce unique challenges not present in synthetic data, such as biological variability, motion artifacts, and complex, unknown noise distributions. These factors make them essential for evaluating how well denoising methods generalize beyond controlled synthetic settings to practical, real-world applications. Using even a subset of these recordings helps assess the robustness and applicability of our models under realistic experimental conditions.

By including both synthetic and real data in our experiments, we aim to provide a comprehensive comparison: the synthetic data offers precise benchmarks for signal recovery accuracy, while the real data tests each method's effectiveness in preserving meaningful neural dynamics under natural imaging conditions. This balanced evaluation strategy helps ensure that our findings are both scientifically grounded and applicable to real neuroscience workflows.

4.2 Experimental Conditions

To ensure a consistent and fair comparison between the three denoising method, we designed the experimental setup around a controlled data split.

We employed a 20:80 data split, using the first 20% of the frames as the test set and the remaining 80% as the training set. This choice was driven by the nature of SUPPORT, which can be benefit from access to temporal context [3]. By placing the test set at the beginning of the time series, we ensure that no future frames leak into the model training, which would otherwise lead to inflated performance metrics due to temporal overlap [23].

PMD, being a traditional matrix factorization-based method, does not require training and can be applied directly on the test set. While PMD operates independently of training data, it is evaluated using the same test segment (first 20% of frames) for consistency.

4.3 Implementation Details

To maintain a consistent and reproducible evaluation pipeline, no hyperparameter tuning was performed for any of the methods. Instead, we adopted the default training hyperparameters recommended by the original authors of SUPPORT and DeepCAD-RT. For specific settings such as learning rates, loss functions, and optimizer choices, we refer readers to the respective publications [3, 4].

Both SUPPORT and DeepCAD-RT were trained for 20 epochs, which we found sufficient for stable convergence under the prescribed settings.

All experiments were conducted with an Intel Core i7-13650HX CPU and an NVIDIA GeForce RTX 4060 GPU.

4.4 Evaluation Metric

To quantitatively assess the performance of denoising methods on voltage imaging data, we employ three complementary metrics: Peak Signal-to-Noise Ratio (PSNR), the Structural Similarity Index Measure (SSIM) [24], temporal Signal-to-Noise Ratio (tSNR). These metrics were selected to capture both spatial and temporal aspects of denoising quality, which are critical in the context of voltage imaging, where preserving fine cellular structures and fast transient dynamics are equally important. PSNR is applied exclusively to the Optosynth dataset, where a clean ground-truth signal is available, enabling direct evaluation of reconstruction accuracy. SSIM is used only on the real dataset, where no ground truth exists, to assess the preservation of structural content relative to the raw input. tSNR is computed for both datasets to evaluate temporal noise suppression across methods.

4.4.1 Peak Signal-to-Noise Ratio (PSNR)

PSNR is a widely used metric in image processing that quantifies the fidelity of a denoised image relative to a reference by comparing the mean squared error [3, 4, 6]. It is expressed in decibels (dB), with higher values indicating closer agreement between the denoised and reference images. It is used in datasets where a clean ground truth is available (in the Optosynth dataset), where it serves as a direct measure of reconstruction quality. For a clean signal X and a noisy signal Y, it is defined as:

$$\mathrm{PSNR}\left(X,Y\right) = 10\log_{10}\left(\frac{I_{\max}}{\frac{1}{WH}\sum_{x=1}^{W}\sum_{y=1}^{H}\left[X(x,y)-Y(x,y)\right]^2}\right) \tag{1}$$

where W is the width of the signal, H is the height of the signal, and I_{max} is the maximum possible value of signal intensity.

4.4.2 Structual Similarity Index Measure (SSIM)

SSIM [24] evaluates image similarity based on luminance, contrast, and structural information. It ranges from 0 to 1, with values closer to 1 indicating higher structural similarity. In the absence of ground-truth images (as in real datasets), SSIM can be computed relative to the raw input to estimate how much structural information is preserved during denoising. This helps identify methods that are overly aggressive and may smooth out important spatial details.

4.4.3 Temporal Signal-to-Noise Ratio (tSNR)

Inspired by a previous work [19], tSNR is used to measure the temporal performance. It measures the stability of pixel-wise signals over time by computing the ratio of the temporal mean to the temporal standard deviation. Formally,

$$\text{Temporal SNR}(x) = \frac{\mu}{\sigma} = \frac{\frac{1}{T}\sum_{t=1}^{T}x_t}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(x_t-\mu)^2}} \tag{2}$$

where μ is the temporal mean and T is the time series. It quantifies how effectively a denoising method reduces temporal fluctuations while retaining meaningful signal variations. In the context of voltage imaging, where rapid changes in membrane potential are captured over time, a higher tSNR indicates better temporal clarity and reduced noise in the recordings.

4.4.4 Computational Efficiency

In addition to denoising performance, we assessed the computational efficiency of each method, as this is a critical factor for real-world applications, especially in high-throughput or real-time imaging scenarios. For each method, we measured the runtime per sample (one frame of 128×512 from the Optosynth dataset [21]) needed to denoise the test set, as well as the runtime per patch and peak GPU memory usage during training for the model-based approaches (SUPPORT and DeepCAD-RT). Since both models are applied patch-wise during training and inference, the overall computational cost scales approximately linearly with the number of patches, which in turn scales with video resolution and length. To ensure a fair comparison, the patch size was fixed to $150 \times 100 \times 100 (t \times x \times y)$ for both training and inference, following recommendations from previous work [3, 4]. The overlap factor was set to 0.25 while training and 0.6 while inference, as recommended in [4]. These measurements provide insight into the practical deployment costs of each approach. Including computational efficiency in the evaluation allows us to weigh the trade-offs between denoising performance and resource consumption, thereby providing a more holistic comparison across methods.

5 Results

5.1 Results on Optosynth dataset

Table 1: PSNR results in dB on Optosynth data for different denoising methods across different SNR levels

SNR Level	Method	Mode PSNR Gain(dB)	Mean PSNR Gain(dB)
1	SUPPORT	23.87	24.68
	DeepCAD-RT	18.19	18.13
	PMD	21.74	21.3
2	SUPPORT	23.93	24.34
	DeepCAD-RT	16.42	16.35
	PMD	16.71	16.33
3	SUPPORT	23.72	23.24
	DeepCAD-RT	14.22	14.02
	PMD	15.15	14.78
4	SUPPORT	23.57	22.06
	DeepCAD-RT	12.77	12.35
	PMD	20.46	19.86

Table 1 summarizes the PSNR gain in decibels for three denoising methods, SUPPORT [3], DeepCAD-RT [4], and PMD [5], across varying SNR levels on the Optosynth dataset. SUPPORT consistently outperformed the other methods, achieving the highest mode and mean PSNR gains across all noise levels. Its performance remained remarkably stable, with mean PSNR gains ranging from 22.06 dB to 24.68 dB. In contrast, DeepCAD-RT showed a noticeable decline in PSNR gain as the SNR level increased, with its mean gain decreasing from 18.13 dB at SNR level 1 to 12.35 dB at SNR level 4. PMD demonstrated intermediate performance, with more fluctuation across SNR levels; notably, it achieved a relatively high mean gain of 19.86 dB at SNR level 4, surpassing DeepCAD-RT at higher noise levels. These results suggest that SUPPORT is robust to varying noise conditions, while PMD may be preferable to DeepCAD-RT when denoising data with moderate to high SNR.

Table 2 presents the mean tSNR gain in decibels achieved by three denoising methods across varying SNR levels on the Optosynth dataset. PMD consistently delivered the highest tSNR gain across all noise conditions, with gains ranging from 7.31 dB at SNR level 4 to 16.62 dB at SNR level 1. SUPPORT showed strong performance at low SNRs, particularly at SNR level 1 and 2, but its tSNR gain declined more sharply at higher SNR levels. DeepCAD-RT exhibited a similar decreasing trend, with its tSNR gain dropping from 12.54 dB at SNR level 2 to 6.05 dB at SNR level 4. Overall, the results indicate that while all methods improve temporal signal quality, PMD is most effective across all SNR levels, particularly at the lower end, whereas the benefits of SUPPORT and DeepCAD-RT diminish more noticeably

Table 2: tSNR results in dB on Optosynth data for different denoising methods across different SNR levels

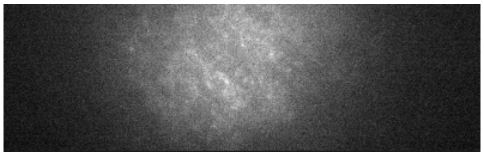
SNR Level	Method	Mean tSNR Gain (dB)
1	SUPPORT	12.66
	DeepCAD-RT	11.24
	PMD	16.62
2	SUPPORT	12.28
	DeepCAD-RT	12.54
	PMD	15.87
3	SUPPORT	10.08
	DeepCAD-RT	8.69
	PMD	11.28
4	SUPPORT	7.01
	DeepCAD-RT	6.05
	PMD	7.31

as the underlying data becomes less noisy.

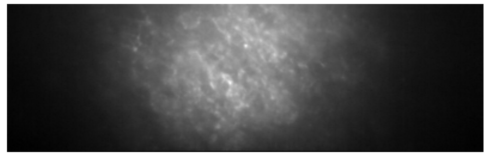
Considering both PSNR and tSNR results (Tables 1 and 2), a clear trade-off emerges between spatial and temporal denoising performance among the three evaluated methods. SUPPORT consistently delivers the highest spatial denoising quality, with the largest PSNR gains across all SNR levels, maintaining strong performance even as noise decreases. However, its tSNR gains are more moderate and decline significantly at higher SNR levels, indicating reduced effectiveness in preserving temporal dynamics in cleaner data. In contrast, PMD excels in temporal performance, achieving the highest tSNR gains across all SNR levels, particularly under strong noise conditions. Although its spatial performance is generally lower than SUPPORT's, it remains competitive, especially at higher SNR levels where it even outperforms DeepCAD-RT. DeepCAD-RT shows the weakest performance overall, with both PSNR and tSNR gains decreasing as SNR improves, suggesting it is less robust under varying noise conditions. These findings highlight a fundamental trade-off: SUPPORT prioritizes spatial fidelity, PMD favors temporal consistency, and DeepCAD-RT underperforms in both dimensions under these conditions. Depending on the application, the choice of denoising method should balance this trade-off to align with the primary objective, whether accurate spatial structure or stable temporal signal is more critical.

5.2 Results on real dataset

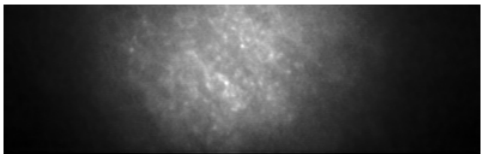
Figure 3 presents a visual comparison of denoising results from the three methods on the same example frame, alongside the corresponding raw signal. The raw signal appears noisy, with significant background fluctuations obscuring the underlying structures. All three denoising methods effectively suppress the noise and enhance the visibility of neuronal features. SUPPORT and PMD produce smoothened outputs with improved clarity over the raw signal, while DeepCAD-RT stands out by revealing slightly sharper and more distinct neuronal structures. This suggests that DeepCAD-RT not only reduces noise but also preserves fine details more effectively, offering improved visual fidelity in the denoised output.



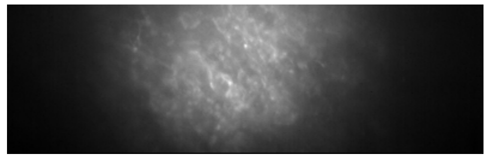
(a) Example frame of raw signal



(b) Example frame of SUPPORT denoised signal



(c) Example frame of DeepCAD-RT denoised signal



(d) Example frame of PMD denoised signal

Figure 3: Example raw and denoised frames of real dataset

Table 3: tSNR and SSIM results comparison on the real dataset for different denoising methods

Method	**Mean tSNR Gain (dB)**	**Mean SSIM**
SUPPORT	4.37	0.915
DeepCAD-RT	4.36	0.922
PMD	4.42	0.918

Table 3 presents a comparison of denoising performance on the real dataset using mean tSNR and mean SSIM. All three denoising methods substantially improve the temporal SNR compared to the raw input, with PMD achieving the highest mean tSNR gain (4.42 dB), followed closely by SUPPORT (4.37 dB) and DeepCAD-RT (4.36 dB). While the differences in tSNR are minor, they suggest that all methods are similarly effective at suppressing temporal noise. Regarding structural preservation, DeepCAD-RT yields the highest mean SSIM (0.922), indicating it best preserves spatial structures relative to the noisy input. PMD and SUPPORT also perform well, with SSIM values of 0.918 and 0.915, respectively. Overall, PMD offers the best balance between temporal denoising and spatial fidelity, while DeepCAD-RT is slightly more conservative in denoising but excels at retaining structural information. SUPPORT, while still effective, appears to be marginally more aggressive, potentially sacrificing some spatial detail for denoising strength.

5.3 Computational Efficiency

Table 4: Training peak memory used and speed for SUPPORT and DeepCAD-RT

Method	**Peak Memory (GB)**	**Training Speed (patch/s)**
Support	5.5	1.21
DeepCAD-RT	3.7	3.12

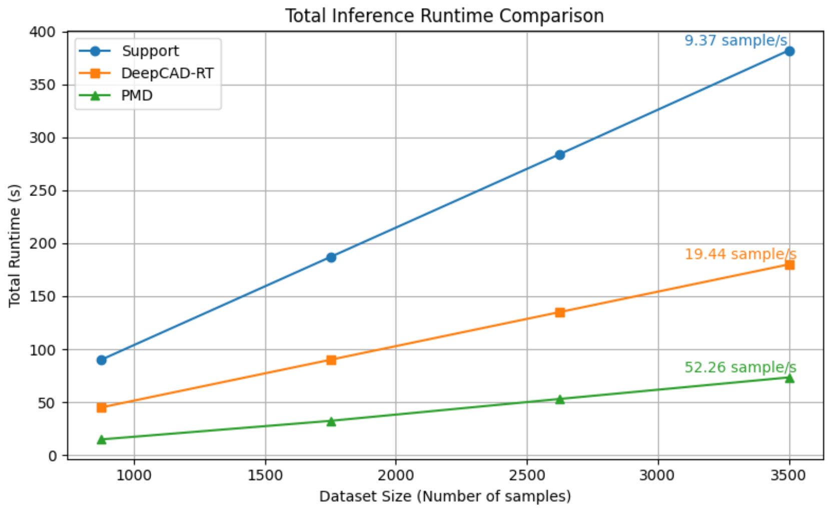


Figure 4: Inference runtime per sample (s) for different denoising methods across dataset sizes

Table 4 compares the training performance of SUPPORT and DeepCAD-RT. DeepCAD-RT exhibits notably better efficiency, with a lower peak memory usage of 3.7 GB and a faster training speed of 3.12 patches per second, compared to SUPPORT's 5.5 GB and 1.21 patches per second. This indicates that DeepCAD-RT is more resource-efficient and better suited for faster model development. Figure 4 shows the inference speeds of different denoising methods. DeepCAD-RT achieves an inference speed of 19.44 samples per second, outperforming SUPPORT, which runs at 9.36 samples per second. PMD records the highest inference speed at 52.26 samples per second; however, as a non-model-based method, it is not directly comparable in terms of training performance. Nonetheless, its fast inference makes it attractive for scenarios where speed is prioritized over model-based learning. Overall, DeepCAD-RT provides a strong balance between training efficiency and fast inference, making it a practical choice for real-time or large-scale denoising tasks.

6 Responsible Research

In conducting this study, we adhered to principles of transparency, reproducibility, and ethical integrity, as emphasized in the ACL Responsible Research Checklist [25]. All datasets used in this work are publicly available, and appropriate citations are provided to acknowledge the original sources. In particular, we clearly distinguish between synthetic and real datasets to avoid any ambiguity in evaluation scenarios, and we make no unjustified claims about generalizability beyond the tested data.

We have ensured that the methodological comparisons are fair and unbiased. The experimental setup was carefully designed so that each method was evaluated under equivalent conditions, with no selective tuning or omission of results. Hyperparameter settings followed defaults from the respective original publications, and performance metrics were computed on shared test sets to allow valid comparisons. Wherever subjective evaluation (such as visual inspection) was involved, it was supplemented by quantitative metrics to reduce bias.

We are concerned about reproducibility. The implementations of all denoising methods used in this study are publicly available through their respective official repositories. Detailed descriptions of model architectures, training procedures, and hyperparameters can be found in the original publications. To ensure consistency and facilitate replication of our results, we fixed all random seeds across experiments and documented the complete experimental setup. At the end of this section, we provide access to all relevant repositories and datasets needed to reproduce this study.

With regard to potential societal impacts, this research is focused on improving signal quality in voltage imaging, a domain with implications in neuroscience and biomedical research. We acknowledge that although denoising enhances interpretability, it may also risk introducing artifacts or over-smoothing, particularly in the absence of ground truth data. We encourage careful, domain-aware interpretation of results when applying these methods to real-world experimental data.

No personally identifiable information or human subjects data were used in this study. Our work poses minimal ethical risk and remains within the bounds of responsible computational research.

Code and data availability

Python implementation of SUPPORT: `https://github.com/NICALab/SUPPORT/` [3]. Python implementation of DeepCAD-RT: `https://github.com/cabooster/DeepCAD-RT/` [4]. Python implementation of PMD: `https://github.com/paninski-lab/funimag/` [5]. The Optosynth dataset can be accessed from the Google Cloud bucket found at gs://broad-dsp-cellmincer-data [6]. The real dataset can be accessed at `https://zenodo.org/records/10020273` [22], specifically *00_ 02.tif, 00_ 03.tif, 01_ 01.tif* in voltage_HPC2.zip/HPC2.

7 Discussion

7.1 Comparative Analysis

The results presented provide a comprehensive comparison of three prominent denoising methods, SUPPORT [3], DeepCAD-RT [4], and PMD [5], evaluated on both synthetic and real voltage imaging datasets. Each method offers unique strengths and limitations, and their performance varies depending on the metric and dataset context.

SUPPORT demonstrates superior spatial denoising performance, achieving the highest PSNR gains across all SNR levels in the Optosynth dataset, and competitive tSNR and SSIM on real data. This suggests that SUPPORT is highly effective in enhancing image clarity and reducing pixel-level noise without significantly compromising temporal fidelity, particularly at lower SNRs. This aligns with the conclusion of previous research [3, 6, 18]. However, its declining tSNR performance in higher SNR conditions, along with its relatively high computational demands, indicates a trade-off that may limit its scalability or applicability in real-time imaging workflows.

PMD, by contrast, excels in temporal denoising, as evidenced by the highest tSNR gains across all noise regimes. This makes it an appealing option for analyses that rely on the stability and integrity of time-series signals. While its PSNR is lower than that of SUPPORT, its competitive SSIM [24] and rapid inference speed highlight its utility in applications where quick processing is essential. Nonetheless, as a non-trainable, factorization-based method, PMD lacks adaptability and may struggle with datasets that deviate from its implicit assumptions about noise or signal structure.

DeepCAD-RT shows relatively lowest performance in both PSNR and tSNR on synthetic data, particularly as SNR improves. However, it achieves the highest SSIM on real data, indicating that it preserves structural integrity effectively, an important consideration for qualitative analyses or visual inspection tasks. Additionally, its lower memory footprint and faster training and inference times make it a practical choice in scenarios where computational efficiency is critical. While it underperforms in raw denoising metrics, its architectural simplicity and speed offer significant advantages, especially if further optimized.

7.2 Limitations and Future Works

One key limitation highlighted in our study is that DeepCAD-RT was used with default parameters and without task-specific optimization. While previous research has shown that DeepCAD-RT underperforms compared to SUPPORT and PMD, the performance gap may be narrower with proper fine-tuning [3, 4, 6]. Given its relatively lightweight design and potential for fine-tuning, future work could explore whether targeted training strategies, architectural enhancements, or loss function modifications might boost its performance, particularly for temporal signal preservation. For instance, incorporating temporal regularization or hybrid losses that jointly optimize for both spatial and temporal fidelity could potentially improve DeepCAD-RT's applicability in voltage imaging.

Another important consideration is the generalizability of the evaluated methods. While the Optosynth dataset [21] offers high realism and diversity, and real datasets provide biological genuineness, the scope of the evaluation remains limited to specific types of neuronal structures, imaging conditions, and signal dynamics. Expanding future evaluations to include more diverse datasets, such as recordings from different brain regions, imaging modalities (e.g., two-photon vs. widefield), or different types of voltage indicators, would provide a more robust assessment of model performance under varied biological and technical conditions.

To enhance future comparative evaluations, studies can consider incorporating labeled datasets with annotated regions of interest (ROIs). ROI-based ground truth enables more precise and biologically meaningful benchmarking of denoising methods, particularly within regions where neural signals are expected to occur. This approach allows for assessing how effectively each method preserves signal fidelity within relevant structures, rather than relying solely on global metrics such as PSNR or SSIM. Such targeted evaluations offer a more functionally relevant basis for comparison, better reflecting the practical utility of denoising in neuroscience applications.

Moreover, while PSNR, tSNR, and SSIM [24] are valuable for quantifying denoising performance, they may not fully capture the biological relevance of the output. In practical neuroscience applications, the end goal of denoising is often not just visual clarity, but accurate interpretation of neural signals, such as spike detection, population dynamics, or connectivity inference. Future studies should incorporate task-specific downstream metrics such as event detection accuracy, spike timing precision, or classification accuracy in behaviorally relevant tasks. These functional metrics would provide a more application-driven perspective on denoising quality and reveal whether improvements in conventional image metrics translate to enhanced scientific insight.

7.3 Summary

In summary, our comparative analysis underscores the multifactorial nature of denoising in voltage imaging. Each method represents a distinct point in the trade-off space between spatial accuracy, temporal fidelity, structural preservation, and computational efficiency. Although SUPPORT remains the most robust in terms of spatial denoising, PMD offers unmatched speed and temporal performance, and DeepCAD-RT offers limited benefits overall, but remains notable for structural preservation and computational efficiency.

8 Conclusions and Future Work

This study conducted a comparative analysis of three modern denoising methods, SUPPORT, DeepCAD-RT, and PMD, for voltage imaging data. Through evaluations on both synthetic and real datasets, we examined each method's strengths in terms of spatial fidelity, temporal stability, and computational efficiency. The results highlight that no single method is universally best. Instead, each offers specific advantages depending on the denoising objective. SUPPORT emphasizes spatial clarity, PMD excels in temporal consistency and speed, and DeepCAD-RT shows a balance with efficient, structure-preserving denoising.

As discussed, the choice of method should be guided by the specific priorities of the imaging task, whether that is preserving fine structural detail, stabilizing time-series signals, or ensuring fast and resource-efficient processing. This reinforces the importance of evaluating denoising tools not only by traditional metrics but also by how well they serve practical goals in real neuroscience workflows.

Looking forward, future work can build on the limitations and opportunities identified in the discussion. In particular, refining evaluation metrics to include task-specific or biologically grounded criteria, exploring ROI-based benchmarking, and expanding the scope of datasets would offer more functionally relevant insights. Moreover, further development of existing models, especially lightweight architectures like DeepCAD-RT, could enhance performance through targeted optimization. These directions will help bridge the gap between algorithmic improvements and their real-world scientific utility.

References

[1] St-Pierre F Yang HH. Genetically encoded voltage indicators: Opportunities and challenges. *J Neurosci*, 2016.

[2] Alle Meije Wink and Jos B.T.M. Roerdink. *Enhancing Functional Neuroimages: Wavelet Denoising as an Alternative to Gaussian Smoothing.* University of Groningen, Johann Bernoulli Institute for Mathematics and Computer Science, 2002. Relation: http://www.rug.nl/informatica/organisatie/overorganisatie/iwi Rights: University of Groningen. Research Institute for Mathematics and Computing Science (IWI).

[3] Minho Eom, Seungjae Han, Pojeong Park, Gyuri Kim, Eun-Seo Cho, Jueun Sim, Kang-Han Lee, Seonghoon Kim, He Tian, Urs L. Böhm, Eric Lowet, Hua an Tseng, Jieun Choi, Stephani Edwina Lucia, Seung Hyun Ryu, Márton Rózsa, Sunghoe Chang, Pilhan Kim, Xue Han, Kiryl D. Piatkevich, Myunghwan Choi, Cheol-Hee Kim, Adam E. Cohen, Jae-Byum Chang, and Young-Gyu Yoon. Statistically unbiased prediction enables accurate denoising of voltage imaging data. *Nature Methods*, 20, 2023.

[4] Xinyang Li, Yixin Li, Yiliang Zhou, Jiamin Wu, Zhifeng Zhao, Jiaqi Fan, Fei Deng, Zhaofa Wu, Guihua Xiao, Jing He, Yuanlong Zhang, Guoxun Zhang, Xiaowan Hu, Xingye Chen, Yi Zhang, Hui Qiao, Hao Xie, Yulong Li, Haoqian Wang, Lu Fang, and Qionghai Dai. Real-time denoising enables high-sensitivity fluorescence time-lapse imaging beyond the shot-noise limit. *Nature Biotechnology*, 41, 2022.

[5] E. Kelly Buchanan, Ian Kinsella, Ding Zhou, Rong Zhu, Pengcheng Zhou, Felipe Gerhard, John Ferrante, Ying Ma, Sharon Kim, Mohammed Shaik, Yajie Liang, Rongwen Lu, Jacob Reimer, Paul Fahey, Taliah Muhammad, Graham Dempsey, Elizabeth Hillman, Na Ji, Andreas Tolias, and Liam Paninski. Penalized matrix decomposition for denoising, compression, and improved demixing of functional imaging data. *bioRxiv*, 2019.

[6] Brice Wang, Tianle Ma, Theresa Chen, Trinh Nguyen, Ethan Crouse, Stephen Fleming, Alison S. Walker, Vera Valakh, Ralda Nehme, Evan W. Miller, Samouil L. Farhi, and Mehrtash Babadi. Robust self-supervised denoising of voltage imaging data using cellmincer. *bioRxiv*, 2024.

[7] Yoav Adam, Jeong J Kim, Shan Lou, Yongxin Zhao, Michael E Xie, Daan Brinks, Hao Wu, Mohammed A Mostajo-Radji, Simon Kheifets, Vicente Parot, et al. Voltage imaging and optogenetics reveal behaviour-dependent changes in hippocampal dynamics. *Nature*, 569(7756):413–417, 2019.

[8] Ian Jolliffe and Jorge Cadima. Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374:20150202, 04 2016.

[9] Toru Aonishi, Ryoichi Maruyama, Tsubasa Ito, Hiroyoshi Miyakawa, Masanori Murayama, and Keisuke Ota. Imaging data analysis using non-negative matrix factorization. *Neuroscience Research*, 179:51–56, 2022. Lighting up cosmic neuronal networks with transformative in vivo calcium imaging.

[10] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7):3142–3155, 2017.

[11] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2Noise: Learning image restoration without clean data. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 2965–2974. PMLR, 10–15 Jul 2018.

[12] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2void - learning denoising from single noisy images, 2019.

[13] Joshua Batson and Loic Royer. Noise2Self: Blind denoising by self-supervision. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 524–533. PMLR, 09–15 Jun 2019.

[14] Yide Zhang, Yinhao Zhu, Evan Nichols, Qingfei Wang, Siyuan Zhang, Cody Smith, and Scott Howard. A poisson-gaussian denoising dataset with real fluorescence microscopy

images. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11702–11710, 2019.

[15] Xinyang Li, Guoxun Zhang, Jiamin Wu, Yuanlong Zhang, Zhifeng Zhao, Xing Lin, Hui Qiao, Hao Xie, Haoqian Wang, Lu Fang, et al. Reinforcing neuron extraction and spike inference in calcium imaging using deep self-supervised denoising. *Nature methods*, 18(11):1395–1400, 2021.

[16] Jérôme Lecoq, Michael Oliver, Joshua H Siegle, Natalia Orlova, Peter Ledochowitsch, and Christof Koch. Removing independent noise in systems neuroscience data using deepinterpolation. *Nature methods*, 18(11):1401–1408, 2021.

[17] Guoxun Zhang, Xiaopeng Li, Yuanlong Zhang, Xiaofei Han, Xinyang Li, Jinqiang Yu, Boqi Liu, Jiamin Wu, Li Yu, and Qionghai Dai. Bio-friendly long-term subcellular dynamic recording by self-supervised image enhancement microscopy. *Nature Methods*, 20(12):1957–1970, 2023.

[18] Jelena Platisa, Xin Ye, Allison M Ahrens, Chang Liu, Ichun Anderson Chen, Ian G Davison, Lei Tian, Vincent A Pieribone, and Jerry L Chen. High-speed low-light in vivo two-photon voltage imaging of large neuronal populations. *Nature methods*, 20(7):1095–1103, 2023.

[19] Chang Liu, Jiayu Lu, Yicun Wu, Xin Ye, Allison M. Ahrens, Jelena Platisa, Vincent A. Pieribone, Jerry L. Chen, and Lei Tian. Deepvid v2: Self-supervised denoising with decoupled spatiotemporal enhancement for low-photon voltage imaging. *bioRxiv*, 2024.

[20] Chang Liu, Jiayu Lu, Yicun Wu, Xin Ye, Allison M. Ahrens, Jelena Platisa, Vincent A. Pieribone, Jerry L. Chen, and Lei Tian. Deepvid v2: Self-supervised denoising with decoupled spatiotemporal enhancement for low-photon voltage imaging. *bioRxiv*, 2024.

[21] Mehrtash Babadi. Optosynth. `https://github.com/cellarium-ai/Optosynth`, 2021.

[22] Yuki Bando, Raghav Pillai, Akihiro Kajita, Faisal Abdul Hakeem, Yann Quemener, Hsien-An Tseng, Kiryl Piatkevich, Cheng Linghu, Xue Han, and Edward S. Boyden. Real-time neuron segmentation for voltage imaging, 2023.

[23] Ravi K. Samala, Heang-Ping Chan, Lubomir Hadjiiski, and Sathvik Koneru. Hazards of data leakage in machine learning: a study on classification of breast cancer using deep neural networks. In Horst K. Hah nand Maciej A. Mazurowski, editor, *Medical Imaging 2020: Computer-Aided Diagnosis*, volume 11314, page 1131416. International Society for Optics and Photonics, SPIE, 2020.

[24] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[25] Acl rolling review - a peer review platform for the association for computational linguistics. `https://aclrollingreview.org/responsibleNLPresearch/`. Accessed: 2025-06-10.