CellMincer: Self-Supervised Denoising of Functional Imaging

by

Brice Wang

Bachelor of Science in Computer Science and Engineering,
Massachusetts Institute of Technology (2020)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2022

Author .....
Department of Electrical Engineering and Computer Science
May 17, 2022

Certified by .....
Mehrtash Babadi
Associate Director of Machine Learning
Thesis Supervisor

Certified by .....
Caroline Uhler
Associate Professor
Thesis Supervisor

Accepted by .....
Katrina LaCurts
Chair, Master of Engineering Thesis Committee

CellMincer: Self-Supervised Denoising of Functional Imaging

by

Brice Wang

Submitted to the Department of Electrical Engineering and Computer Science
on May 17, 2022, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

All-optical electrophysiology offers accessibility and scalability in observing neuronal activity beyond what can feasibly be achieved with patch clamp techniques. However, imaging platforms like Optopatch suffer from excessive detection noise, photobleaching, and an inability to organically segment and isolate neurons of interest. These drawbacks preclude its use as a true substitute for direct electrophysiological measurement, but recent advances in deep neural network inference may enable computation to recover the difference in data quality. To date, few robust denoising algorithms have been designed and implemented for voltage imaging data, in part because the lack of ground truth imaging complicates the task of training such a model. This thesis introduces CellMincer, a self-supervised deep neural network for denoising functional imaging. By exploiting a combination of spatiotemporally local contexts and precomputed global features, CellMincer outperforms comparable algorithms at denoising several modes of optical electrophysiology on a range of metrics, including measures of biologically relevant features.

Thesis Supervisor: Mehrtash Babadi
Title: Associate Director of Machine Learning

Thesis Supervisor: Caroline Uhler
Title: Associate Professor

Acknowledgments

First, I would like to thank my research supervisor Mehrtash Babadi. Despite my joining our first Zoom conversation with little more than enthusiasm and a passing knowledge of deep learning and biology, he welcomed me into his lab group and connected me with a fulfilling research experience. In particular, his contributions in creating the original CellMincer codebase as well as Optosynth were critical in setting this project in motion. Over the fifteen months working with him, he has been unfailingly patient and overwhelmingly generous with his time and energy. My understanding of deep learning, biology, and the research process have all grown considerably under his instruction.

I would also like to thank my MEng thesis advisor Caroline Uhler. She graciously agreed to co-advise my project with Mehrtash and was incredibly helpful in connecting me with the resources I needed from the MIT EECS Department. My final thesis could not have been possible without her aid.

I would like to thank my academic advisor Justin Solomon. My MEng journey was not without logistical hurdles, and he led the great task of coordinating multiple fronts to secure the research assistantship funding I needed. He was very encouraging throughout my undergraduate and graduate experience and helped me thrive academically.

I would like to thank my labmates Luca D'Alessio, Stephen Fleming, and Nick Barkas. As a new member to their lab, they made me feel welcome in both an academic and personal sense. Over the course of my project, they contributed innumerable insights and ideas to my work, and they took special care in making their own work discussions accessible to me by providing ample context and answering my questions.

I would like to thank Sami Farhi and his lab group in the Optical Profiling Platform for generously providing me with Optopatch datasets, offering avenues to explore the biological implications for optical electrophysiology denoising, and using their efforts to discover their own applications of this result.

I would like to thank Evan Miller for graciously sharing his curated and annotated paired optical-ephys datasets with me, without which a key pillar of this thesis would not be possible.

I would like to thank my friends in the MIT community as well as my old friends from the IMSA community for keeping me sane during a prolonged quarantine.

I would like to thank my sister Shuyu for introducing me to Mehrtash and his work, for following my research developments with enthusiasm, for advising me on my future, and for being unconditionally supportive in my endeavors. She is the reason I cultivated a passion for STEM research and has helped me the most to get to where I am today.

Finally, I would like to thank my parents Zhenyong and Xiaozhu. They have helped me discover my passions, given me the space to make my mistakes, and directed me back on track without judgement. Their encouragement has helped me dare to tackle challenges and dream big, and for that they have my gratitude.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Measuring neuronal activity

Neurons are widely known to carry electrical signals through the body and are intricately involved in coordinating brain functions, bodily processes, and stimulus relays, among many other activities necessary for survival. To prime itself to release such a signal, a neuron establishes an ion gradient across its membrane using sodium-potassium pumps. By asymmetrically displacing NA^+ and K^+ cations with active transport, the cell maintains a negative resting potential. A neuron uses voltage-gated ion channels to restore equilibrium, initiating a rapid membrane depolarizing event that cascades down the cell's axon. This event, referred to as an action potential, can contribute to the activating stimulus of a downstream neuron, continuing the signal transmission.

To observe neuronal activity, a patch clamp[12] physically interfaces with the cell membrane to measure transmembrane voltage (Figure 1-1). Widely accepted as the gold standard for collecting electrophysiological (ephys) data, patch clamps are notably laborious to handle and difficult to scale—a lab technician may only be able to patch clamp a handful of cells every hour. In cases where many neurons need to be analyzed in tandem, notably in certain *in vivo* settings, the patch clamp technique cannot feasibly be applied. In addition, a patch clamp can only measure one signal from one location, which can fail to capture the full spatial variance of neuronal

electrical activity[16]. Capturing these dynamics requires a higher probe density which can both be labor-prohibitive and disruptive to the tissue of interest[15].

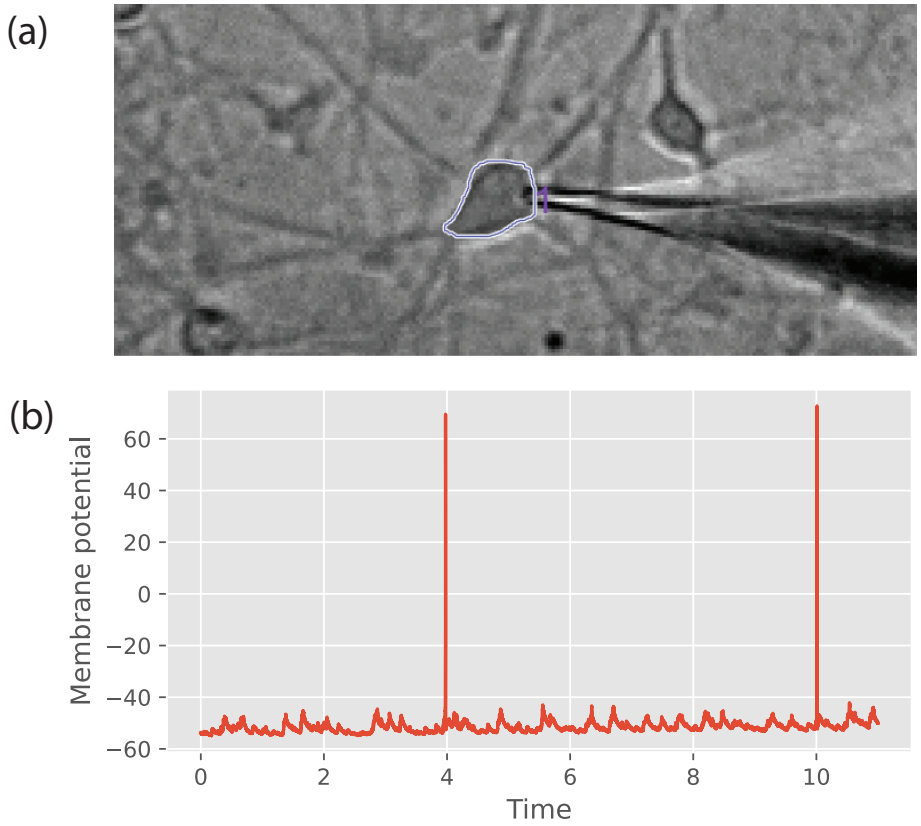

Figure 1-1: (a) Image of a patch clamp embedded in the membrane of a neuron. (b) Electrophysiological readout of the patch clamp during neuron stimulus.

Optopatch[9] is an all-optical electrophysiology platform which uses a system of membrane-bound proteins to image electroactive cells. To prepare the neuron, genetic information encoding the CheRiff channelrhodopsin actuator and the QuasAr archaerhodopsin voltage indicators is introduced, and the cell expresses them on its surface. A blue light stimulus activates CheRiff to initiate depolarization, while a red light stimulus triggers QuasAr fluorescence (Figure 1-2). These light emissions can be observed with a camera-like detector, such as a sCMOS, to produce a fluorescence recording[13]. With Optopatch, a plate of neurons can be prepared, perturbed, and observed in parallel.

A similar configuration for optical electrophysiology uses BeRST_1[14], a far-red fluorescent indicator which exhibits photostable characteristics and produces a linear

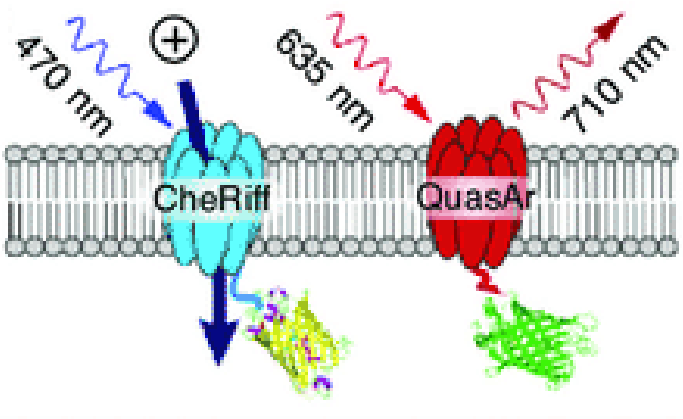

Figure 1-2: A diagram of CheRiff and QuasAr response to stimulation, reproduced from ref.[3]

response to voltage. Neurons with localized BeRST$_1$ fluorophores can be measured via patch clamp in tandem, producing a simultaneous optical and electrophysiological signal. These datasets feature prominently in analysis detailed in Chapter 5.

Compared to patch clamping, Optopatch-like imaging platforms offer easier, scalable access to measuring neuronal activity, and the inclusion of spatial resolution of voltage can reveal dynamics like signal propagation within and between neurons. However, the use of fluorescence as a proxy for membrane voltage introduces significant observation noise. At the scale of individual cells, the combination of shot noise from photon counting variance and Gaussian noise from background fluctuation produces a Poisson-Gaussian noise[18] of significant magnitude. Figure 1-3 depicts a normalized frame of such a recording.

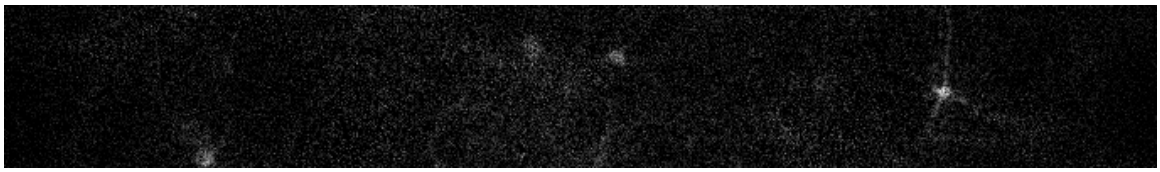

Figure 1-3: A normalized frame from a raw Optopatch dataset.

The signal-to-noise ratio can be improved by increasing the power of the reporter laser. While the cell would emit brighter fluorescence as a result, the amplified stimulus accelerates the degradation of the fluorophore molecules, impairing and ultimately inactivating the cell's voltage-reporting capability. This effect, referred to as photo-

bleaching, occurs at all reporter laser intensities and limits the recording duration when applying a meaningful level of stimulus to no more than 1-2 minutes. This falls short of relevant timescales in applications for behavioral analysis, a field which could potentially make use of high-resolution functional imaging.

Finally, Optopatch does not contain a mechanism for segmenting neurons imaged in a single FOV. Indeed, because it is both difficult to precisely position neurons on a plate and to assess the viability of the Optopatch modification in a particular neuron, even counting the number of neurons present is a challenge. The task of segmenting an Optopatch recording can only be done as a post-processing step and is not easily automated due to the inherent noise.

1.2 Proposing an Optopatch denoiser

1.2.1 Limitations of supervised learning

Denoising an Optopatch recording can potentially mitigate many of the issues associated with the process. By improving the signal-to-noise ratio, a neuron's underlying electrophysiological signal can more easily be recovered via simple intensity-averaging over a region of interest. Stimulating the fluorophore molecules at a lower power reduces the photobleaching effect, extending the time limit of an Optopatch recording, and the deficit in quality could be balanced out with denoising. Finally, because denoising such a dataset would strengthen the correlation between same-neuron pixels, it may be easier both to detect fluorescing neurons and to automatically segment an FOV with some form of components analysis.

Several factors complicate the task of producing such a denoiser. As an Optopatch dataset takes the form of a fluorescence movie, a natural approach would be to apply a deep denoising network for videos like ViDeNN[2]. In practice, Optopatch operates on an intensity scale far removed from that of a conventional video. As a demonstration of scale, the baseline intensity begins in the thousands and shifts drastically over time as a result of photobleaching, while a neuron action potential yields an increase of no

more than 1-2% of that value.

In addition, many frameworks for image denoising assume the inclusion of ground truth with its training data, which does not exist for voltage imaging. Attempting to predict the activity of a plate before imaging cannot be done reliably, and it is not feasible to control the imaging process in a way that produces a noise-free recording. Because of photobleaching, it is also not possible to repeatedly record the same sample under identical conditions so that the ground truth can be approximated by averaging over a large sample size.

With the possibility of generating ground truth directly using Optopatch excluded, other avenues for generating clean datasets should be considered. While elements of the Optopatch recording process can be simulated *in silico* to produce synthetic data, the diversity of reporter dye characteristics, stimulation protocols, and recording quality invalidate the use of models trained on synthetic data in the general setting. As a consequence, an ideal denoiser would need to be trained on data resembling its target distribution.

Finally, collecting a paired electrophysiological signal with patch clamp to use as ground truth would be prohibitively time-consuming, particularly if every new dataset series requires some paired recordings to initiate training. Furthermore, the two techniques record at vastly differing frequencies, and it is not obvious how spatial resolution could be extrapolated from the electrophysiological signal. This paired data, however, can be used to validate a denoising model; such efforts are detailed in Chapter 5.

For these reasons, the most viable application of deep learning to the task of Optopatch denoising requires a training scheme that relies only on raw Optopatch data with no ground truth. Self-supervised learning lends itself well to a task with this specification.

1.2.2 Self-supervised learning

Self-supervised learning differs from supervised learning in that the data itself, rather than some external objective truth, assumes the supervisory role in training. The

following algorithms apply this framework by using a prediction task as a pretext for denoiser training.

Noise2Noise (N2N)[7] is a self-supervised denoising algorithm that trains on pairs of independent realizations of the same signal source. In its conception, it operated on images but can be extended to any data modality. It performs most effectively in a regime where the same sample can be observed twice, but not enough times relative to its dimensionality to take advantage of simple averaging. A N2N model is trained on the task of predicting one image from its twin. Even though the target image of its prediction task is just as noisy as its input, the key insight is that the two noise realizations are completely uncorrelated while the shared signals are correlated. More precisely, because the input image carries no information about the target noise which can be assumed to have zero mean, the predictor's loss is minimized only by outputting the average of all possible realizations. This average is simply the image without noise.

In fact, this principle still holds with certain regularizing assumptions even when the noise has nonzero mean. For example, in the case where fewer than half the pixels, in expectation, in a set of images are corrupted with some nonzero mean noise, it is more likely that a randomly sampled pixel is correct than incorrect. Because applying L2 loss would average the noise into the true signal, the images can be recovered by training a Noise2Noise model on L1 loss, which attracts to the median of the noisy distribution.

Noise2Self (N2S)[1], alternatively called Noise2Void (N2V)[5], is a related algorithm which learns to denoise from individual noisy images by exploiting the pixel-independence of noise in an image. To gather intuition, consider a pixel $\hat{x}$ in a noisy image, which can be decomposed into the "true" value x and noise ϵ. Like the scenario described above, assuming that ϵ has zero mean allows many common loss functions to suffice:

$$\hat{x} = x + \epsilon \ ; \quad \mathbf{E}[\epsilon] = 0 \tag{1.1}$$

Note that the variance of ϵ need not be independent of x, as is not the case with Poisson-Gaussian noise in Optopatch data. Now consider an ideal predictive model tasked with predicting the noisy value $\hat{x}$, and input a version of the image where $\hat{x}$ is occluded. Because the image has pixel-independent noise, it is impossible to reconstruct ϵ, so at best, this model could infer the distribution of $\hat{x}$. For a suitable choice of loss function like L_2, this model minimizes loss by predicting $\mathbb{E}[\hat{x}]$. But this is exactly the ground truth value x. It follows that the ideal predictor of a noisy image is also its ideal denoiser:

$$\mathbb{E}[\hat{x}] = \mathbb{E}[x + \epsilon] = x + \mathbb{E}[\epsilon] = x \tag{1.2}$$

In the simplified example above, the model would be trained by applying a donut filter to each pixel and tasking it with predicting the single pixel from all other pixels. This construction works because all pixels have independent noise realizations. While this is sufficient for the purpose of denoising Optopatch data, it can be shown that a broader class of functions using more exotic filtering modes can be trained in a self-supervised manner. These functions require the property of $\mathcal{J}$-invariance. For some partitioning $\mathcal{J}$ of the data dimensions, a function is $\mathcal{J}$-invariant iff the function evaluation on the data, after being restricted to some dimensional subset $J \in \mathcal{J}$, is independent of the original data restricted to J. In the above case, $\mathcal{J}$ partitions the data into individual pixels so that each occupies an independent dimension.

This principle lays the foundation for training an Optopatch denoiser. Individual pixels can be occluded and then predicted from their spatiotemporally local contexts. By occluding the pixel, the model learns to predict pixels without relying on their actual noisy values in the input image. At evaluation time, even without any occlusion, the model would predict each pixel from its neighborhood only, thus discarding the noise component from the pixel itself.

1.3 U-Net as image-processing architecture

A U-Net[11] is a deep learning convolutional architecture which takes in as input, as well as outputs, full-resolution images. In the downward path of the U-Net, the input image is progressively convolved and downsampled to produce a deep embedding. In the upward path, the embedding is then convolved and upsampled in turn. The two paths are of equal length so that the final output is an image with the same resolution as the input. Notably, at each step of the upward path, the embedding is concatenated with one of the intermediate results from the downward path. Because each upsampling restores one level of resolution, the first upward step is concatenated with the last downward result, the second upward step with the second to last downward result, and so on.

The motivation for these lateral connections is that the deep embedding describes the low-dimensional signal contained in the image but lacks spatial positioning information. Pairing this signal with intermediate convolutions of the image restores the missing spatial context to produce the desired output image. In this manner, a U-Net combines learned low-dimensional embeddings with the aggregation of spatial information at multiple length scales. U-Nets feature prominently in deep learning algorithms for biomedical image processing and can be trained on tasks like denoising, masking, and segmentation. Indeed, the U-Net architecture lends itself well to incorporating various borrowed augmentations like local attention maps and latent representations.

An attention U-Net[10] augments the lateral skip connections with local attention gates. Rather than directly concatenating the downward path features to the upsampled image, each step in the upward path first uses the pre-upsampled image as a gating factor. Both the gating factor and the downward path features are projected to high dimensional space, added together, and passed through an activation function to produce a resampling map, from which the new skip connection features are computed. This mechanism, called an additive activation gate, infers self-attention-like associations within the data at progressively localized length scales. This U-Net is

used to segment biomedical imaging and identify outlines of organ structures.

A probabilistic U-Net[4] applies a conditional variational autoencoder to learn a posterior latent space from the image. At evaluation, the model samples from this space to augment the final U-Net layer to realize different segmentations of the image. This construction enables the model to identify a collection of likely segmentations on ambiguous images. In examples of biomedical imaging where one of multiple structures in the data could be cancerous, the model identifies a probable distribution of outcomes.

1.4 Thesis outline

This thesis introduces CellMincer, a self-supervised deep learning model for denoising voltage imaging data. Chapter 2 explains the model and training scheme of CellMincer, then it presents the series of optimization experiments used to tune its hyperparameters. Chapter 3 describes the process by which CellMincer is optimized. Chapter 4 details the benchmarking experiments used to compare the tuned CellMincer model with comparable denoising algorithms. These experiments use synthetic data so that the performance of each algorithm can be assessed on the "true" optical signal. Chapter 5 extends the benchmarking experiments to real voltage imaging data with paired electrophysiology, where the algorithms are assessed on their performance in mimicking the electrophysiological signal with the associated denoised optical signal. Chapter 6 presents the conclusions and describes additional ideas for assessing CellMincer's efficacy, methods for interpreting the denoised optical data, and potential applications in downstream analysis.

1.5 Individual contributions

Among the various efforts detailed in this thesis, the author contributed to the following:

- Developing the CellMincer model into a full Python CLI tool with all relevant

customization options accessible through configuration files

- Modifying and optimizing the model implementation to handle datasets of any dimension and of a wide range of formats
- Enabling multi-GPU training through use of the PyTorch Lightning framework
- Developing Workflow Description Language (WDL) scripts to integrate CellMincer with Terra.bio, a database and workflow manager maintained by the Broad Institute, allowing for asynchronous training and denoising operations
- Integrating CellMincer with Neptune.ai, a metadata organizer for ML training
- Designing and running optimization experiments to tune CellMincer's hyperparameters
- Designing and running the Optosynth benchmarking experiments on CellMincer and its competing algorithms
- Designing and running the benchmarking experiments on paired optical-ephys data, including the process for optical-ephys signal alignment and the peak-calling analysis

Chapter 2

CellMincer: Architecture and Implementation

CellMincer is a self-supervised deep neural network for denoising Optopatch-like data. It is robust to variations in the recording protocol and data quality, and it offers a significant improvement in the recovered signal-to-noise ratio from other standard denoising algorithms. CellMincer aggregates information from the entire target dataset into a compact global feature map. Using these global features, a single-frame U-Net extracts pixel-level features from individual frames in a fixed-length window, and a temporal post-processor convolves these features to produce a denoised middle frame. Figure 2-1 depicts the components of the CellMincer model.

2.1 Global features

Unlike training-free methods which can operate on the entire dataset in CPU memory, deep neural network models are most efficiently trained and evaluated in GPU memory. Furthermore, to use a meaningful batch size to increase the forward pass throughput and reduce gradient variance, the network must be configured to process an appropriately small input. This limits the size of the data tensor that a neural network would be able to process in one batch, which suggests an approach in which the network is constrained to view only several frames at a time.

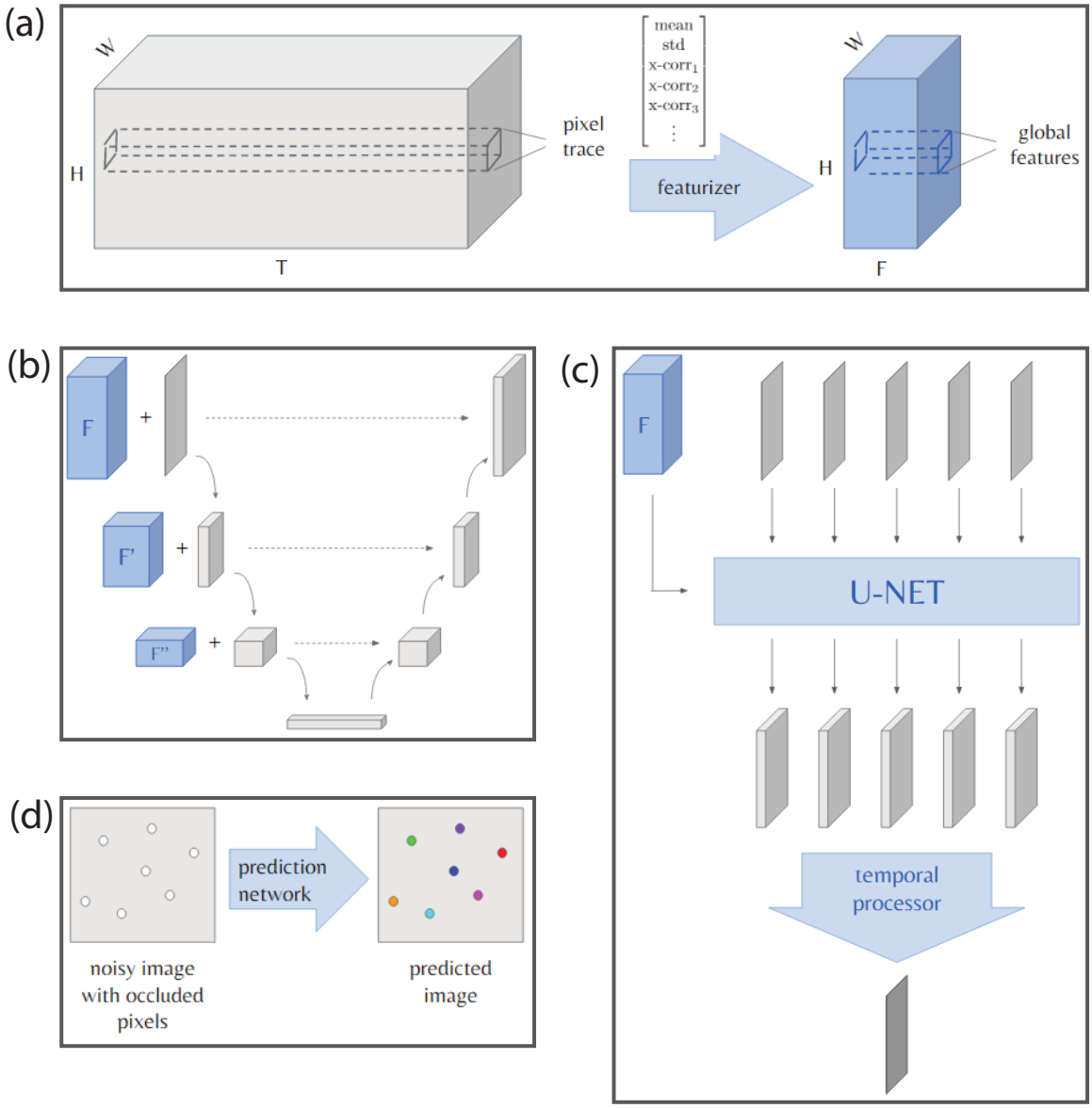

Figure 2-1: CellMincer design diagram. (a) Pixel traces are featurized to produce global feature maps. (b) U-Net architecture with global features processes an individual frame. (c) Full CellMincer model pipeline: frames within a temporal window are processed with the U-Net using global features, and the resulting frame features are convolved to produce a denoised middle frame. (d) Self-supervised random pixel occlusion and frame prediction.

While this approach may suffice for datasets in which the constituent images are independently distributed, a continuous Optopatch recording exhibits strong correlations throughout its length. Most notably, plated neurons under imaging do not drift in most circumstances, resulting in same-neuron pixels being highly correlated. Even with high Optopatch noise, such a correlation would emerge upon examining the thousands of frames in the dataset. As such, a strong denoising model would need to exploit these correlations without being able to compute on the entire dataset. This motivates the idea of precomputed global features.

Before training or denoising with the proposed model, a deep global feature map is computed on the entire dataset. From each pixel trace, various moments and cross-correlations with neighboring pixels are extracted and written to a separate tensor object. These feature vectors are one-to-one with the spatial dimensions of the dataset, so featurizing a $T \times W \times H$ dataset would produce a $F \times W \times H$ feature map, where T corresponds to the time dimension and F the feature vector length. In this implementation of CellMincer, $F = 73$. Importantly, F is fixed and is small enough for the entire feature map to fit in GPU for forward passing.

2.2 U-Net featurizer

CellMincer processes individual frames using a U-Net architecture. To incorporate the precomputed global features, CellMincer appends these features to the input image, creating a deep $(F+1) \times W \times H$ input tensor. After each stage of convolution and downsampling in the downward path, the tensor is reaugmented with a similarly downsampled global feature map. Because the global features are fully spatially resolved, downsampling them in the spatial dimensions reveals pixel-pixel correlations at longer length scales. This information further informs the network's embedding learning and subsequent spatial featurizing of the image.

One caveat of the U-Net is that while the output has full resolution, it corresponds only to a centered crop of the input, and the input edges are lost to output. This is due to the sequence of convolution operations iteratively trimming the edges of

the image. To counteract this effect, the input image must be appropriately padded so that the image is trimmed to the desired output size. While images are conventionally reflection-padded to minimize discontinuities, this approach creates a unique issue when paired with a Noise2Self-like training scheme. Using training data with narrow dimensions will capture both the boundary pixels and their reflections in every training window. As a result, the model can cheat by using the unoccluded reflections to predict the occluded boundary pixels. Trained models which do not control for this effect degenerate to predicting the identity of the data's edge pixels. CellMincer handles this exploit by occluding all padding pixels for the length of training, in addition to the randomly occluded interior pixels.

In contrast with other convolutional networks that expect particular image dimensions, this U-Net implementation accepts variable-size images. Because each downsampling operation must occur on a tensor with even spatial dimensions, the input padding size is specifically chosen to correct all downsampling parities. This condition is sufficient to successfully pass the image through the U-Net. This offers flexibility in training the network, where small crops of frames can be used as input batches, and in evaluating the network, where it is possible to train on one dataset size and denoise other differently sized datasets.

CellMincer's U-Net also offers the option to enable local attention maps as formulated in the attention U-Net architecture[10]. Same-neuron pixels, whether spatially local in the soma region or distally located along dendritic spines, are generally highly correlated throughout the recording due to their shared action potentials. It follows naturally that attention maps may explicitly identify and amplify these correlations for use in inference.

2.3 Temporal post-processor

To denoise a given frame, CellMincer selects a short temporal window centered on that frame and passes each frame in that window through the U-Net, along with the global features computed on the entire recording. These feature outputs are concatenated

along a time dimension and given to the temporal post-processor, which applies a series of 3D 1×1 convolutions. These convolutions separately act on each pixel to collapse the time and feature dimensions, ultimately producing the desired denoised middle frame. Selecting a temporal window length corresponding to the approximate timescale of individual events (e.g. spikes) will improve the post-processor's ability to identify and reconstruct these structures in denoising.

This final step in the CellMincer pipeline marries the two core principles established above: exploiting global features and constraining the model's focus to small segments of the recording. It is argued below that this construction would outperform algorithms that only exhibit one of these characteristics.

2.4 Self-supervised training

CellMincer is trained on noisy Optopatch recordings in a self-supervised manner because there does not exist a viable method to construct the ground truth for a recording or to produce a sufficiently noise-free recording to substitute for ground truth. As such, CellMincer does not require a clean reference for its training data. Even with only a single recording, CellMincer can be trained on only that recording and subsequently used to denoise it.

Adapting the training scheme of Noise2Self, detailed above, for CellMincer yields the following training procedure. Each iteration, a random crop of a random frame is selected, and some of the pixels within the crop are occluded. In this scheme, occluded pixels are replaced with a realization of a random Gaussian variable with mean and variance equal to that of the pixel's empirical distribution over the entire recording. This approach, as opposed to zeroing out the occluded pixels, obscures which pixels are occluded so that the model treats all pixels as occluded. Once the crop is denoised with the network, only the denoised occluded pixels are compared with their respective noisy values in the input crop, and the network is updated with gradient descent on the resulting loss. When the network is actually used to denoise a noisy frame, no occlusion is performed. The network will attempt to predict the

value of each pixel without relying on its actual value in the input, producing a fully denoised frame.

2.5 Conceptual improvements over comparable algorithms

Penalized matrix decomposition (PMD)[17] approximates a matrix as a low-rank factorization with regularization. Each optical electrophysiology recording can be approximated as a noisy union of independent static components (neurons), each associated with a pattern of activations. Applying L1-regularization biases the factorization to select sparse components, particularly suitable for such datasets where the number of neurons is greatly dwarfed by the number of degrees of freedom in the data. As a training-free algorithm, PMD is limited to using the information in the target dataset to compute a solution, and its matrix factorization, by design, imposes a heavy bias on its output. By contrast, CellMincer can learn patterns of inference from multiple datasets, including its denoising target, and its deep neural network architecture confers greater power in modeling nonlinear behavior.

Noise2Self[1] is a basic self-supervised deep neural network that denoises images and can be trained to individually denoise the collection of frames in an optical electrophysiology dataset. While Noise2Self offers the expressiveness of nonlinear modeling, its inability to use other frames in its inference task prevents it from exploiting the full range of spatiotemporal correlations present. CellMincer, initially in a similar position due to its fixed temporal denoising window, addresses this drawback with the inclusion of global features which aggregate information from the entire dataset.

DeepCAD[8] is a self-supervised deep neural network for denoising calcium imaging. In a training window, every 2nd frame is concatenated as input, and the model is trained to interpolate between them and predict the missing frames. This formulation of the prediction task is well-suited to calcium imaging, which operates on a much coarser timescale and is considerably less volatile from frame to frame. By contrast,

CellMincer retains the entire temporal neighborhood around a frame for denoising and uses the most temporally local datapoints for inference.

DeepInterpolation[6] is another self-supervised deep neural network for denoising various forms of functional imaging. In training, the model removes a central frame from a temporal window and predicts the entire frame from its local context. While it accesses more temporally local data than DeepCAD, a notable drawback is its inability to use same-frame pixels for inference. Certain pixels in the same timestep are expected to be highly correlated to the target pixel and are particularly important in inference on voltage imaging, where spiking events can occur over 1-3 frames. By individually occluding individual pixels at random as opposed to discarding the entire target frame, CellMincer learns to exploit all same-frame information for inference.

2.6 Implementation

The CellMincer model is implemented using the PyTorch framework and wrapped in PyTorch Lightning, which consolidates the training logic and allows for GPU-agnostic training. CellMincer's protocol is packaged into four CLI commands: *preprocess*, *feature*, *train*, and *denoise*. Each command takes various dataset file and directory paths as input and writes its output to a designated path.

The preprocess command takes a raw dataset file and learns a baseline trend for each pixel trace. It subtracts this per-pixel trend from the dataset, producing a trend and a residual component. Optionally, it removes jitter from the dataset caused by inadvertent strobing of the reporter laser or undesired motion in the camera. The trend and residual components are written to a designated dataset directory. The trend, which roughly corresponds to the background fluorescence, is left out of downstream processing and added back only at the end of denoising.

The feature command computes a predetermined set of global features on the residual component of a dataset. These features include various moments of the trace distribution as well as cross-correlations with neighboring pixels. If the dataset has defined periods of stimulus and dormancy, only the stimulated frames are used.

These features are written to the corresponding dataset directory to be used in model training and denoising.

The train command trains a CellMincer model from scratch on one or more preprocessed and featurized datasets. It takes in a configuration file that supplies various model specifications and training settings. During training, it periodically outputs a checkpoint file which describes the full model architecture, parameters, and training status. This checkpoint can be used to restart training with no loss of work, particularly helpful when running on preemptible virtual machines. The final checkpoint is written to a specified directory and can be used for model evaluation.

The denoise command takes in a preprocessed, featurized dataset and a model checkpoint and evaluates the model on the full dataset. It then readds the subtracted trend and outputs this final denoised result as well as a normalized video rendering of the data for visualization purposes.

CellMincer is a robust, production-grade software package that performs well on various modes of optical electrophysiology, accepts a range of file formats, and can be easily configured through its compartmentalized configuration files. This contrasts with other software packages which often need significant upfront time to transform an input dataset to fit within the constraints of their implementation.

2.7 Sample denoising results

Figure 2-2 depicts two normalized sample frames from two datasets, before and after denoising.

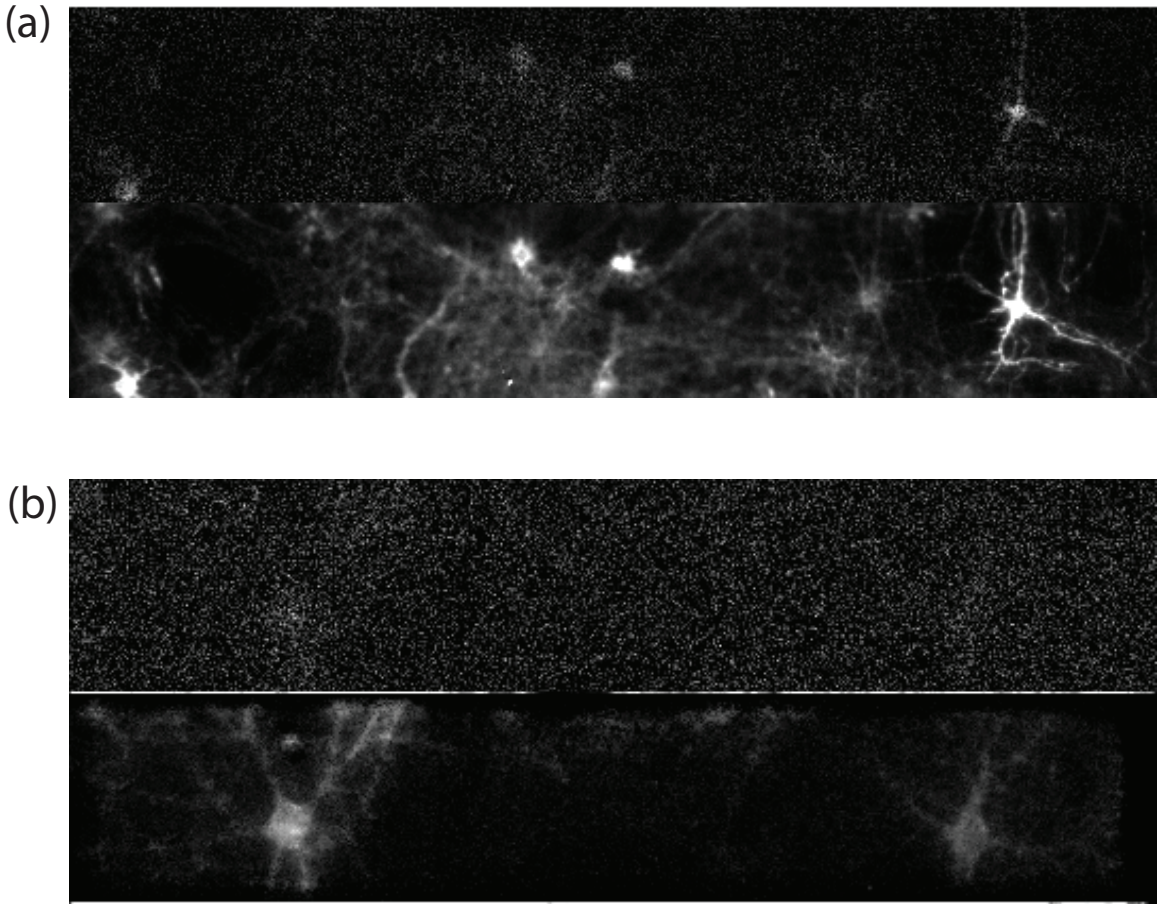

Figure 2-2: Each image depicts a sample dataset frame twice: raw (above) and denoised (below). (a) Dataset from Farhi Lab. (b) Dataset from Miller Lab.

Chapter 3

Model Optimization

To tune the hyperparameters of CellMincer, datasets containing ground truth are needed to quantitatively measure the accuracy with which the model reconstructs the optical recording. Because Optopatch data collection with ground truth is not feasible, especially at scale, a synthetic alternative to Optopatch is needed. Optosynth, a synthetic generator of Optopatch-like data, fulfills this need. The data generated by Optosynth is used to train variations of the CellMincer model, whose performances are measured by denoising Optosynth datasets and comparing the outputs to the ground truth recordings. Through this process, the strongest-performing variation of CellMincer is selected for use on real Optopatch datasets.

3.1 Optosynth: synthetic Optopatch data generator

Optosynth is a generator for synthetic voltage imaging data based on the single-neuron database maintained by the Allen Brain Institute. From this database consisting of individually measured human and mouse neurons, electrophysiological data and morphologies are sampled, and Optosynth simulates plating these "neurons" together. This plating incorporates variance in light diffusibility as a result of unequal depth positioning. The simulation of Optopatch stimulus accounts for propagation delay, non-uniform archaerhodopsin expression, and dendritic backpropagation, among other biological phenomena. The resulting fluorescence data is distorted by applying

PSF blurring, Poisson-Gaussian camera noise, and background fluorescence to mimic the complications in Optopatch imaging. Figure 3-1 depicts sampled electrophysiological and morphological data from the Allen Brain database, as well as several stages of Optosynth data generation.

3.2 Performance metric

The primary metric for reconstruction quality used in the simulated data case is peak signal-to-noise ratio (PSNR). This quantity, proportional to the log of inverse pixel MSE, describes the fidelity of a corrupted image with respect to its ground truth. It is measured in decibels, where an increase of 2 dB corresponds to a hundred-fold increase in the signal-to-noise ratio (SNR). The PSNR of a denoised recording with respect to its ground truth is computed by averaging the PNSR between each corresponding pair of frames.

3.3 Partitioning of hyperparameter search space

The large number of tunable hyperparameters, which collectively influence the model architecture, loss computation, and training scheme, makes a complete grid search unreasonable. Thus, to tune the model, the hyperparameter space is partitioned into the three aforementioned facets of CellMincer training, and each facet is separately perturbed while fixing the other hyperparameters.

3.3.1 Model architecture

Among the first key decisions in model architecture is the temporal post-processing of features produced by the single-frame U-Net into one denoised frame. Under initial consideration was the 3D convolutional approach detailed previously as well as a "timestep-as-features" approach which concatenates all frames along the feature dimension and performs a series of 1×1 convolutions to produce the output frame. The latter approach, while simpler, proves to be less powerful on account of its lack of

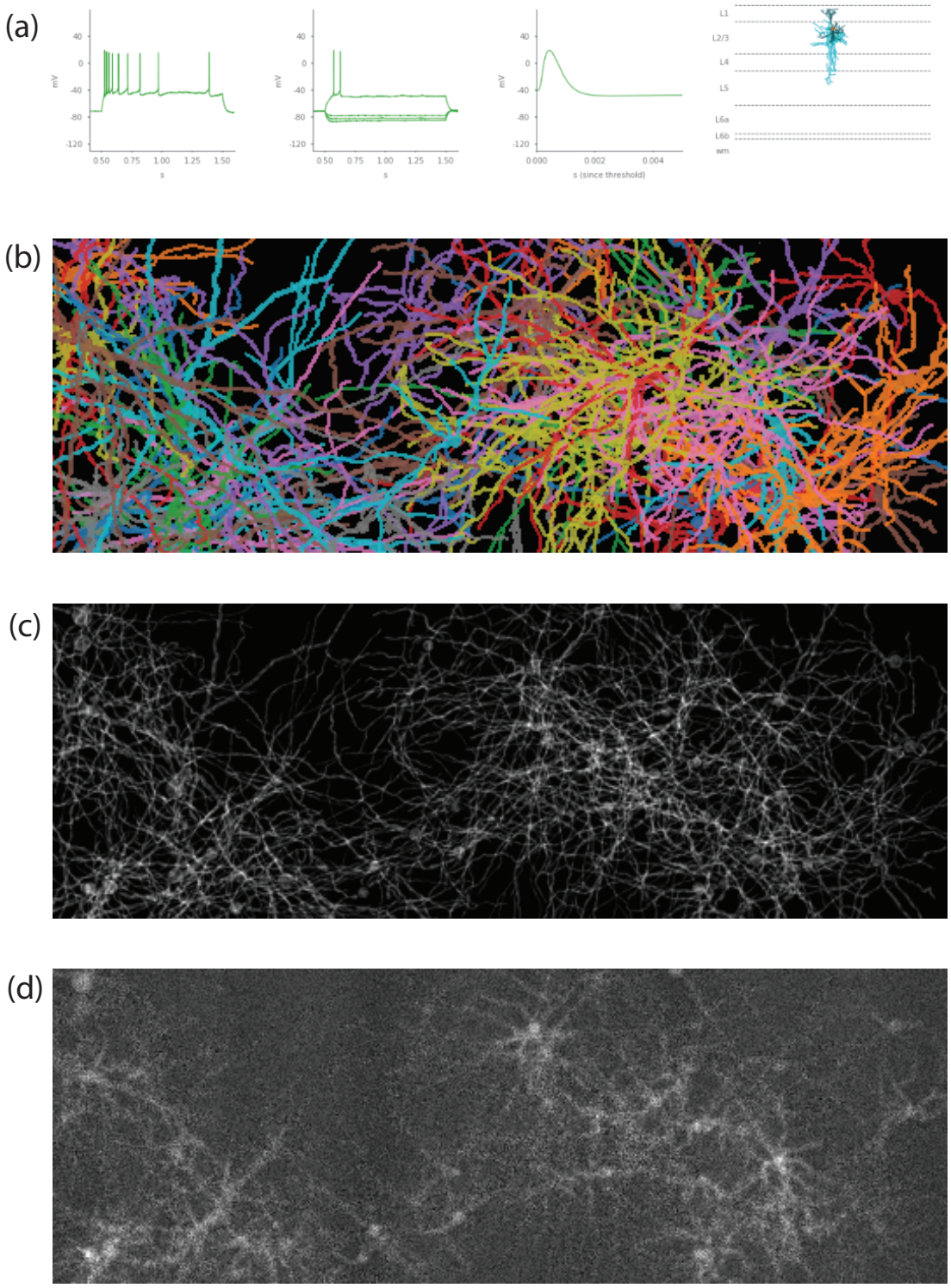

Figure 3-1: An overview of the Optosynth data generation pipeline. (a) A sample neuron electrophysiology and morphology from the Allen Brain single-neuron database. (b) A synthesized "plating" of neurons to be imaged. (c) A sample frame of synthesized fluorescence. (d) The same frame after applying PSF, Poisson-Gaussian camera noise, and background fluorescence.

explicit encoding of temporal sequence, as all input frames are treated as symmetric components of the feature map. In subsequent experimentation, this architecture is removed from consideration.

The chief architectural decision of interest is the use of global features, a primary differentiating factor between CellMincer and comparable denoising algorithms. As discussed previously, the driving intuition behind the introduction of global features into the single-frame U-Net processing is that an Optopatch frame is too strongly correlated with the activity occurring in the rest of the recording to be effectively denoised independently of the other frames. To supplement the U-Net's limited temporal view, global features are calculated between entire pixel traces spanning the length of the recording and used to augment the U-Net input. These global features include summary statistics on individual traces and cross-correlations with its neighbors, which serve as much stronger indicators that certain pixels belong to the same neuron than the noisy realizations from an individual frame. Three settings for global features are considered: no use, single concatenation with the U-Net input, and repeat concatenation with the U-Net's downward path by downsampling the features at each stage to match the dimensions of the transformed data.

The temporal order is the length of the temporal window from which the single middle frame is denoised. A larger temporal order draws information from a wider context but, past some sensible limit, can also dilute the model with temporally distal data that lacks relevance. Intuition hints that an appropriate temporal order would encompass a complete event in the data, like an individual spike, to offer the model enough context to understand how the frame of interest fits into said event. An immediate consequence of this reasoning is that the optimal temporal order must depend on intrinsic aspects of the data like the frame rate and the type of functional imaging. As an example, calcium imaging operates over much longer timescales than voltage imaging, which can capture millisecond-long spiking. These factors determine the number of timesteps that contain such a complete event. Because frame rate is kept consistent throughout many implementations of Optopatch imaging, this limitation of CellMincer architecture does not cause an issue within the scope of this thesis. The

default temporal order is 9 frames, with orders of 5, 13, 17, and 21 tested.

The depth and width of the U-Net broadly influence its capacity to extract latent features from convolution, and it is generally unsurprising that doubling or quadrupling the number of U-Net parameters would improve the model's inference power. As a result, considerations about the U-Net's computational complexity include hardware limitations, as certain choices of complexity, when combined with other architectural options, result in models too large to fit on available GPUs. The default depth and width of the U-Net are 3 layers and 32 channels respectively, so that the downward and upward convolution paths are each of length 3. Experiments include 2 layers and 64-wide channels as alternatives.

In the temporal post-processing step, the features of the denoising window are collapsed into a single value for each pixel over a series of 1×1 convolutions. Convolving the feature dimension over multiple steps increases the parameter space of this post-processor, producing a more complex network. With n being the starting number of features, the default feature reduction path is $n \to n/2 \to 1$, with alternatives being $n \to n \to n/2 \to 1$ and $n \to n \to n \to 1$.

Finally, the use of local attention and batch normalization are included as non-specific augmentations to the U-Net. Local attention is implemented as previously described in Chapter 2. Batch normalization, present in many neural network formulations applied to a range of tasks, generally improves training stability when the input distribution varies widely between batches. This is often the case with Optopatch data, where frames recorded during stimulation are considerably brighter than background frames where no activity occurs. By default, CellMincer excludes these options, with experiments testing the inclusion of attention, batch normalization, and both.

3.3.2 Loss computation

The tunable hyperparameters involved in the computation of reconstruction loss are the loss function and the pixel censoring probability.

The initial loss functions under consideration are L2, L1, annealed L0, and Poisson-

Gaussian. Of these four functions, the first three differ only in the power to which the absolute pixel error is raised, with L0 loss annealing the power from 2 to 0 over the course of training. Poisson-Gaussian loss differs from these functions in that the variance of pixel error scales with the ground truth of said pixel. Without access to ground truth during training, this value is estimated from the noisy recording. The use of Poisson-Gaussian loss is motivated by the Optopatch camera noise being Poisson-Gaussian in distribution, so that errors are more tolerated on brighter pixels. In practice, Poisson-Gaussian and L0 loss produce significantly lower PSNR than L2 and L1. This result can be explained by the intuition that PSNR varies inversely with MSE, and L2/L1 loss are most effective at minimizing MSE. As a result, models in subsequent experiments are trained with L2 or L1 loss only.

The latter hyperparameter, which specifically relates to the self-supervised training strategy used by Noise2Self and by CellMincer, specifies the expected density of censored pixels in the training window. During the training window cropping process, the pixels in the middle frame are censored according to a Bernoulli process parameterized by this probability. Increasing this probability increases the model training speed by directly increasing the number of pixels on which loss is computed. However, this increase also further corrupts the middle frame and reduces the model's capacity to use pixels in the same frame for inference. By default, models were trained with censoring probability 0.05, with choices of 0.01, 0.1, and 0.2 tested in experimentation.

3.3.3 Training scheme

The learning rate schedule, iteration count, and optimizer constitute the tunable components of the training scheme. Because these facets of training are insensitive to the choice of model specification, they were decided prior to fully optimizing the model architecture and loss. The optimal learning rate schedule was determined to have a linear warmup to $\eta = 1\text{e}{-4}$ that spans 10% of the total iterations, followed by an annealed cosine decay to 0 over the remainder of training. Between the choices of 50,000 and 100,000 iterations, the latter yields an insignificant improvement overall, so all models are trained to 50,000 iterations. The Adam optimizer parameterized by

$\beta_1 = 0.9, \beta_2 = 0.999$ significantly outperformed the SGD optimizer with momentum 0.9.

3.4 Initial optimization experiments

A base version of the CellMincer model is specified by the aforementioned default hyperparameters, and variants of the model are constructed by changing a single hyperparameter from the default to one of the experimental values. The data used to train and validate these models are five Optosynth datasets generated with the same neuron count and fluorescence level. Each CellMincer variant is trained on the first three datasets for 50,000 iterations, and the trained model is then evaluated on the three training datasets as well as the two unseen datasets as validation.

For each denoised recording, the PSNR is separately calculated between each stimulation frame and its corresponding ground truth. The same is done for the raw recordings to establish a baseline PSNR. Notably, frames when no stimulation is happening are excluded from the comparison as denoising them is trivial and they do not contain any meaningful neuronal activity to analyze. Each denoised frame PSNR is subtracted by the corresponding raw frame PSNR to compute the gain achieved through denoising. The per-frame PSNR gains of the training datasets and validation datasets are separately pooled and their distributions reported in Figure 3-2.

To briefly summarize the results of this experiment, immediately evident is the observation that the use of global features in any form produces a 4dB increase in PSNR gain, with repeat feature use marginally outperforming single use. This gain is also accompanied by a five-fold drop in PSNR variance. This improvement dwarfs those of the subsequent perturbations, motivating a extension of the same experiments on global feature-enabled models.

Among the temporal window sizes, the models perform progressively better with larger windows, with the trend extending as far as 21 frames. More complex U-Net architectures perform better overall, with the effect diminishing marginally in the out-of-sample evaluations. The effect of temporal post-processor complexity is negligible.

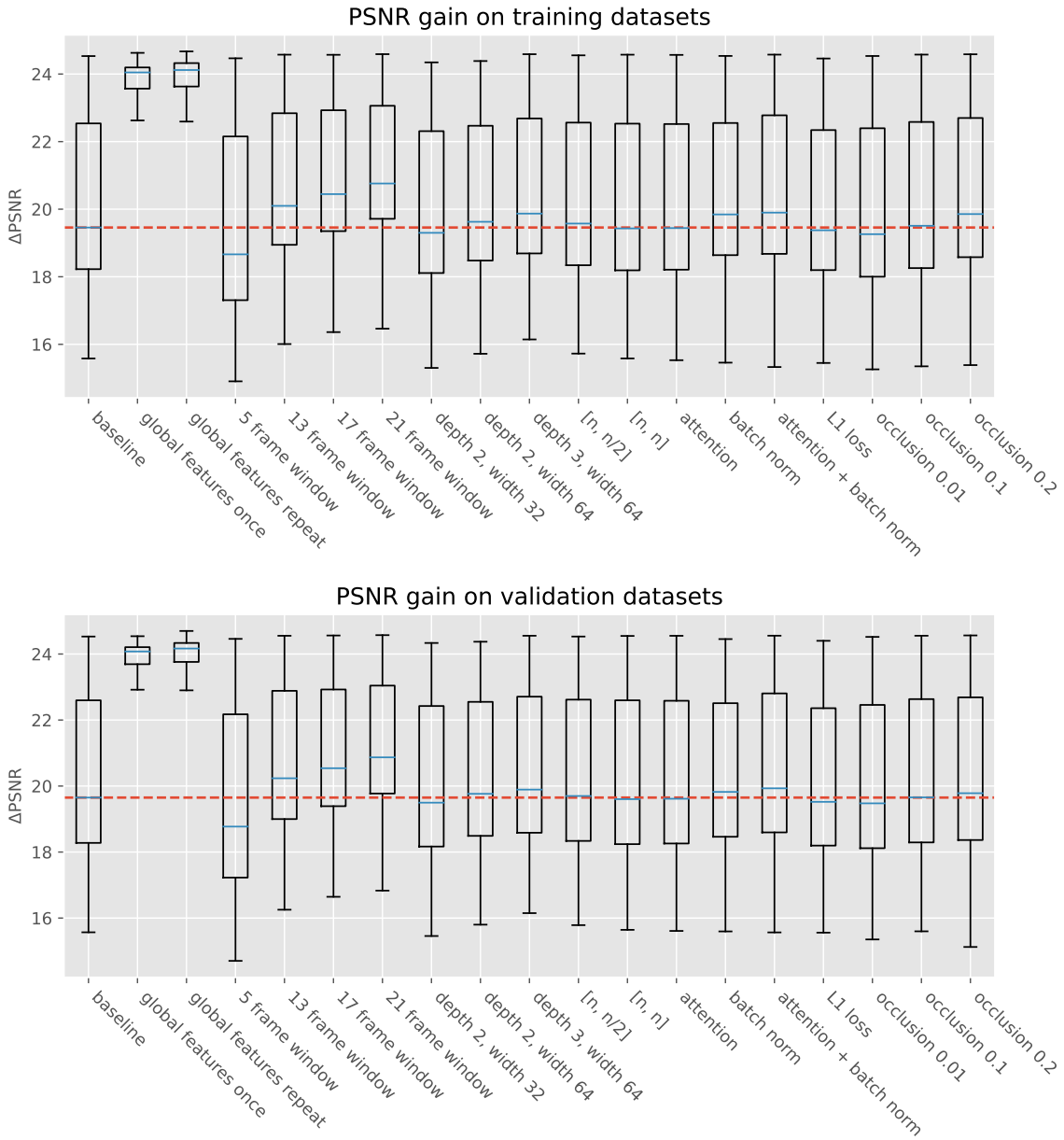

Figure 3-2: PSNR distributions on training datasets and validation datasets for each CellMincer variant. The median PSNR of the baseline distribution is marked.

The model improves when adding local attention layers and batch normalization to the U-Net. L2 loss performs marginally better than L1 loss over all datasets. Finally among the choices for censoring probabilities, the higher values produce better performing models, with 0.2 as the best performing value.

Because the inclusion of global features radically improves the model's performance, the effects of the secondary hyperparameters may be better understood in the context of using global features. To these ends, the above experiments are repeated after incorporating repeat global features into the baseline, and their results are reported in Figure 3-3. Most apparent is the change in PSNR trend when varying the temporal order. When using global features, increasing the temporal window size does not meaningfully increase the resulting PSNR gain after the 9-frame baseline window. An explanation for this change may be that by supplying global features computed on the entire recording, the model no longer needs larger windows to accurately denoise. In other words, the larger window size serves as a proxy for summary features on the full recording. This result reinforces the idea that the global features, representing long-range information from the recording, specifically confer additional inference power to the model.

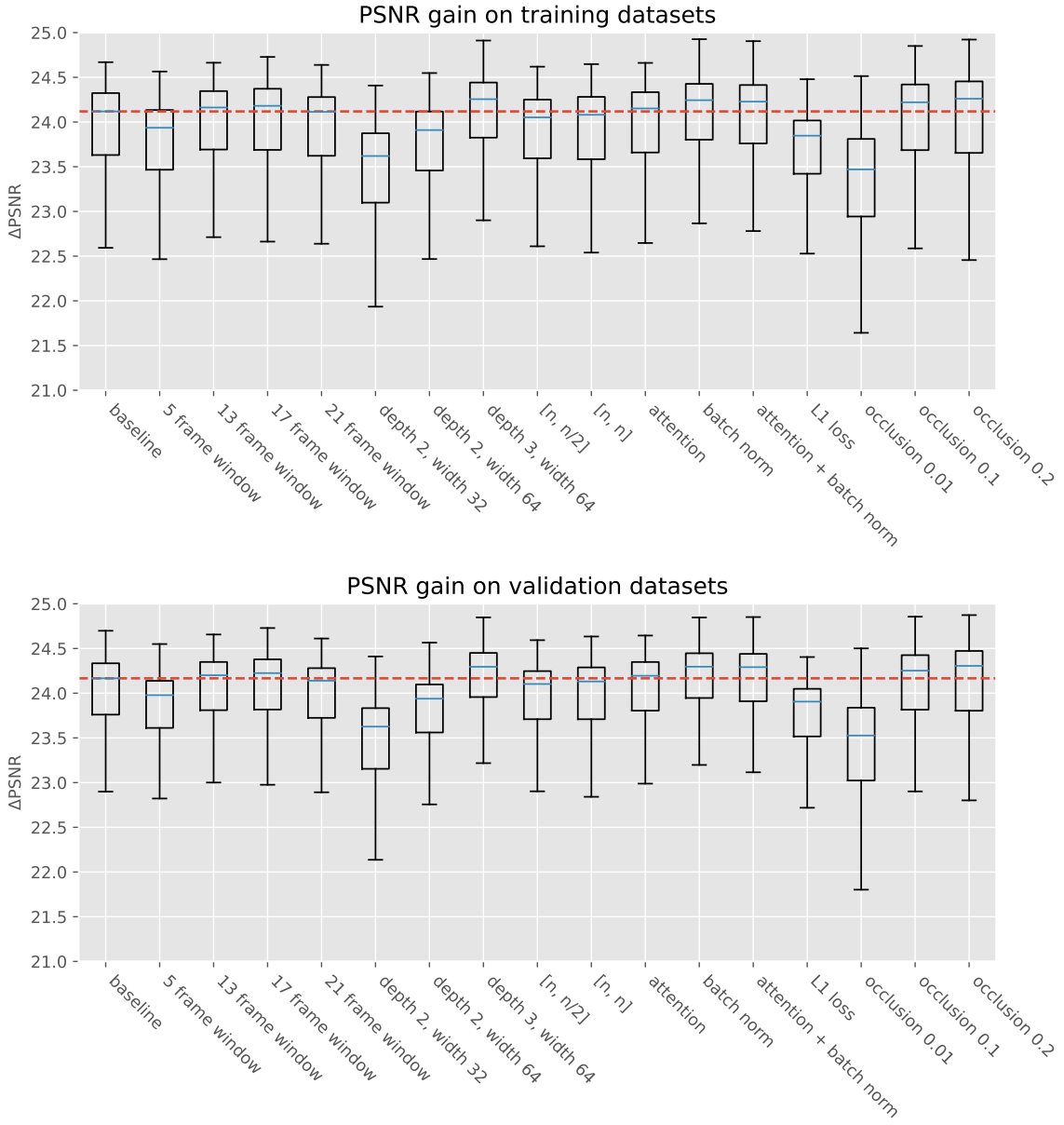

Figure 3-3: PSNR distributions on training datasets and validation datasets for each CellMincer variant using repeat global features. The median PSNR of the baseline distribution is marked.

Chapter 4

Benchmarking CellMincer on Optosynth

To demonstrate CellMincer's efficacy in denoising various forms of Optopatch data, a suite of performance benchmarking experiments are run between two variations of CellMincer, with and without global spatial features, and other image denoising algorithms. From these experiments, it is shown that CellMincer produces a significantly greater gain in PSNR on Optosynth data over other methods. Furthermore, for paired datasets containing an Optopatch recording and associated patch clamp electrophysiology, CellMincer produces a denoised optical recording that more closely aligns with the electrophysiology signal, particularly in the regime of subthreshold neuronal activity.

Of the four algorithms with which direct comparisons have been drawn with CellMincer in Chapter 2, PMD[17] and single-frame Noise2Self[1] feature in the following benchmarking experiments. In initial trials, DeepCAD[8] could not be configured in a manner that produced results of comparable quality on the provided voltage imaging datasets. This may have been a consequence of DeepCAD having been designed for calcium imaging, which operates on a significantly different timescale. DeepInterpolation[6], which was built using the TensorFlow framework, could not be made compatible with available virtual machine configurations. For these reasons, DeepCAD and DeepInterpolation are excluded from benchmarking.

4.1 Performance on Optosynth data

Using the same configuration for training and evaluation on Optosynth as previously detailed, the four algorithms are evaluated and their PSNR distributions shown in Figure 4-1. On both training and validation datasets, CellMincer, when using global features, produces the highest average PSNR gain with the lowest variance. Without global features, CellMincer lacks long-range temporal context and performs significantly worse but remains an improvement over single-frame N2S. PMD, which computes a regularized matrix factorization over the entire recording, performs second best on validation data but falls short of CellMincer even though the model has never seen the validation datasets.

To examine the source of denoising error, Figure 4-2 shows the signed residual heatmaps between the denoised and clean recordings at frame 6500 of dataset 1, which lies in a stimulation period. As one may expect, the erroneous regions concentrate at the locations of fluorescent neurons. It is evident that N2S denoises poorly throughout the frame, whereas errors by CellMincer and PMD occur at particularly active areas, such as the overlap between firing neurons. By plotting the trace of a chosen soma pixel, the difference in reconstruction quality between CellMincer and PMD becomes more apparent. PMD does not effectively distinguish between the high-frequency spiking events and the low-frequency subthreshold activity, resulting in the persistent high-frequency noise seen in Figure 4-2. CellMincer with global features, by contrast, produces a significantly stabler trace.

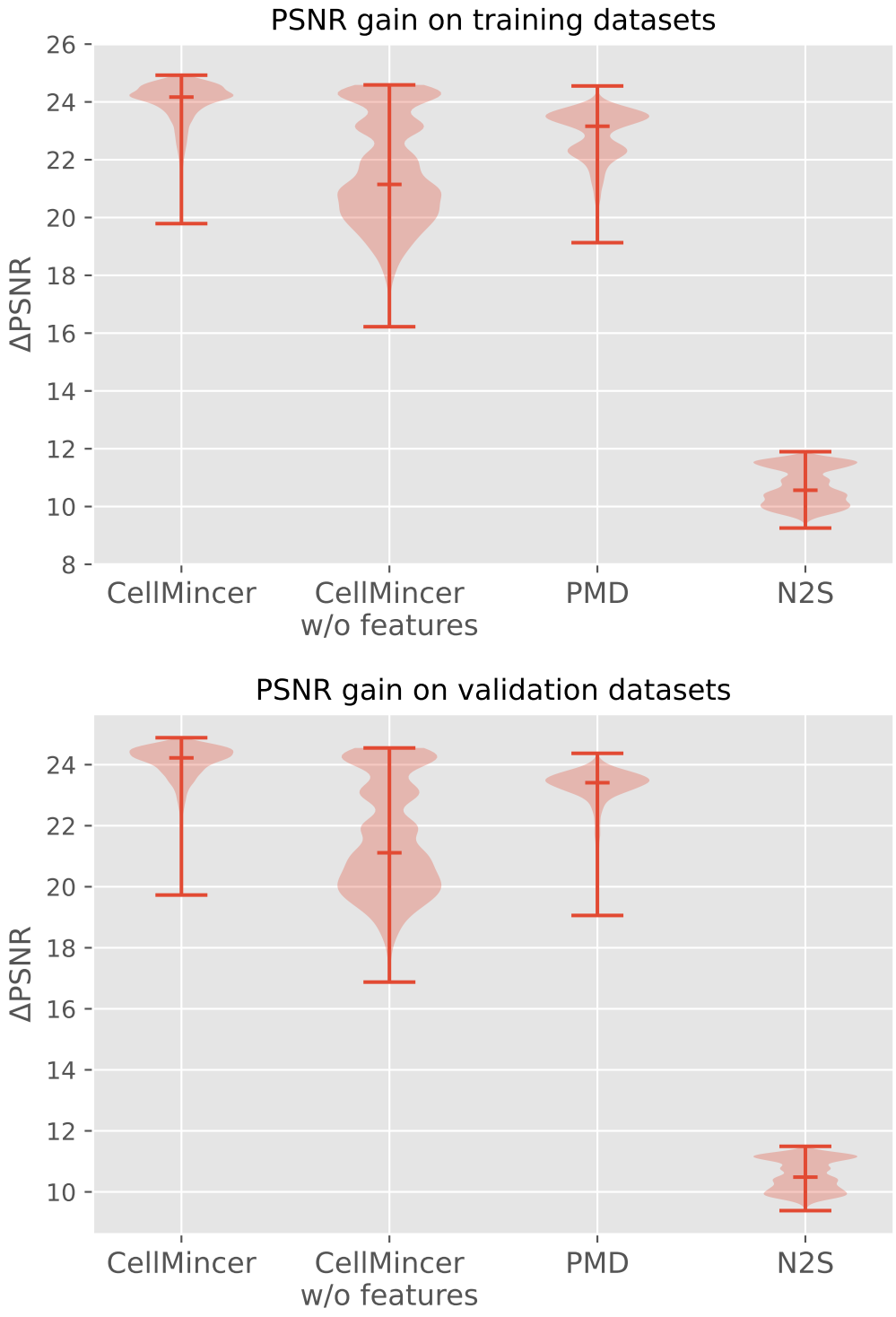

Figure 4-1: PSNR gain distributions with labeled medians over in-sample and out-of-sample Optosynth datasets for CellMincer and benchmark algorithms.

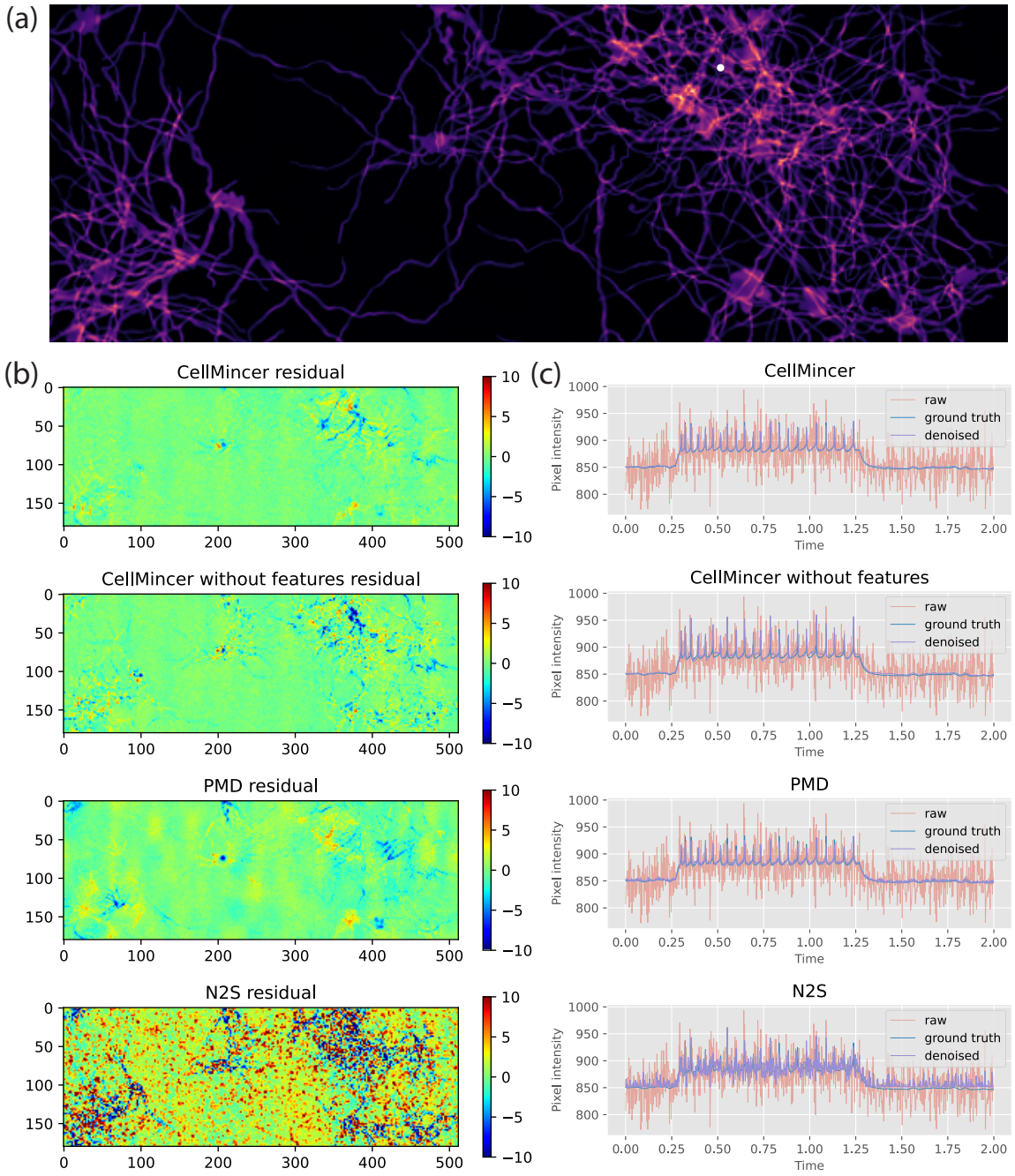

Figure 4-2: (a) Ground truth frame 6500 of Optosynth dataset 1, from which the residual heatmaps are computed. The soma pixel from which the traces are extracted is marked in white. (b) Signed residual heatmaps between each denoised and clean frame, where values are clipped between -10 and 10. (c) Overlaid soma pixel traces from the denoised data, the ground truth, and the raw data.

Chapter 5

Benchmarking CellMincer on Paired Optical-Ephys Data

To record a paired optical-ephys dataset, a single Optopatched neuron is patch clamped, and under stimulation, the neuron activity is measured optically and through the patch clamp probe. Figure 5-1 depicts the components of such a dataset. Because the optical signal produced by the neuron is a noisy linear transformation of the ephys signal, an ideal denoiser would be capable of inferring a signal from the optical data that can be transformed into the ephys signal. The objective of these experiments is to evaluate the quality of alignment between the denoised optical data produced by each algorithm and the ephys signal.

5.1 Denoised alignment of paired optical-ephys data

For each of three paired datasets, each algorithm is trained on the dataset and then used to denoise it. Using an image of the neuron, the denoised optical data is averaged over a manually selected region traced over the soma, producing a single optical trace. This optical trace is aligned to the ephys trace through the following series of signal processing operations.

The optical data is recorded at 500Hz, while the patch clamp electrophysiology is recorded at a significantly higher 50kHz. To transform the ephys signal to 500Hz, the

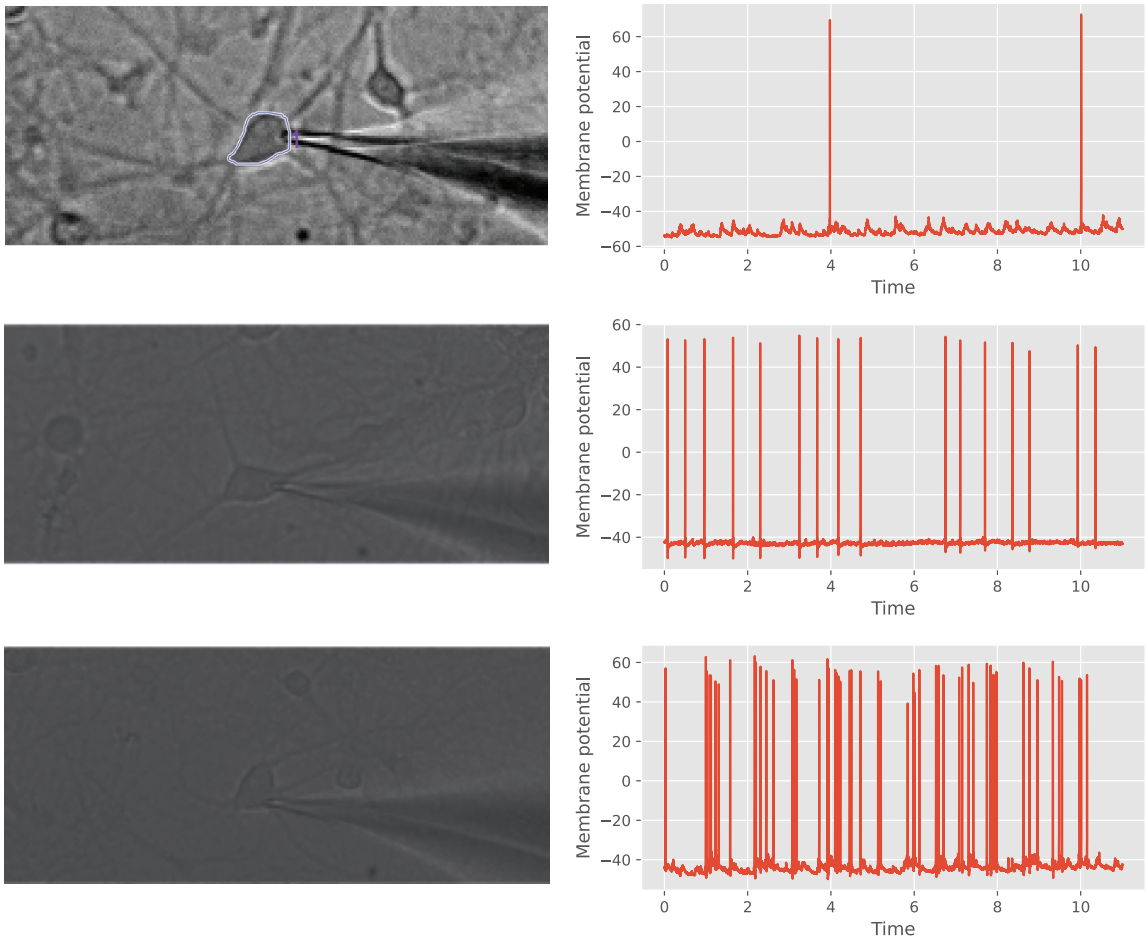

Figure 5-1: Visualizations of three paired optical-ephys datasets, each consisting of the image of the patch clamped neuron and the associated ephys sweep.

signal is first lowpass filtered to 500Hz and then downsampled at a 1:100 rate. Because the recording start and end times vary between the two signals, the spike locations in both signals are determined with manually tuned peak-finding, and a temporal shift minimizing the cumulative distance between the two sets is applied. To remove low frequency baseline movement, both optical and ephys signals are highpass filtered to 2Hz, producing signals O_{lo} and E_{lo} respectively.

After these signal processing operations, O_{lo} is effectively an affine transformation of E_{lo}, with a notable exception at the spike locations. This is because the ephys signal maintains its maximum spike values for a significantly shorter duration than the time interval between optical datapoints, so the optical data fails to capture the true spike maxima. To control for this discrepancy when computing the aforementioned affine transformation, the signals are additionally lowpassed to 20Hz to produce O_{hi} and E_{hi}, effectively smearing out the spikes. Using linear regression, the affine transformation minimizing the distance between O_{hi} and E_{hi} is computed, and this transformation aligns O_{lo} to E_{lo} in Figure 5-2. In particular, the difference in alignment accuracy is most pronounced in the subthreshold regions of the signal, depicted in the figure insets.

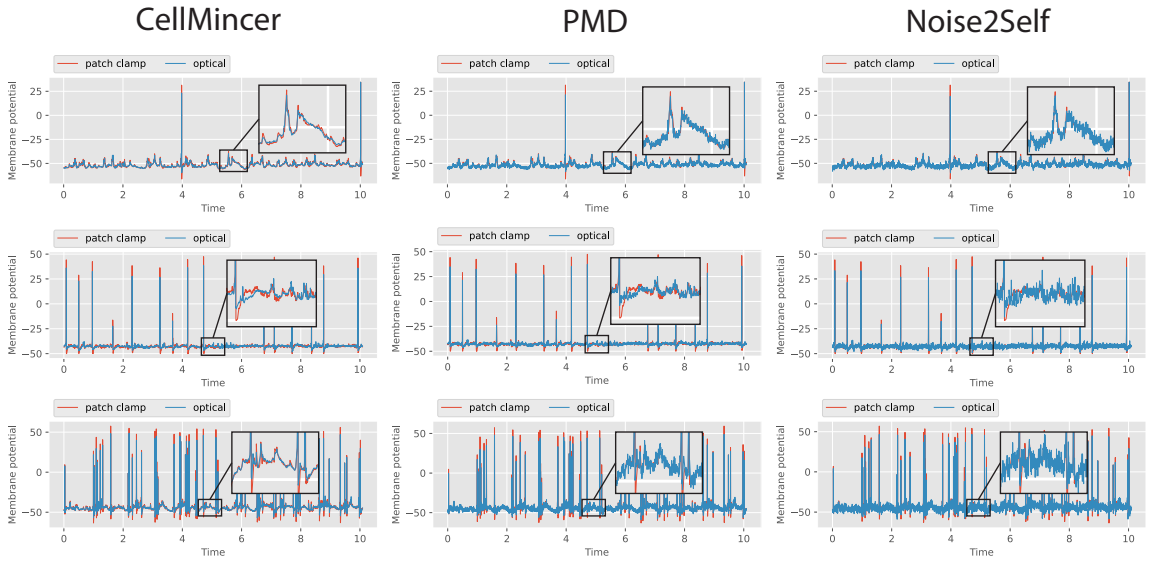

Figure 5-2: Optical-ephys alignments for three paired datasets each denoised by three algorithms.

5.2 Metrics for alignment accuracy

The MSE between the two aligned signals of a paired dataset describes the overall error over all timepoints along the signal. However, a significant component of this error for all datasets and denoising algorithms is the discrepancies in spike height, which as described above is not reconstructible from the optical signal. This error is largely contained in the spiking regions, defined here as the datapoints within 0.04 seconds of a datapoint exceeding a manually determined spike threshold. To control for this property of the data, a subthreshold MSE is computed on only the datapoints outside these spiking regions. Both full and subthreshold MSEs over all datasets and denoising algorithms are reported in Table 5.1.

Dataset #	CellMincer	PMD	N2S
1	**0.55**	1.11	1.70
	0.47	0.97	1.59
2	**2.19**	2.43	3.22
	1.11	1.37	2.06
3	**3.47**	5.65	7.75
	1.67	3.81	5.92

Table 5.1: Full MSE and subthreshold MSE between denoised optical and aligned ephys signals for each of three paired datsets denoised by each of three algorithms.

All denoising algorithms are effective at reconstructing the spike locations, as the signal is easily discernible even before denoising the optical signal. A notable advantage of denoising with CellMincer is the increased fidelity with which its denoised optical signal models the ephys signal in the subthreshold regime. In the alignments depicted in Figure 5-2, the benchmark algorithms PMD and N2S retain a persistent high-frequency noise in the subthreshold regions, while CellMincer's denoised optical signal more smoothly tracks the ephys signal. This difference can be visualized through a comparison of denoised optical spectrograms and quantified by measuring optical peak-calling accuracy.

5.2.1 Spectrogram analysis

Paired dataset 1, which exhibits sparse spiking and predominant subthreshold activity, is a suitable case in which to visualize the removal of high-frequency noise in CellMincer's output. Figure 5-3 shows the spectrograms of the ephys signal and of each denoised optical signal as heatmaps, where each column depicts the distribution of frequency energies over a short temporal window of the signal. In the ephys spectrogram, the higher frequencies are present only at the two spiking moments and, to a lesser degree, some of the subthreshold peaking. This property is largely retained in CellMincer's spectrogram, while the higher frequencies are present throughout the recording in those of PMD and N2S.

5.2.2 Peak-calling accuracy with progressive prominence thresholding

As discussed previously, a neuron ephys signal exhibits spiking in which the neuron receives sufficient stimulus to cause an action potential, as well as subthreshold activity where a neuron's membrane potential shifts but does not fully depolarize. The latter activity produces smaller, less prominent peaks in the recording. The neuron spikes, which are the most prominent peaks in the signal, are easily identified by all denoising algorithms, but qualitatively, it is evident that only CellMincer accurately reconstructs the subthreshold peaks. In PMD's and N2S's output, by contrast, the high-frequency noise in the subthreshold regime introduces many spurious peaks. This observation motivates the framing of optical denoising as a peak-calling task on the aligned ephys data as a method of measuring the quality of subthreshold reconstruction.

This analysis can be formalized as follows: for a choice of prominence threshold p, the peak locations in the optical and ephys signals with prominence exceeding p are respectively assigned as the sets of predicted and actual peaks. Viewed as a binary classification problem, the predicted and actual peak sets produce an F1-score. Rather than require an exact timepoint match, the average time interval w between

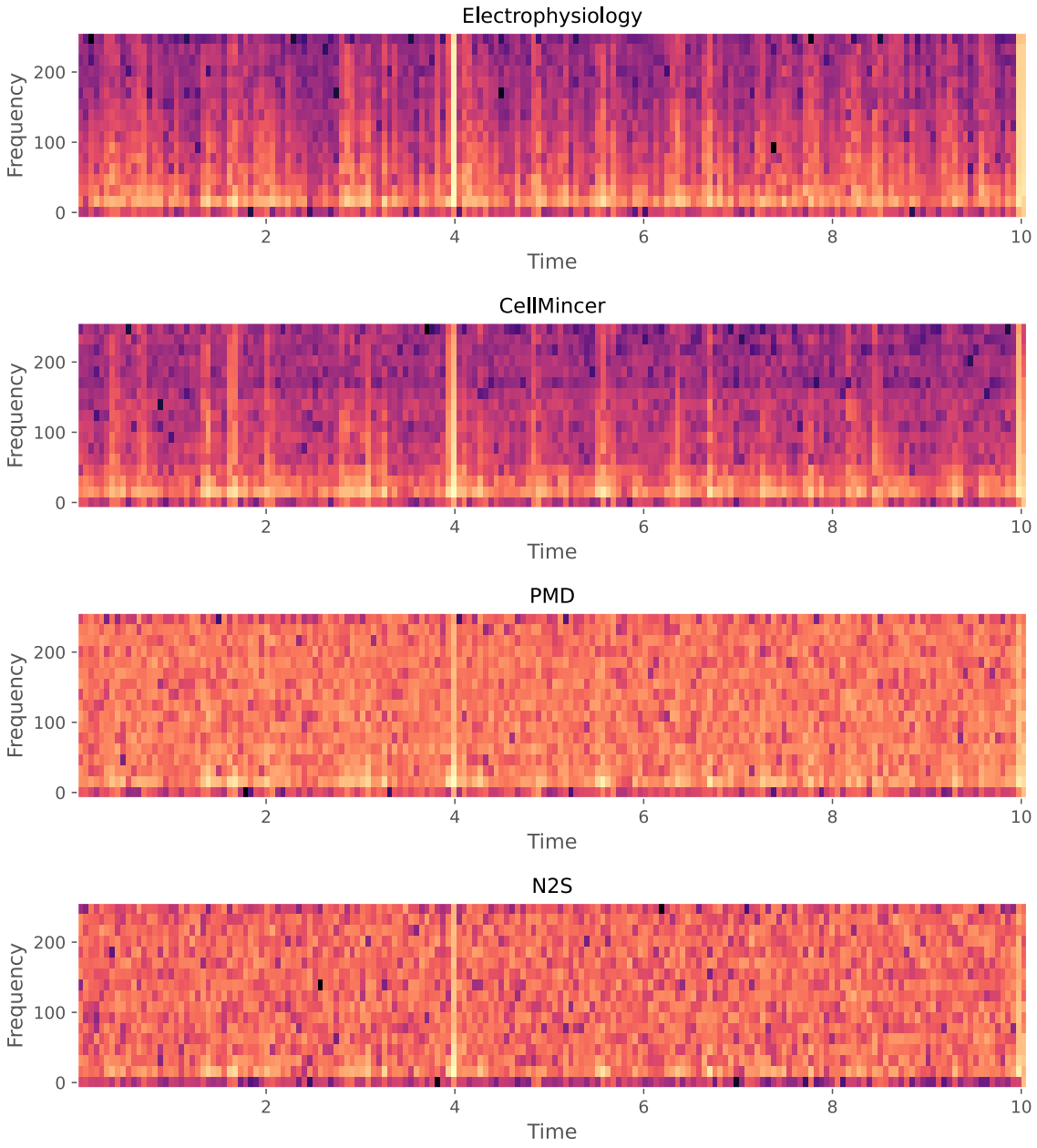

Figure 5-3: Spectrograms of the ephys signal and each of the aligned optical signals on dataset 1.

consecutive ephys peaks is computed, and a true positive is defined as an optical peak that occurs within $0.01w$ of an ephys peak. This adaptive peak-calling tolerance accounts for minor deviations in peak location when a high prominence threshold is set while removing the leniency when all peaks, which can occur every 3-5 timesteps, are counted. This approach can be repeated for various choices of p, producing an F1 curve. The F1 curve for each denoising algorithm and dataset is shown in Figure 5-4.

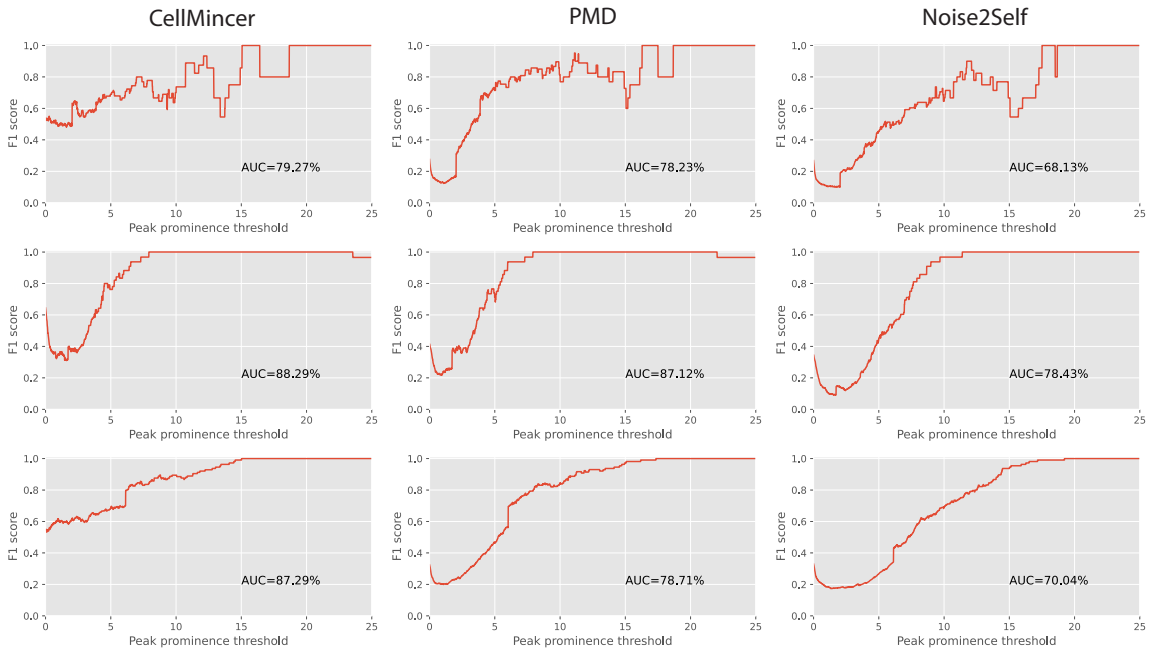

Figure 5-4: Peak-calling F1-score as a function of peak prominence threshold for optical-ephys alignments from three paired datasets each denoised by three algorithms.

For large p that filter out all but the spikes, every denoising algorithm can achieve perfect accuracy. As p falls below 5, roughly corresponding to the inclusion of sub-threshold peaking events, the peak-calling performance of both PMD and N2S quickly degrade, while CellMincer's performance maintains stable. Even as p approaches 0, CellMincer's peak-calling F1-score remains above 0.5, while PMD and N2S achieve a score of 0.2, approximately equal to the performance achieved by randomly sampling peaks at that density. Integrating the F1-score over a suitable range of p produces an AUC ratio that describes the algorithm's cumulative peak-calling performance. An

AUC of 100%, for reference, would indicate perfect peak-calling at every prominence threshold.

One notable limitation of this analysis is that in the high-prominence, sparse peak regime, the F1-score is sensitive to small deviations in certain regions of the subthreshold activity. In the case of dataset 1, it is readily apparent from the data that all denoising algorithms recognize two spikes emerging above the subthreshold activity. Because of slight variations around the third most prominent peak, each algorithm calls the third peak beginning at a different prominence threshold, which produces a significant drop in F1-score when not in agreement with the ephys signal. An example of this sensitivity can be found between $16.5 < p < 18.5$ in CellMincer's F1 curve on dataset 1.

Chapter 6

Conclusions

Through analysis of the experiment results, CellMincer is shown to be the top performing denoising algorithm for denoising voltage imaging recordings. It surpasses training-free methods like PMD in generalizability and expressiveness, as well as deep neural network approaches that lack the capacity to aggregate long-range information. When denoising paired optical-ephys datasets, CellMincer produces optical signals that closely mimic the features of the associated electrophysiological signals. This hints at a powerful potential application in which some biologically significant modes of analysis can rely solely on measuring and denoising an optical electrophysiology signal, saving substantial time and labor that would have been spent patch clamping neurons.

One possible future direction is analyzing the effect of training data SNR on the performance and generalizability of CellMincer. Most typical applications of CellMincer would involve training a model on curated examples of the type of dataset the model is expected to denoise. However, in the case where data is sparse, it may be beneficial to maintain a generalized CellMincer model for finetuning. This leads to the question of how such a generalized model would be best trained. At one extreme where the SNR approaches 0, there is no signal for the model to learn. At the other where the SNR approaches infinity, the model would presumably learn the identity function to minimize L2 loss. This thought experiment suggests that an ideal denoiser would need to be trained at some intermediate SNR.

Another future direction involves computing biologically meaningful features from the denoised optical trace. While this signal cannot recreate the precise curvature of the true electrophysiological signal due to an inferior sampling rate, it may be possible to reliably extract features like the action potential count, the spike half-width, and the interspike interval. Computing features like these may be sufficient to run analyses of electrophysiological activity using optical electrophysiology as a proxy measurement.

On examining the full spatial optical data, it is evident that CellMincer reveals much clearer structures within the FOV than are discernible in the raw data. The simplest such consequence of this improvement is being able to identify far more distinct neurons in the data and with higher confidence. The task of identifying neurons that once had to be left to the trained eye of a lab technician could be automated and performed more reliably on the denoised data.

Bibliography

[1] Joshua Batson and Loic Royer. Noise2self: Blind denoising by self-supervision, 2019.

[2] Michele Claus and Jan van Gemert. Videnn: Deep blind video denoising. *CoRR*, abs/1904.10898, 2019.

[3] Chris Hempel, Christopher Werley, Graham Dempsey, and David Gerber. Targeting neuronal function for cns drug discovery. *Drug Discovery Today: Technologies*, 23, 05 2017.

[4] Simon A. A. Kohl, Bernardino Romera-Paredes, Clemens Meyer, Jeffrey De Fauw, Joseph R. Ledsam, Klaus H. Maier-Hein, S. M. Ali Eslami, Danilo Jimenez Rezende, and Olaf Ronneberger. A probabilistic u-net for segmentation of ambiguous images, 2018.

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

[6] 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.

[7] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data, 2018.

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

[9] S. Lou, Y. Adam, E. N. Weinstein, E. Williams, K. Williams, V. Parot, N. Kavokine, S. Liberles, L. Madisen, H. Zeng, and et al. Genetically targeted all-optical electrophysiology with a transgenic cre-dependent optopatch mouse. *Journal of Neuroscience*, 36(43):11059–11073, 2016.

[10] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention u-net: Learning where to look for the pancreas, 2018.

[11] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

[12] B Sakmann and E Neher. Patch clamp techniques for studying ionic channels in excitable membranes. *Annual Review of Physiology*, 46(1):455–472, 1984.

[13] Massimo Scanziani and Michael Häusser. Electrophysiology in the age of light. *Nature*, 461(7266):930–939, 2009.

[14] Alison S. Walker, Benjamin K. Raliski, Kaveh Karbasi, Patrick Zhang, Kate Sanders, and Evan W. Miller. Optical spike detection and connectivity analysis with a far-red voltage-sensitive fluorophore reveals changes to network connectivity in development and disease. *Frontiers in Neuroscience*, 15, 2021.

[15] Guangfu Wang, Daniel R Wyskiel, Weiguo Yang, Yiqing Wang, Lana C Milbern, Txomin Lalanne, Xiaolong Jiang, Ying Shen, Qian-Quan Sun, J Julius Zhu, and et al. An optogenetics- and imaging-assisted simultaneous multiple patch-clamp recording system for decoding complex neural circuits. *Nature Protocols*, 10(3):397–412, 2015.

[16] Stephen R Williams. Spatial compartmentalization and functional impact of conductance in pyramidal neurons. *Nature Neuroscience*, 7(9):961–967, 2004.

[17] D. M. Witten, R. Tibshirani, and T. Hastie. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics*, 10(3):515–534, 2009.

[18] Yide Zhang, Yinhao Zhu, Evan L. Nichols, Qingfei Wang, Siyuan Zhang, Cody J. Smith, and Scott S. Howard. A poisson-gaussian denoising dataset with real fluorescence microscopy images. *CoRR*, abs/1812.10366, 2018.