Fast, Efficient, and Accurate Neuro-Imaging Denoising via Deep Learning

Shivesh Chaudhary[1], Sihoon Moon[1], Hang Lu[1,2,*]

[1]School of Chemical & Biomolecular Engineering, Georgia Institute of Technology, Atlanta, Georgia, United States of America

[2]Petit Institute for Bioengineering and Bioscience, Georgia Institute of Technology, Atlanta, Georgia, United States of America

[*]Correspondence should be addressed to HL: hang.lu@gatech.edu

Volumetric functional imaging is widely used for recording neuron activities *in vivo*, but there exist tradeoffs between the quality of the extracted calcium traces, imaging speed, and laser power. While deep-learning methods have recently been applied to denoise images, their applications to downstream analyses, such as recovering high-SNR calcium traces, have been limited. Further, these methods require temporally-linked pre-registered data with ultrafast rates. Here, we demonstrate supervised deep-denoising methods to circumvent these tradeoffs for several applications, including whole-brain imaging, large field-of-view imaging in freely moving animals, and recovering complex neurite structures in *C. elegans*. Our framework has 30x smaller memory footprint, and is fast in training and inference (50-70ms); it is highly accurate and generalizable, and further, only small, non-temporally-sequential, independently-acquired training datasets (~500 images) are needed. We envision that the framework will enable faster and long-term imaging experiments necessary to study neuronal mechanisms of many behaviors.

Introduction

Fluorescent functional imaging is ubiquitous in neuroscience research in model systems. The persistent goal is to image wider (more cells and larger areas), deeper, and faster, while enhancing signal-to-noise ratio (SNR). For commonly available functional imaging setups such as point-scanning or spinning disk confocal systems, tradeoffs exist between SNR in images and microscopy parameters such as imaging speed (exposure time), field-of-view (FOV), image resolution, length of recording *etc*. While advancements in genetically encoded calcium and voltage indicators and new microscopic techniques [1–10] with high spatiotemporal resolution and large FOV have relaxed the requirements and driven the development of whole-brain imaging methods in several organisms, tradeoffs still exist in several model organism systems. For instance, in the nematode *C. elegans* [3,4,11–13], SNR in images is limited due to the requirement of small exposure time to capture neural dynamics at 3-6 volumes/s and to prevent motion artifacts. While SNR can be improved by increasing laser power, this leads to photo-bleaching of fluorophores and photo-toxicity, thus limiting the length of recordings, especially during longer timescale behavior. To cover neurons in the whole animal, the FOV is further expanded [9], which necessitates lower magnification and higher laser power, but again exacerbates photo-bleaching.

Recently deep learning enhanced microscopic techniques [14–16] have been developed that significantly overcome the tradeoff between imaging speed and SNR in images. However, these techniques either require expertise in characterizing the microscopy system at hand for generating realistic training data, such as the axial light propagation [14] or they require light-field microscopy setups [15,16] that are not commonly available to all researchers. Further, whether these methods can perform at low laser power conditions that are critical to prevent photobleaching and enable long-term recording of neuron activities is not currently shown. Thus, an orthogonal method to enhance SNR, to circumvent the tradeoffs, would thus be enabling in many studies.

An alternative strategy that has been established recently and has achieved state-of-the-art results to overcome tradeoffs in microscopy is deep-learning based image denoising [17–26]. In these methods, a deep neural network is trained to recover high SNR fluorescent images from low SNR images acquired with low exposure time or low laser power conditions. These include supervised [17–22] and unsupervised [23,24,26,27] methods. Unsupervised methods offer the benefit of training on the data to be denoised itself thus no training data collection is needed. Despite the success of denoising methods, their application on downstream analyses such as high SNR calcium trace extraction from videos has been show in only a few model organisms and

microscopic techniques, all using unsupervised methods. For instance, DeepInterpolation [27] and DeepCAD [26], demonstrate high-quality calcium trace extraction on 2D two-photon imaging data in mice. While impressive, these methods do require large training data set (~100,000 frames for DeepInterpolation and 3,500 frames for DeepCAD); further, pre-registration of the images (or images with minor deviations) before training are required, which also necessitates ultrafast imaging rates. DeepCAD also shows decreasing accuracy for data acquired at slower imaging rates, demonstrating that information in temporally linked images is important for denoising. Practically, these models also have a large memory requirement for training and inference. While new advances in microscopy greatly improve imaging speed and field-of-view [10], generating such large-scale ultrafast recordings for 3D imaging in models organisms is currently not feasible for all researchers with commonly available confocal systems. Additionally, training these methods on calcium activity recordings in moving animals would require a non-trivial pre-registration step and training results would be contingent on the accuracy of registration step.

Compared to unsupervised methods, supervised methods for image processing are expected to achieve higher denoising accuracy and are more generalizable. Currently, supervised methods have not been used for video data denoising and extracting calcium traces. This is likely due to several factors. For instance, if supervised methods are to be trained using temporally linked data, akin to unsupervised methods, custom microscope setups will be needed that can collect low and high SNR video data simultaneously. In contrast, if supervised methods are to be trained with non-temporally linked data, it is not immediately apparent whether the temporal structural features in the dynamical data (as in calcium imaging experiments) can be preserved from independently denoised images. It is also not obvious to what extent the supervised models can be generalized. The wide deployment of these models will also be dependent on several practicalities such as model size, inference speed, and memory requirement on the computation. Here we show that supervised deep denoising can achieve high accuracy in extracting high-SNR calcium traces from noisy videos. Our optimized models are 20-30X smaller in memory footprint, 3-4X faster inference speeds and can be trained with as few as 500 single images that are temporally independent and collected across different samples. With the use of temporally independent data for training, fast imaging rate for training data collection and pre-registered data are not required; further, networks can be trained with a variety of images across animals with different posture configurations, neuron morphology, cell labelling techniques (soma, membrane etc.) and marker (RFP, GCaMP etc.), thus improving the generalizability across conditions and noises.

Optimized deep neural networks for denoising images

To address the challenges of extracting clean calcium traces from noisy calcium imaging videos in common applications, we designed Neuro-Imaging Denoising via Deep Learning (NIDDL), a convolutional neural network (CNN) pipeline that can be trained using only a small set of training non-video data (Fig. 1). The ability to work with independently acquired image training pairs (not from videos) greatly improves the generalizability because of the much relaxed requirements in data acquisition. For instance, to obtain the ground truth for training, images can be acquired for immobilized samples, with little photobleaching (by using independent samples), at different times, and possible across different biological conditions (e.g. different strains). This enables more researchers using a wider set of instruments and in wider biological settings to denoise neural images and recordings. The pipeline takes in independent pairs of noisy (acquired either with low laser-power or short exposure-time) and high SNR image stacks, acquired across samples and reagents (Fig. 1A). Subsequently, efficient denoising convolutional neural networks are trained using the non-video data. In application phase, trained networks are applied to denoise video data by independently denoising each volume in the video. Finally, high quality calcium traces are extracted from denoised video using conventional calcium signal extraction pipeline in *C. elegans* that involves cell segmentation, cell tracking, and signal extraction (Fig. 1A). As an example, microscopic conditions used for whole-brain calcium activity recordings lead to significant loss of SNR in images (Fig. 1B), thus making densely packed nuclei in images barely distinguishable (Supplementary Figure 1). Low SNR in images can significantly reduce the accuracy of intermediary tasks such cell segmentation and tracking, thus making downstream analysis of neuron activity data extremely slow and challenging. We demonstrated that trained networks can significantly recover nuclei structure from these noisy images (Fig. 1B, Supplementary Figure 1).

To achieve a fast, small memory footprint, and data-efficient CNN, we optimized several network hyper-parameters (Online methods – Network Optimization). For instance, starting with vanilla UNet [17,28] and Hourglass architectures [29], we tested several design choices such as kernel size, channel depth, depth of architecture, and presence or absence of residual connections (Supplementary Figure 2,3). Additionally we compared architectures across L2 and L1 loss functions used commonly in image restoration tasks [17,20] (Supplementary Figure 4) and three different training modes (Supplementary Figure 5) including 2D mode, 2.5D mode and 3D mode (Methods section). The optimal models significantly reduce the number of parameters and memory footprint by fixing channel depth across all layers. This allows 1) networks to be

deeper, i.e. have more convolutional blocks compared to CARE using default parameters [17], and 2) use residual connections within each convolutional block that are not present in default UNet. Compared to previously established methods such as CARE [17], RCAN [19], and default UNet and Hourglass architectures, our optimized architectures are 20-30X smaller in memory footprint, have 3-5X faster inference time, are 2-3X faster in training (Fig. 1B,C). We show that for whole-brain imaging applications, model accuracy plateaus at training with 500-600 images (corresponding to 25-40 whole-brain stacks) (Fig. 1D, Supplementary Figure 6), which is much smaller than number of images used for training in recent methods DeepCAD (3,500 frames) and DeepInterpolation (~100,000 frames). Thus, networks can be easily trained in individual labs specific to individual experimental and instrumentation conditions.

We have also tested CNNs trained with L2 or L1 loss and show that they achieve similar accuracy (Supplementary Figure 4), with L1 loss training being more stable across different instances of training. Further, we noticed that L1 loss performs better in RMSE and PSNR metrics whereas L2 loss performs better in SSIM metric. This could be because L1 loss is more suitable to handle the type of noise present in experimental data whereas L2 loss is more suitable to preserve structural information. Finally, we tested three modes of training that differ in 3D spatial context used by networks for denoising (Supplementary Figure 5). These modes include 1) 2D mode where input and output to the networks are 2D images, 2) 2.5D mode where input to the network is a 3D stack consisting of z-planes above and below the image to be denoised and output is the middle denoised 2D image, and 3) full 3D mode where input to the networks is 3D stack and output is also 3D stack. Comparisons showed that training with 2D images, rather than 3D stacks, is sufficient (Supplementary Figure 5), possibly because more training data is needed for 3D mode of training. Practically, 2D images can be acquired easily using commonly available setups thus simplifying the training step. Importantly, these memory-efficient and fast models can be used widely without expensive GPUs. When comparing inference time of models without GPUs, NIDDL achieves an average inference time of 1.25 s whereas CARE and RCAN denoise images in much longer time of 2.67 s and 7.29 s respectively (Fig. 1E).

To characterize the performance of NIDDL, we worked with *C. elegans* strains with whole-brain neuronal labels. We took advantage of microfluidic immobilization of animals to avoid the complex image pre-registration step across image pairs before training the networks, and to acquire data in high-throughput manner [30]. Acquired pairs of non-sequential image data across samples is used as inputs to train the CNN. Trained networks are then applied to noisy video frames independently to recover clean images. Subsequently high SNR calcium traces are

extracted from denoised videos. We show that deep denoising recovers structures in noisy whole-brain images with clear distinction of nuclei (Supplementary Figure 1, Fig. 2, Supplementary Figure 7, Supplementary Video 1), which can greatly improve the nuclei segmentation performance [31], and thus the accuracy of downstream tasks such as cell identity annotation [32]. We bench-marked NIDDL's performance with those from other approaches (Online methods – Comparison against other methods). Representative images show that NIDDL produces cleaner denoised images closer to the ground truth images, while simple denoising methods such as Median and Gaussian filtering, as well as advanced non-deep learning based methods such as NLM and BM3D, suffer from either blurring artifact or not recovering information (Supplementary Figure 7). Quantitatively, the optimized NIDDL model achieve high accuracy on held-out datasets, outperforming traditional denoising methods, non-deep learning based methods such as NLM and BM3D, and deep learning methods such as RCAN[19] (Fig. 2B, Supplementary Figure 8). While the recently published algorithm CARE produces similar accuracy as NIDDL, the advantage of NIDDL is smaller model size and real-time inference time (Fig. 1C), which would be important for applications that would require near real-time feedback, e.g. closed-loop optogenetic interventions.

To test the generalizability of the approach, we trained separate network instances on data collected across a variety of conditions and compared within-condition accuracy with across-condition accuracy. These include two whole-brain imaging strains with different levels of fluorophore expression labelling all cells, three levels of laser powers, and three independent experiments on different days. Models trained on independent experiments and strains are particularly generalizable across conditions (Fig. 3A-D). As an example, denoised images output by networks, when networks were either trained on the same strain or a different strain, visually appear indistinguishable (Fig. 3A, 3B). In both cases, networks significantly recover distinguishable nuclei structure from noisy images (Fig. 3B). When accuracies are characterized, cross-strain model performance also appears similar to that of within-strain models (Fig. 3C). Our results do show that models are sensitive to image-acquisition laser power (Fig. 3E, Supplementary Figure 10). In comparison, models generalize with high degree of accuracy across independent experiments (Fig. 3D, Supplementary Figure 9). In parallel, we conducted an *in silico* experiment to characterize the robustness of the optimized CNNs against noise levels; we generated realistic 3D synthetic data with densely packed nuclei (Online Methods – Synthetic whole-brain data generation) across a range of signal levels (photon counts), corrupted by Poisson shot noise and Gaussian readout noise. We show that NIDDL consistently and efficiently denoise the images, better than traditional methods (Supplementary Figure 11). We hypothesized

that as long as a minimum requirement for SNR is met, NIDDL can produce efficient denoising, and that the corruption of the signal by noise beyond a certain threshold cannot be rescued by denoising. Indeed, this notion is corroborated by the characterizations of the SNR in the actual experiments (Fig. 3F) where the SNR levels across laser powers vary vastly, those across strains vary less, and across independent experiments sessions have similar SNR levels. These results demonstrate that as long as the imaging experiments meet a minimum SNR threashold (~20), NIDDL can efficiently denoise. This points to the advantages of NIDDL, where training data sets can be gathered in a distributed manner and from varied conditions (including from different strains), which would greatly lower the barriers for use in practice.

High SNR calcium trace recovery using NIDDL

While denoising images in itself can improve accuracy of many tasks in whole-brain imaging, including segmentation, tracking, and identification, the critical goal is to extract clean calcium traces. We next denoised a whole-brain video (Online Methods – Calcium imaging data collection) that was held out from the training (Fig. 4A, Supplementary Video 2) and extracted traces (Methods – Denoising and extracting calcium traces). We note that methods used for calcium signal extraction from two-photon recordings of spiking neurons [33–35] differ from standard methods used for *C. elegans* [11,36–38]. The deep denoised video provides much cleaner traces compared to the original noisy video (Fig. 4B) and correlated neuron activity is detectable visually. Since NIDDL is trained using non-video data, denoising each frame of video independently could introduce artifacts in calcium traces. To establish that NIDDL recovered calcium traces do not contain artifacts, we compared the traces extracted from denoised video to traces extracted from high-SNR ground-truth video for the same recording. Denoised traces show the same temporal structure in neuron activity as present in high SNR video thus denoising does not introduce artifacts (Fig. 4B). Furthermore, denoised traces show much lower mean absolute error (Fig. 4C) and higher correlation to the traces from the ground-truth low-noise video (Fig. 4D). This demonstrates that denoising by NIDDL greatly improves SNR in the frames independently. Further and perhaps more importantly, denoising with NIDDL recovers correlational structure among neuron activities (Fig. 4E), crucial for downstream analyses and interpretation such as PCA based latent activity recovery [11] commonly used in whole-brain data analysis pipelines [39]. We further tested the robustness of NIDDL against different noise levels by denoising and extracting traces from semi-synthetic videos across a range of SNR levels (Online method – Semi synthetic video data generation). Deep denoising significantly removes noise from traces (Supplementary Figure 12A, C) and performs better than traditional methods across all SNR

levels (Supplementary Figure 13D). Lastly, we show that NIDDL denoised traces significantly improve the performance of PCA analysis commonly used for analyzing whole-brain recording datasets [11]. Neural activity trajectory in low dimensional space show smooth dynamics in ground-truth video; however, such structure is lost in noisy video (Supplementary Figure 12B). NIDDL denoised video successfully recovers the smooth dynamics (Supplementary Figure 12B). Taken together, these results demonstrate that denoising using NIDDL requires a small set of training data, is forgiving in many experimental constraints, and yet provides excellent performance in accuracy, robustness, and generalizability while using minute inference time potentially enabling on-line feedback manipulations from calcium dynamics.

Next, we sought to demonstrate denoising on large field-of-view (FOV) data acquired at low magnification (Online Methods – Calcium imaging data collection). The advantage of large FOV is to capture more cells simultaneously. The challenges with large FOV recording, however, is low spatial resolution so that each cell corresponds to only a few pixels, and this necessitates higher laser power to boost SNR. Here, we imaged simultaneously many ventral cord (VC) motor neurons in *C. elegans*. To avoid photo-bleaching, we also used low laser power, which results in worse SNR as compared to imaging at 40x (higher NA) [Fig. 5A, Supplementary Figure 13]. We trained NIDDL with temporally independent (i.e. non-video data) pairs of low and high-laser-power images of ventral cord neurons expressing GCaMP. NIDDL was able to remove much of the noise, enabling the detection of cells barely noticeable in noisy images (Fig. 5A, Supplementary Figure 13). Quantitative comparisons show, NIDDL significantly outperforms traditional denoising methods and advanced non-deep learning based methods (Fig. 5A, 5B, Supplementary Figure 14, Supplementary Figure 15) and achieve similar accuracy as CARE and RCAN. Next, we denoised low-SNR videos held out from training and extracted calcium traces from them. Again, NIDDL enables extraction of high-quality calcium traces from noisy videos, making it much easier to detect coordinated neuron activities (Fig. 5C, Supplementary Figure 16, Supplementary Video 3) barely visible in traces extracted from noisy videos. This demonstration suggests that NIDDL is truly enabling tool for large FOV applications where SNR levels in images are very low, and each cell corresponds to only a few pixels in images thus hindering extraction of clean calcium traces. Further NIDDL can avoid photobleaching in large FOV by enabling imaging at low laser power conditions.

While a simple use of the large FOV and deep denoising is to increase the number of cells observed simultaneously and increase the throughput of experiments by enabling imaging multiple animals simultaneously (Fig. 5C), the technique is truly enabling for imaging moving

samples where low exposure time (to reduce blurring type motion artifact) is critical (Fig. 5D). Conventionally for freely moving animals, neural activities are imaged at high magnification, while behavior is tracked with a second low-magnification light path. Here, with large-FOV low-magnification imaging and deep denoising, animals can be tracked directly on the fluorescent images, while deep-denoising can extract clean calcium traces from these videos, with more cells, without compromising the imaging quality. We demonstrate this by imaging motor neurons' along the ventral nerve cord of freely moving animals. Deep denoising by NIDDL significantly removes noise from calcium traces, resulting in clear bouts of neural activities (Fig. 5D, Supplementary Video 4). Ext, we correlated activities of motor neurons to local body curvature of the animal as it roams. Motor neuron activity recovered by NIDDL showed enhanced correlation to animal curvature (Supplementary Figure 17B-D) compared to traces extracted from noisy videos. Thus, NIDDL enables recordings where samples move significantly by enabling imaging using low exposure time conditions. By requiring only low light, this approach will also enable more prevalent longer-term imaging with behavior.

Complex neurite structure recovery with NIDDL

Another application of deep denoising is in imaging subcellular features such as the dendritic processes, which are typically dim and difficult to quantify compared to imaging the soma. Because denoising neurites presents different challenges, we sought to optimize network hyper-parameters specifically for neurites (Supplementary Figure 18) and chose L2 loss for due to slightly better performance. Optimized network recovers structure of neurites from noisy images (Fig. 6A, Supplementary Figure 18, Supplementary Figure 19, Supplementary Video 5) showing distinct processes barely visible in noisy images. Further, NIDDL enables quantitative characterization of neurite morphology as recovered neurite structure significantly improves neurite segmentation performance using simple methods (Online Methods – Neurite segmentation) (Fig. 6D, Supplementary Figure 20). Compared to non-deep learning based methods, NIDDL again performed better on accuracy (Fig. 6B, Supplementary Figure 21) while only using pairs of training images, rather than video data. Further NIDDL achieved similar accuracy to test generalizability across neurite morphology, we tested the performance across two strains labelling neurons with distinct structures (the gentle touch neurons ALM, AVM, and PLM, and the multimodal sensory neuron PVD in *C. elegans*). Models trained only on one strain's data achieved equivalent accuracy across other strain (Fig. 6C, Supplementary Figure 23). We envision NIDDL being applied to study calcium signal distribution in complex morphologies of mechanosensory neurons.

Discussion

In this work, we present an easy-to-train, fast, data-efficient, and generalizable deep-learning framework for denoising calcium activity volumetric recordings. While our method has similarities to recently developed supervised learning methods for restoring images [17,20], applications of supervised methods for extracting calcium traces from volumetric recordings in model organisms have not been shown. Here, we demonstrate the utility of supervised denoising methods for various calcium imaging application in *C. elegans*, and highlight key advantages over previous methods, which make them attractive for researchers to adopt easily. First, we demonstrate that networks trained with temporally independent (non-video) data collected across animals, strains, and imaging conditions can be used to recover high-quality calcium traces from video data, thus providing several experimental simplifications. For instance, ultrafast imaging rates for training data collection are avoided, thus enabling more labs to collect data with commonly available microscopy setups. Additionally, since networks are trained with non-video data, complex pre-registration of images before training is circumvented, making the method suitable for motile animals, such as *C. elegans.*

Second, we demonstrate that networks can be trained with order of magnitude smaller training data (~500 images) compared to previous methods DeepInterpolation and DeepCAD. Temporally sequential data used in these previous methods closely resemble the approach of Noise2Noise [23] where multiple samples of an image with independent noise present in samples are used to recover high SNR image. This is because consecutive images in pre-registered data from ultrafast recordings can be thought of as coming from one sample with independent noise in each image. Due to the need of multiple images of each sample with independent noise, the amount of training data needed is typically very large. In comparison, supervised learning methods, do not rely on multiple images of each sample and do not make assumptions on noises in the data. Thus, supervised methods are advantageous for dynamic data, such as those from (slow) volumetric functional imaging where consecutive frames may not have correlated signals, and those from moving samples where frames are not already registered; furthermore, supervised methods can also be trained with much smaller training data.

Third, we demonstrate the generalizability of networks trained in supervised manner across different strains and experimental sessions. This is possible because the models are trained with pairs of high SNR and low SNR images across a variety of conditions, animals, strains etc., which capture the distribution of SNR levels across experimental conditions that the researchers may expect under typical experimental conditions. Thus, these supervised methods

can achieve higher denoising accuracy compared to unsupervised methods, making them suitable for high quality calcium signal from new experimental recordings without retraining the networks. In contrast, unsupervised methods trained on one functional recording dataset may need to be trained again for every new video. Furthermore, to eliminate complex pre-registration of images in moving animal recordings during training, supervised training only needs images of stationary samples across various conditions, and trained networks can be applied to images in videos independently to recover high SNR traces. Thus, the pipeline is much more accessible to routine use in calcium imaging in a wide range of scenarios, e.g. in sensory behavior, mating behavior, and social behavior.

Finally, we have optimized networks extensively to significantly reduce the memory footprint and inference time compared to previous deep learning methods. Our current models can achieve real-time denoising speeds making them suitable for experiments requiring real-time feedback such as optogenetic perturbations. Our demonstrations on variety of data sets that include high-magnification whole-brain calcium recordings, low-magnification large field-of-view calcium recordings and recovering complex neurite morphology highlight the utility of deep learning based denoising methods. We imagine that our demonstration of deep learning methods for functional imaging denoising in *C. elegans* will inspire newer experiments in other model systems such as hydra [40–42], Drosophila [43–46], and zebrafish [39,47], where long-term whole-brain and functional recordings are needed to find neuronal bases of behaviors that evolve over long time scales [48–50]. NIDDL facilitates such recordings by use of low laser power and shorter exposure time. Combining this technology with microscopy techniques requiring low light dosage [9], e.g. using light-sheet for developing neurons [47], or other microscopy techniques such as virtual refocusing and light-field reconstruction, will enable recordings of longer durations and fast frame rates previously not possible.

Acknowledgements

The authors thank the Zimmer Lab for providing ZIM504 strain and the Hobert Lab for providing NeuroPAL strains. The authors acknowledge the funding support of the U.S. NIH (R01NS096581, and R01NS115484) and the U.S. NSF (1764406 and 1707401) to HL. Some nematode strains used in this work were provided by the Caenorhabditis Genetics Center (CGC), which is funded by the NIH (P40 OD010440), National Center for Research Resources and the International *C. elegans* Knockout Consortium. The authors declare no competing interest.

Methods

Methods are described in Online Methods section.

References

1. Ahrens MB, Li JM, Orger MB, Robson DN, Schier AF, Engert F, et al. Brain-wide neuronal dynamics during motor adaptation in zebrafish. Nature. 2012;5: 471–477. doi:10.1038/nature11057

2. Ahrens MB, Orger MB, Robson DN, Li JM, Keller PJ. Whole-brain functional imaging at cellular resolution using light-sheet microscopy. Nat Methods. 2013;10: 413–420. doi:10.1038/nmeth.2434

3. Schrödel T, Prevedel R, Aumayr K, Zimmer M, Vaziri A. Brain-wide 3D imaging of neuronal activity in Caenorhabditis elegans with sculpted light. Nat Methods. 2013;10: 1013–1020. doi:10.1038/nmeth.2637

4. Prevedel R, Yoon Y-G, Hoggmann M, Pak N, Wetzstein G, Kato S, et al. Simultaneous whole-animal 3D-imaging of neuronal activity using light-field microscopy. Nat Methods. 2015;11: 727–730. doi:10.1038/nmeth.2964.Simultaneous

5. Stirman JN, Smith IT, Kudenov MW, Smith SL. Wide field-of-view, multi-region, two-photon imaging of neuronal activity in the mammalian brain. Nat Biotechnol. 2016;34: 857–862. doi:10.1038/nbt.3594

6. Prevedel R, Verhoef AJ, Pernía-Andrade AJ, Weisenburger S, Huang BS, Nöbauer T, et al. Fast volumetric calcium imaging across multiple cortical layers using sculpted light. Nat Methods. 2016;13: 1021–1028. doi:10.1038/nmeth.4040

7. Skocek O, Nöbauer T, Weilguny L, Martínez Traub F, Xia CN, Molodtsov MI, et al. High-speed volumetric imaging of neuronal activity in freely moving rodents. Nat Methods. 2018;15: 429–432. doi:10.1038/s41592-018-0008-0

8. Weisenburger S, Tejera F, Demas J, Chen B, Manley J, Sparks FT, et al. Volumetric Ca2+ Imaging in the Mouse Brain Using Hybrid Multiplexed Sculpted Light Microscopy. Cell. 2019; doi:https://doi.org/10.1016/j.cell.2019.03.011

9. Voleti V, Patel KB, Li W, Perez Campos C, Bharadwaj S, Yu H, et al. Real-time volumetric microscopy of in vivo dynamics and large-scale samples with SCAPE 2.0. Nat Methods. 2019;16: 1054–1062. doi:10.1038/s41592-019-0579-4

10. Demas J, Manley J, Tejera F, Barber K, Kim H, Traub FM, et al. High-speed, cortex-wide volumetric recording of neuroactivity at cellular resolution using light beads microscopy.

Nat Methods. 2021;18: 1103–1111. doi:10.1038/s41592-021-01239-8

11. Kato S, Kaplan HS, Schrödel T, Skora S, Lindsay TH, Yemini E, et al. Global Brain Dynamics Embed the Motor Command Sequence of Caenorhabditis elegans. Cell. 2015;163: 656–669. doi:10.1016/j.cell.2015.09.034

12. Venkatachalam V, Ji N, Wang X, Clark C, Mitchell JK, Klein M, et al. Pan-neuronal imaging in roaming Caenorhabditis elegans. [Internet]. Proceedings of the National Academy of Sciences of the United States of America. 2016. doi:10.1073/pnas.1507109113

13. Nguyen JP, Shipley FB, Linder AN, Plummer GS, Liu M, Setru SU, et al. Whole-brain calcium imaging with cellular resolution in freely behaving Caenorhabditis elegans. Proc Natl Acad Sci U S A. 2015; 33. doi:10.1073/pnas.1507110112

14. Wu Y, Rivenson Y, Wang H, Luo Y, Ben-David E, Bentolila LA, et al. Three-dimensional virtual refocusing of fluorescence microscopy images using deep learning. Nat Methods. 2019;16: 1323–1331. doi:10.1038/s41592-019-0622-5

15. Wang Z, Zhu L, Zhang H, Li G, Yi C, Li Y, et al. Real-time volumetric reconstruction of biological dynamics with light-field microscopy and deep learning. Nat Methods. 2021;18: 551–556. doi:10.1038/s41592-021-01058-x

16. Wagner N, Beuttenmueller F, Norlin N, Gierten J, Boffi JC, Wittbrodt J, et al. Deep learning-enhanced light-field imaging with continuous validation. Nat Methods. 2021;18: 557–563. doi:10.1038/s41592-021-01136-0

17. Weigert M, Schmidt U, Boothe T, Müller A, Dibrov A, Jain A, et al. Content-aware image restoration: pushing the limits of fluorescence microscopy. Nat Methods. 2018;15: 1090–1097. doi:10.1038/s41592-018-0216-7

18. Wang X, Yu K, Wu S, Gu J, Liu Y, Dong C, et al. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. ECCV Workshops. 2018.

19. Chen J, Sasaki H, Lai H, Su Y, Liu J, Wu Y, et al. Three-dimensional residual channel attention networks denoise and sharpen fluorescence microscopy image volumes. Nat Methods. 2021;18: 678–687. doi:10.1038/s41592-021-01155-x

20. von Chamier L, Laine RF, Jukkala J, Spahn C, Krentzel D, Nehme E, et al. Democratising deep learning for microscopy with ZeroCostDL4Mic. Nat Commun. 2021;12: 2276. doi:10.1038/s41467-021-22518-0

21. Qiao C, Li D, Guo Y, Liu C, Jiang T, Dai Q, et al. Evaluation and development of deep neural networks for image super-resolution in optical microscopy. Nat Methods. 2021;18: 194–202. doi:10.1038/s41592-020-01048-5

22. Zhang H, Fang C, Xie X, Yang Y, Mei W, Jin D, et al. High-throughput, high-resolution deep learning microscopy based on registration-free generative adversarial network. Biomed Opt Express. OSA; 2019;10: 1044–1063. doi:10.1364/BOE.10.001044

23. Lehtinen J, Munkberg J, Hasselgren J, Laine S, Karras T, Aittala M, et al. Noise2Noise: Learning image restoration without clean data. 35th International Conference on Machine Learning, ICML 2018. 2018. pp. 4620–4631.

24. Krull A, Buchholz TO, Jug F. Noise2void-Learning denoising from single noisy images. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2019. pp. 2124–2132. doi:10.1109/CVPR.2019.00223

25. Lecoq J, Oliver M, Siegle JH, Orlova N, Koch C. Removing independent noise in systems neuroscience data using DeepInterpolation. bioRxiv. 2020; 2020.10.15.341602. doi:10.1101/2020.10.15.341602

26. Li X, Zhang G, Wu J, Zhang Y, Zhao Z, Lin X, et al. Reinforcing neuron extraction and spike inference in calcium imaging using deep self-supervised denoising. Nat Methods. 2021; doi:10.1038/s41592-021-01225-0

27. Lecoq J, Oliver M, Siegle JH, Orlova N, Ledochowitsch P, Koch C. Removing independent noise in systems neuroscience data using DeepInterpolation. Nat Methods. 2021;18: 1401–1408. doi:10.1038/s41592-021-01285-2

28. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Miccai. 2015; 234–241. doi:10.1007/978-3-319-24574-4_28

29. Newell A, Yang K, Deng J. Stacked hourglass networks for human pose estimation. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2016. pp. 483–499. doi:10.1007/978-3-319-46484-8_29

30. Lee H, Kim SA, Coakley S, Mugno P, Hammarlund M, Hilliard MA, et al. A multi-channel device for high-density target-selective stimulation and long-term monitoring of cells and subcellular features in C. elegans. Lab Chip. 2014;14: 4513–4522. doi:10.1039/C4LC00789A

31. Toyoshima Y, Tokunaga T, Hirose O, Kanamori M, Teramoto T, Jang MS, et al. Accurate Automatic Detection of Densely Distributed Cell Nuclei in 3D Space. PLoS Comput Biol. 2016;12. doi:10.1371/journal.pcbi.1004970

32. Chaudhary S, Lee SA, Li Y, Patel DS, Lu H. Graphical-model framework for automated annotation of cell identities in dense cellular images. Calabrese RL, editor. Elife. eLife Sciences Publications, Ltd; 2021;10: e60321. doi:10.7554/eLife.60321

33. Pnevmatikakis EA, Soudry D, Gao Y, Machado TA, Merel J, Pfau D, et al. Simultaneous Denoising, Deconvolution, and Demixing of Calcium Imaging Data. Neuron. 2016;89: 299. doi:10.1016/j.neuron.2015.11.037

34. Giovannucci A, Friedrich J, Gunn P, Kalfon J, Brown BL, Koay SA, et al. CaImAn an open source tool for scalable calcium imaging data analysis. Elife. 2019;8. doi:10.7554/eLife.38173

35. Pachitariu M, Stringer C, Dipoppa M, Schröder S, Rossi LF, Dalgleish H, et al. Suite2p: beyond 10,000 neurons with standard two-photon microscopy. bioRxiv. 2016; 061507. doi:10.1101/061507

36. Tokunaga T, Hirose O, Kawaguchi S, Toyoshima Y, Teramoto T, Ikebata H, et al. Automated detection and tracking of many cells by using 4D live-cell imaging data. Bioinformatics. 2014;30: i43-51. doi:10.1093/bioinformatics/btu271

37. Chaudhary S, Lu H. Point-set registration framework with Conditional Random Fields for automatic tracking of neurons in C . elegans whole-brain videos Point-set registration framework with Conditional Random Fields for automatic tracking of neurons in C . elegans whole-brai. Work Worm Neural Inf Process 31st Conf Neural Inf Process Syst (NIPS 2017), Long Beach, CA, USA. 2018; Available: https://www.researchgate.net/profile/Shivesh-Chaudhary/publication/327056521_Point-set_registration_framework_with_Conditional_Random_Fields_for_automatic_tracking_of_neurons_in_C_elegans_whole-brain_videos/links/5b756b1345851546c909d79d/Point-set-registration-framework-with-Conditional-Random-Fields-for-automatic-tracking-of-neurons-in-C-elegans-whole-brain-videos.pdf

38. Nguyen JP, Linder AN, Plummer GS, Shaevitz JW, Leifer AM. Automatically tracking neurons in a moving and deforming brain. 2016; Available: http://arxiv.org/abs/1610.04579

39. Marques JC, Li M, Schaak D, Robson DN, Li JM. Internal state dynamics shape brainwide

activity and foraging behaviour. Nature. 2019; doi:10.1038/s41586-019-1858-z

40. Lovas JR, Yuste R. Ensemble synchronization in the reassembly of Hydra's nervous system. Curr Biol. 2021;31: 3784–3796.e3. doi:https://doi.org/10.1016/j.cub.2021.06.047

41. Dupre C, Yuste R. Non-overlapping Neural Networks in Hydra vulgaris. Curr Biol. 2017;27: 1085–1097. doi:https://doi.org/10.1016/j.cub.2017.02.049

42. Weissbourd B, Momose T, Nair A, Kennedy A, Hunt B, Anderson DJ. Functional modules within a distributed neural network control feeding in a model medusa. bioRxiv. 2021; 2021.02.22.432372. doi:10.1101/2021.02.22.432372

43. Pacheco DA, Thiberge SY, Pnevmatikakis E, Murthy M. Auditory activity is diverse and widespread throughout the central brain of Drosophila. Nat Neurosci. 2021;24: 93–104. doi:10.1038/s41593-020-00743-y

44. Mann K, Gallen CL, Clandinin TR. Whole-Brain Calcium Imaging Reveals an Intrinsic Functional Network in Drosophila. Curr Biol. 2017;27: 2389–2396.e4. doi:10.1016/j.cub.2017.06.076

45. Münch D, Goldschmidt D, Ribeiro C. Distinct internal states interact to shape food choice by modulating sensorimotor processing at global and local scales. bioRxiv. 2021; 2021.05.27.445920. doi:10.1101/2021.05.27.445920

46. Sayin S, De Backer JF, Siju KP, Wosniack ME, Lewis LP, Frisch LM, et al. A Neural Circuit Arbitrates between Persistence and Withdrawal in Hungry Drosophila. Neuron. 2019;104: 544–558.e6. doi:10.1016/j.neuron.2019.07.028

47. Koyama M, Druckmann S, Wan Y, Wei Z, Looger LL, Koyama M, et al. Single-Cell Reconstruction of Emerging Population Activity in an Entire Developing Circuit. Cell. 2019; 1–18. doi:10.1016/j.cell.2019.08.039

48. Ji N, Madan GK, Fabre GI, Dayan A, Baker CM, Kramer TS, et al. A neural circuit for flexible control of persistent behavioral states. Zimmer M, Sengupta P, editors. Elife. eLife Sciences Publications, Ltd; 2021;10: e62889. doi:10.7554/eLife.62889

49. Susoy V, Hung W, Witvliet D, Whitener JE, Wu M, Park CF, et al. Natural sensory context drives diverse brain-wide activity during C. elegans mating. Cell. 2021;184: 5122–5137.e17. doi:https://doi.org/10.1016/j.cell.2021.08.024

50. Wirak GS, Florman J, Alkema MJ, Connor CW, Gabel C V. Age-associated changes to neuronal dynamics involve a loss of inhibitory signaling in C. elegans bioRxiv. 2021; 2021.07.07.451497. doi:10.1101/2021.07.07.451497

51. Stiernagle T. Maintenance of C. elegans. WormBook : the online review of C. elegans biology. 2006. pp. 1–11. doi:10.1895/wormbook.1.101.1

52. Yemini E, Lin A, Nejatbakhsh A, Varol E, Sun R, Mena GE, et al. NeuroPAL: A Multicolor Atlas for Whole-Brain Neuronal Identification in C. elegans. Cell. 2021;184: 272–288.e11. doi:https://doi.org/10.1016/j.cell.2020.12.012

53. Buades A, Coll B, Morel JM. A non-local algorithm for image denoising. Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005. 2005. pp. 60–65. doi:10.1109/CVPR.2005.38

54. Dabov K, Foi A, Katkovnik V, Egiazarian K. Image denoising with block-matching and 3D filtering. Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning. 2006. p. 606414. doi:10.1117/12.643267

55. Dabov K, Foi A, Katkovnik V, Egiazarian K. Image restoration by sparse 3D transform-domain collaborative filtering. Image Processing: Algorithms and Systems VI. 2008. p. 681207. doi:10.1117/12.766355

On-line methods:

C. elegans *culture*

For all experiments, animals were cultured using standard techniques [51]. A detailed list of strains used in this work is provided below.

Name	Genotype	Experiments	Reference
ZIM504	mzmEx199[Punc-31::NLSGCaMP5K; Punc-122::gfp]; lite-1 (xu7)	Used for 1) collecting training data for whole-brain images, 2) test denoising accuracy across imaging sessions (days), 3) test denoising accuracy across SNR in images (laser power settings), and 4) demonstrate high quality neuron activity trace extraction by denoising noisy whole-brain videos	[3,11]
OH16230	otIs672 [rab-3::NLS::GCaMP6s + arrd-4:NLS:::GCaMP6s]. otIs670 provides a healthier alternative to otIs669, performing better in a variety of phenotypic assays. otIs670 [UPN::NLS::TagRFP-T + acr-5::NLS::mTagBFP2::H2B + flp-1::NLS::mTagBFP2::H2B + flp-6::NLS::mTagBFP2::H2B + flp-18::NLS::mTagBFP2::H2B + flp-19::NLS::mTagBFP2::H2B + flp-26::NLS::mTagBFP2::H2B + gcy-18::NLS::mTagBFP2::H2B + ggr-3::NLS::mTagBFP2::H2B + lim-4::NLS::mTagBFP2::H2B + pdfr-1::NLS::mTagBFP2::H2B + srab-20::NLS::mTagBFP2::H2B + unc-25::NLS::mTagBFP2::H2B + cho-1::NLS::CyOFP1::H2B + flp-13::NLS::CyOFP1::H2B + flp-20::NLS::CyOFP1::H2B + gcy-36::NLS::CyOFP1::H2B + gpa-1::NLS::CyOFP1::H2B + nlp-12::NLS::CyOFP1::H2B + nmr-1::NLS::CyOFP1::H2B + ocr-1::NLS::CyOFP1::H2B + osm-9::NLS::CyOFP1::H2B + srh-79::NLS::CyOFP1::H2B + sri-1::NLS::CyOFP1::H2B + srsx-3::NLS::CyOFP1::H2B + unc-8::NLS::CyOFP1::H2B + acr-2::NLS::mNeptune2.5 + ceh-2::NLS::mNeptune2.5 + dat-1::NLS::mNeptune2.5 + dhc-3::NLS::mNeptune2.5 + eat-4::NLS::mNeptune2.5 + flp-3::NLS::mNeptune2.5 + gcy-35::NLS::mNeptune2.5 + glr-1::NLS::mNeptune2.5 + gcy-21::NLS::CyOFP1::H2B::T2A::NLS::mTagBFP2::H2B + klp-6::NLS::mNeptune2.5::T2A::NLS::CyOFP1::H2B + lim-6::NLS::mNeptune2.5::T2A::NLS::CyOFP1::H2B + mbr-1::NLS::mNeptune2.5::T2A::NLS::mTagBFP2::H2B + mec-3::NLS::CyOFP1::H2B::T2A::NLS::mTagBFP2::H2B + odr-1::NLS::mNeptune2.5::T2A::NLS::mTagBFP2::H2B + srab-20::NLS::mNeptune2.5::T2A::NLS::mTagBFP2::H2B] V	Used for 1) collecting training data for whole-brain images, and ventral cord motor neurons, 2) test denoising accuracy across strains, 3) demonstrate high quality neuron activity trace extraction by denoising noisy videos of ventral cord neurons in restrained and freely moving animals	[52]
GT372	aSi31[lox2272 Cbr-unc-119(+) lox2272 + mec-7p::GCaMP7F::ras-2CAAX::SL2::mScarlet-I::ras-2CAAX] II; unc-119(ed3) III	Used for 1) collecting training data for gentle touch neurons' neurite denoising, 2) demonstrate neurite denoising and segmentation	this work
GT366	unc-119(ed3) III; aEx45[ser-2p3b::GCaMP7F::ras-2CAAX::SL2::mScarlet-I::ras-2CAAX + pDSP2(Cbr-unc-119(+)]	Used for 1) collecting training data for harsh touch neuron PVD neurite denoising, 2) demonstrate neurite denoising and segmentation	this work

Training data collection

All imaging was performed using Brucker Opterra II Swept field confocal microscope, with an EMCCD camera. Objective lenses used for each type of data and other imaging parameters are described below.

1. Whole-brain data – Whole brain data was collected using ZIM504 and OH16230 strains. Animals were synchronized to L4 stage and were immobilized in a microfluidic array device to prevent motion. Two 3D stacks (25-30 z planes with 1 µm spacing) were acquired for each animal, one at low laser power and one at the highest laser power setting available in microscope, 10ms exposure time, and Plan Fluor 0.75 NA 40x air objective. Low laser power image specifies the noisy (low SNR) image and high laser power image specifies the clean (high SNR) image. Neural networks were trained to predict high SNR image from low SNR image as described in section Network Training. To quantify prediction generalizability across days, independent datasets were collected for strain ZIM504 on different days using the same strategy. In this case, all datasets were collected at same laser power setting. To quantify prediction generalizability across image SNR levels, additional datasets were acquired using ZIM504 strain at very low and intermediate low laser power levels. To quantify prediction generalizability accuracy across strains with nuclear localized markers, data collected across three strains, ZC392 (nuclear localized RFP expression in glutamatergic cells), OH16230 (nuclear localized GCaMP and CyOFP expression in all neurons) and ZIM504 ((nuclear localized GCaMP5K expression in all neurons) were used. In this case, data for all strains was collected at same laser power settings.
2. Ventral cord neurons data – Images of ventral cord motor neurons were collected using strain OH16230. Animals were synchronized to L4 stage and were immobilized in a microfluidic array device. 3D stacks (40 z planes with 1 um spacing) were collected at 10ms exposure time, using SPlan Fluor ELWD 0.45 NA, 20X air objective. Two stacks were acquired for each animal, one at low laser power and one at the highest laser power setting available in microscope.
3. Neurite data - Images of neurites were collected using strain GT372 and GT366. These strains label different cells with different neurite morphology. GT372 labels gentle touch cells neurites that are sparser compared to harsh touch neuron PVD's neurites labeled in GT366. Animals were synchronized to L4 stage and were immobilized in a microfluidic array device. 3D stacks (40 z planes with 1 um spacing) were collected at 10ms exposure

time, using Plan Fluor 0.75 NA 40x air objective. Two stacks were acquired for each animal, one at low laser power and one at the highest laser power setting available in microscope.

Synthetic whole-brain image data generation

To generate synthetic image data across a range of SNR levels, 3D stacks (128 by 128 by 30 pixels) were generated. Cells were simulated as 3D Gaussian distributions. Cell positions (mean of Gaussian distributions), cell sizes (3D covariance matrices of Gaussian distributions), and cell intensities (max peak of Gaussian distributions) were randomly generated for 60 cells and 3D intensity profiles of all cells were added together to form the image stack. Intensity profile of the resultant image was scaled to a maximum photon count level to specify the peak signal in image. Six photon count levels (20, 50, 100, 200, 500, 1000) were used. This image specified the ground-truth clean image. To generate the corresponding noisy image, two kinds of noises were added to the clean image, photon shot noise (no parameter needed as the noise depends on each pixel's intensity level) and readout noise (normally distributed with mean 0 and 1 variance).

Semi synthetic whole-brain video data generation

To generate semi synthetic 4D (3D + t) calcium imaging video data, 3D stacks (512 by 512 by 30 pixels) were generated for 100 time points. Here again, cells were simulated as 3D Gaussian distributions (as described in Synthetic image data generation section). However, here cell positions (means of Gaussian distributions) were taken from OpenWorm atlas to mimic cell configuration in *C. elegans* head. 130 cells were randomly selected from OpenWorm atlas and positions of only those cells were used for a specific video. This mimics the fact that typically in whole brain recordings, not all cells are imaged due to low fluorophore expression. Further, temporal intensities for each cell were specified using experimental whole-brain recording datasets published previously [11]. A 100 frame window was randomly selected from published data, cell traces within the selected window were extracted from the published data, and each cell in synthetic video was randomly assigned a trace from the selected chunk. Thus realistic experimental calcium traces were present in synthetic video for each cell. Next, intensities of all frames were scaled to a maximum photon count level (using the maximum and minimum pixel intensity across all frames) to specify the peak signal in video. Four photon count levels (100, 200, 500, 1000) were used. This specified the ground-truth clean video. To generate the corresponding noisy video, two kinds of noises were added to each frame, photon shot noise (Poisson noise) and readout noise (normally distributed with mean 0 and 1 variance).

Calcium imaging data collection

We demonstrate deep denoising framework's capability to extract high quality calcium traces from noisy videos for three applications.

1. **High magnification head ganglion functional imaging -** Data was collected using ZIM504 strain. Animals were synchronized to L4 stage and were immobilized in a microfluidic array device. Video (3D + t) stacks (30 z planes with 1 µm spacing, x time points) were acquired at 10ms exposure time, using Plan Fluor 0.75 NA 40x air objective. Noisy (low SNR) frames were acquired at low laser power. For each noisy stack, a ground-truth (high SNR) stack was acquired alternatively. Thus, the two stacks were not completely synchronous, however the time difference between two stacks was very small (~100ms) compared to the dynamics of calcium signal. We compared the traces extracted from noisy video after denoising it with deep neural network with the traces extracted from ground-truth video to ensure that deep denoising does not introduce artifacts in traces.
2. **Low magnification functional imaging of ventral cord neurons in device -** Data was collected using OH16230 strain. Animals were synchronized to L4 stage and were immobilized in a microfluidic array device. Video (3D + t) stacks (40 z planes with 1 µm spacing, x time points) were acquired with 10ms exposure time and SPlan Fluor ELWD 0.45NA 20x air objective. All stacks were acquired at low laser power settings.
3. **Low magnification functional imaging of ventral cord in freely moving animals -** Data was collected using OH16230 strain. Animals were synchronized to L4 stage and were sandwiched between two agar pads on two cover-slips before imaging. 3D stacks (20 z planes with 1 um spacing, x time points) were collected at 10 ms exposure time using Plan Apo Lambda 0.75NA 20x air objective. All stacks were acquired at low laser power settings. At 20x magnification, animals were tracked easily while imaging using z stage x-y controller and kept in the field of view.

Network optimization

We experimented with UNet[17,28], Hourglass[29] and DFCAN[21] architectures given the past success of these networks shown in several biological image analysis tasks such as image restoration, pose prediction, segmentation etc. Architecture details of networks are shown in Supplementary Figure 1, 2. We tested with three hyper-parameters and training settings as described below. In all cases, the network takes as input a noisy (low SNR) image (512 X 512 X d) and through applications of convolutional layers with non-linear activation (ReLU), max-pooling,

up-sampling, feature concatenation or summation generates an output image (512 X 512 X d). Here d is the depth of input and output images. We experimented with different d values as described below. Parameters in networks were trained using stochastic gradient descent with AdamOptimizer (learning rate 0.001) such that the output image is as close as possible (per some loss function) to the corresponding clean (high SNR) image. Training was performed on computing clusters using 16GB or 32 GB GPUs.

1. Architectures – The following convolutional neural network architectures were tested.
    i. UNet – An architecture very similar to conventional UNet architecture was used with 4 down-sampling/max pooling and 4 up-sampling layers. In this case, the first feature map had 32 channels (i.e. 512 X 512 X 32). Depth (number of channels) of feature maps after each maxpooling based down-sampling doubled and depth of feature maps after each up-sampling layer halved. Similar to conventional UNet, long range residual connections were included that concatenate feature maps in down sampling to the feature maps in up-sampling layers.
    ii. UNet_fixed – This architecture is the same as the Unet architecture. However in this case the depth of all feature maps was fixed to 32. Doing so significantly reduced the model size compared to Unet and decreased the network training and inference time without any decrease in accuracy (Fig. 1B, Supplementary Figure 3).
    iii. Hourglass_wores – An architecture very similar to the conventional Hourglass architecture was used. Compared to the Unet architecture, where long range residual connections are a direct concatenation of feature maps in down-sampling layers to feature maps in up-sampling layers, Hourglass architecture has side blocks with trainable parameters (see Supplementary Figure 2) that extract features from down-sampling layers before max-pooling them and adding them to the feature maps in up-sampling layers. This enables the network to extract relevant information from feature maps in down-sampling layers. In our implementation, different from conventional Hourglass architecture, depth of feature maps within each convolutional block was not kept fixed thus it was not possible to include short range residual connection within convolutional block as it requires summation of input and output with same feature depth. Depth of the first feature map was set as 32. Depth of feature maps after each down-sampling layer doubled and depth of feature maps after each up-sampling layer halved

iv. Hourglass_wres – Architecture same as Hourglass_wores was used. However, in this case depth of feature maps in each layer was kept fixed as 32. Further, short range residual connection within each convolutional block was used. Keeping the depth of feature maps fixed to 32 significantly reduced the model size compared to Hourglass_wores, and decreased the network training and inference time without any decrease in accuracy (Fig. 1B, Supplementary Figure 3).

v. DFCAN – Architecture implementation was borrowed from previously published code[21]. We were not able to train the network when input size was 512 X 512 X 1 with published architecture on 32GB memory GPU due to large GPU memory requirements. Thus, we reduced the feature depth in FCAB (feature channel attention blocks) to 32 compared to 64 in published implementation. Further, we trained the network with 128 X 128 X 1 images instead of 512 X 512 X 1 (used for previous methods in this section) to further reduce memory requirements. The number of RCABs (residual channel attention blocks) was kept as 4 and each RCAB contained 4 FCABs, same as published implementation. By design for our task, the output image size is the same as the input image size; thus, scale factor was set to 1. With these settings, we trained several instances of networks with random selection of same of amount of training data used for previous architectures. However, the network did not train well as the output images of the trained networks looked empty across all training instances. This could be due to not enough training data needed by DFCAN. Thus, DFCAN was not considered for further optimization.

All models were trained on the same set of training data and accuracy was tested on a separate held-out dataset consisting of 600 images. Based on comparable or higher accuracy achieved by UNet_fixed and Hourglass_wres architectures compared to other architectures and much smaller memory footprint of these architectures, we selected these architectures for our applications. Small memory footprint also provides the benefit of faster training and faster inference, thus making models user friendly and enabling real time applications.

Apart from architecture type, we also sought to determine if larger filters in convolutional layers can increase accuracy as they can take into account longer range spatial context in images. To do so, we compared the prediction accuracy of the two selected

architectures for two sizes of convolution filters (3 X 3 vs 5 X 5). Since we did not see significant differences in accuracy when using 5 X 5 filters, and models with 3 X 3 filters have smaller memory footprint, we used 3 X 3 filters.

2. Loss function – Two kinds of loss functions have been used previously for image restoration tasks, L2 loss and L1 loss[17]. We asked if one loss function may achieve higher denoising accuracy on some datasets whereas the other may achieve higher accuracy on others. Thus, we trained the networks with both loss functions and compared the accuracy of models across them for all datasets. For high-magnification head ganglion dataset, we found that accuracy of all architectures was comparable across L2 and L1 losses, with L2 loss performing slightly better in SSIM metric (Supplementary Figure 3). Further, L1 loss showed more stable training, as different rounds of training the network from scratch showed lower variability in accuracy. In comparison, L2 loss-trained network showed greater variability in performance across different rounds of training. For harsh and gentle touch mechanosensory neurons' neurite data, L2 loss performed slightly better than L1 loss (Supplementary Figure 18).
3. 2D vs 2.5D vs 3D training – To identify if depth context in 3D image stacks can improve de-noising performance, we tested several models (Supplementary Figure 4).
    i. 2D models that take as input 1 low SNR image (512 X 512 X 1) and output 1 high SNR image (512 X 512 X 1).
    ii. 2.5D models that take as input a noisy 2D image and d z-planes above it and below it (512 X 512 X (2d + 1)) and outputs 1 high SNR image (512 X 512 X 1). Thus the network uses contextual information in z planes above and below the image to be de-noised. The output of the network corresponds to the center z-plane of the input, i.e. the loss is minimized with respect to the center z plane. We tested two values of d with $d = 1$, and $d = 2$. Higher values of d increases the memory footprint of training.
    iii. 3D models that take as input a 3D image stack consisting of d z-planes and outputs a 3D stack consisting of d z-planes. Thus, all z-planes in the 3D input stack are de-noised simultaneously. Here again we tested two values of d with $d = 1$, and $d = 2$.

All models were trained on the same set of training data, and accuracy was tested on a separate held-out but same for all dataset consisting of 600 images. Across these models we found that 2D models performed best. In principle, taking contextual information into

consideration could improve performances. Our observation can be explained by the following. For 2.5D models, noise in z-planes around the center z plane confused the model to focus on denoising center z-plane; further, for 3D models we had to reduce the batch size while training due to memory constraints, which could reduce training performance.

Denoising and extracting calcium traces

Whole-brain video – Low SNR video collected at low laser power was first de-noised using a network trained on whole-brain image dataset. The trained network takes as input individual noisy z planes (512 X 512 X 1) of 3D image stacks in the video and outputs high SNR z planes (512 X 512 X 1), which were subsequently combined to form the de-noised video. To obtain activity traces, nuclei in ground-truth video were first segmented using a Gaussian mixture based segmentation method. Segmented nuclei were tracked across frames using an automated tracking algorithm. Generated tracks of cells were manually inspected and tracks for cells with minor tracking errors were semi-manually corrected. Single pixel activity traces were extracted using the centers of the tracked segmented masks. The same segmented masks were used to extract activity traces from the noisy video and the de-noised video as well to get consistent activity traces across videos and avoid any comparison artifacts due to differences in cell segmentation procedures across videos.

Ventral cord motor neurons in device – The process is same as that to extracting traces from whole-brain videos except that here we de-noise maximum projection images of 3D stacks in the video instead of whole 3D stacks as in whole-brain video denoising case. Thus, in this case the trained network takes as input a maximum projection image of a noisy stack (512 X 512 X 40 converted to 512 X 512 X 1) in the video and outputs high SNR maximum projection stack (512 X 512 X 1). Neuron activity traces were extracted from the maximum projection denoised output.

Ventral cord motor neurons in freely moving animal – Here again we de-noise maximum projection images of 3D stacks in the video instead of whole 3D stacks. The trained network takes as input a maximum projection image of a noisy stack (512 X 512 X 20 converted to 512 X 512 X 1) in the video; and outputs high SNR maximum projection stack (512 X 512 X 1). Neuron activity traces were extracted from the maximum projection de-noised output. To do so, cells in maximum projection images were tracked manually using ManualTracking plugin in Fiji. Subsequently, single pixel activity traces were extracted from both noisy and de-noised videos using track centers.

Accuracy quantification

1. Image denoising accuracy – We quantify image denoising accuracy using 3 metrics – root mean square error (RMSE), peak signal to noise ratio (PSNR) and Structural Similarity Index (SSIM). For each of these metrics, high SNR (ground-truth) image was taken as the reference, and corresponding low SNR (noisy) and network output (denoised) images were compared to the reference. Since maximum intensity value or dynamic range of low SNR (noisy) images is much lower than those in high SNR (clean) images, we normalized intensity values in all images first before calculating the accuracy metrics to prevent arbitrary inflation of errors. Same methodology was used for network optimization and accuracy analysis across all datasets including high magnification whole-brain dataset, low magnification ventral cord imaging, and high magnification neurite dataset).
2. Activity trace from experimental whole-brain video – Single pixel neuron activity traces were extracted from the noisy video, ground-truth video and deep denoised video (see Online Methods – Calcium imaging data collection for video acquisition details and Online Methods – Denoising and extracting calcium traces for activity extraction details). Accuracy was quantified by
    a. Comparing MAE (mean absolute error) of traces extracted from noisy and denoised videos to the traces extracted from ground-truth video.

$$MAE_{noisy} = \frac{1}{T}\sum_{t=1}^{T} |y_{noisy,t} - y_{gt,t}|$$

$$MAE_{denoised} = \frac{1}{T}\sum_{t=1}^{T} |y_{denoised,t} - y_{gt,t}|$$

    b. Comparing Pearson correlation coefficient of neuron activity traces extracted from noisy and denoised videos to the traces extracted from the ground-truth video.

Neuron activity - curvature correlation in freely moving animal

To calculate the curvature of the body as *C. elegans* moves, a 4th degree polynomial was fitted to the coordinates of tracked ventral cord neurons to get ventral cord backbone. Since some cells go out of field of view during animal motion, cells that were consistently present across all frames were used to extract a backbone chunk and curvature analysis was performed using this backbone chuck only. The backbone chunk was divided into 100 segments (sampled at 100 points) and tangent angles to the backbone were calculated at these points. Neuron activity traces

were cross-correlated to tangent angles at all points (shown in heat maps in Supplementary Figure 17D). To quantify improvement in neuron activity-curvature cross-correlation in deep denoised videos, cell traces were cross-correlated to local tangent angles i.e. tangent angles to the backbone at cell's location, and maximum absolute value of the cross-correlation across cells was compared when activity traces were extracted from noisy videos or denoised videos.

Neurite segmentation

Harsh touch neuron PVD's neurites were segmented in noisy and deep denoised images using custom script in MATLAB. The custom script included basic operations with functionalities available in MATLAB - 1) image was sharpened 2) binarized with adaptive thresholding, 3) morphologically eroded to remove segmented noise 4) small holes were filled in image complement, and 5) structures smaller than fixed pixel size were removed.

Comparisons against other methods

We compared the denoising performance of our optimized architectures with several methods across three accuracy metrics; RMSE, SSIM, PSNR. The methods included traditional methods such as Median Filtering and Gaussian Filtering, advanced non-deep learning based methods such as NLM [53], BM3D [54,55], and deep learning based methods such as CARE [17], and RCAN [19]. Below we provide implementation details of these methods. Median Filtering was implemented using default MATLAB function. Three window sizes (3, 5, and 7) for filtering were tried for each dataset and results were reported for best performing window size. Gaussian Filtering was implemented using default MATLAB function. Three kernel sizes or standard deviation values (1, 3, and 5) were tried for each dataset and results were reported for best performing window size. NLM method was implemented using default MATLAB function. No parameters were set for NLM method as it automatically estimates the degree of smoothing based on standard deviation of noise in the image. BM3D method was implemented using MATLAB implementation available here https://webpages.tuni.fi/foi/GCF-BM3D/. Four different values of noise standard deviation were tried (0.05, 0.1, 0.2, 0.5) and results were reported for best performing value for each data set. CARE was implemented using the code provided at https://github.com/CSBDeep/CSBDeep. Default parameters provided in code were used for training except unet_n_depth was set as 4 to be comparable to vanilla UNet architecture that we tried. RCAN was implemented using code provided at https://github.com/AiviaCommunity/3D-RCAN. Default parameters set in code were used for training.

Inference runtime comparisons - system configuration

To compare inference runtime across various deep learning methods, the following system configurations were used.

1. GPU – Quadro M4000, memoryClockRate(GHz): 0.7725, compute capability: 5.2, totalMemory: 8.00 GiB
2. CPU – Intel® Xeon® CPU E5-1620 v4 @ 3.50GHz, RAM: 32 GB, 64-bit Operating System, x64-based processor

Code availability

Code with example datasets is available at https://github.com/shiveshc/whole-brain_DeepDenoising. Instructions on how to run code on sample datasets and train on new datasets are available in the same repository.

Supplement References

21. Stiernagle T. Maintenance of C. elegans. WormBook : the online review of C. elegans biology. 2006. pp. 1–11. doi:10.1895/wormbook.1.101.1

22. Yemini E, Lin A, Nejatbakhsh A, Varol E, Sun R, Mena GE, et al. NeuroPAL: A Multicolor Atlas for Whole-Brain Neuronal Identification in C. elegans. Cell. 2021;184: 272-288.e11. doi:https://doi.org/10.1016/j.cell.2020.12.012

23. Qiao C, Li D, Guo Y, Liu C, Jiang T, Dai Q, et al. Evaluation and development of deep neural networks for image super-resolution in optical microscopy. Nat Methods. 2021;18: 194–202. doi:10.1038/s41592-020-01048-5

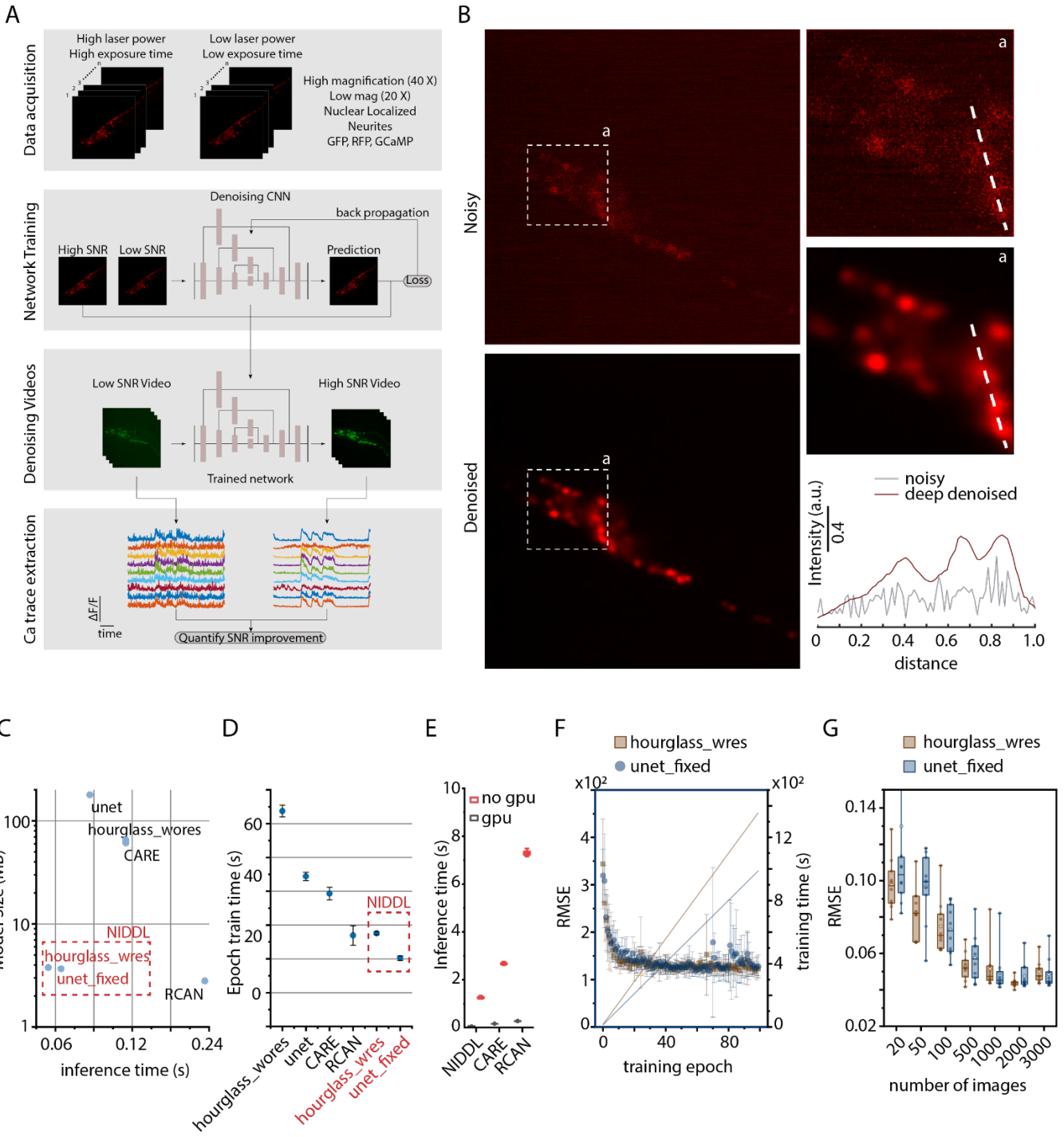


Figure 1. Neuro-Image Denoising with Deep Learning (NIDDL) framework.

A) Overview of the SL framework to recover high SNR from a variety of imaging conditions.

B) An example noisy image (1 zplane from 3D stack) acquired at low laser power and corresponding deep denoised image generated by trained network. Inset 'a' highlights nuclei are difficult to distinguish in

noisy image but can be easily distinguished in denoised image. Right bottom panel shows intensity along dotted lines in noisy and deep denoised images. Data comes from strain OH16230.

C) The optimized neural network architectures 'unet_fixed' and 'hourglass_wres' have 20-30X lower model size (3.77 MB and 3.66 MB) and 3-4X faster inference time (average 48.9 ms and 68.7 ms per 512 X 512 image calculated across 600 images) compared to CARE, RCAN, and non-optimized UNet, Hourglass. Architectures highlighted in red correspond to NIDDL.

D) Per epoch training time comparison across neural network architecture variants with batch size of 50 and epoch training size of 1000 images. Each dot corresponds to average epoch train time across 100 epochs for each instance of trained networks. Error bars indicate standard deviation across 5-10 instances of training with random subset of total data used for training each instance. Architectures highlighted in red correspond to NIDDL.

E) Denoising time comparison of deep learning methods when inference is performed using GPU and without GPU. (n = 50-600 images). Box indicates 25th and 75th percentile, whiskers indicate 5th and 95th percentile

F) Training curves for the optimized neural network architectures and cumulative epoch training time with batch size of 50 images and epoch training size of 1000 images. The optimized architectures train faster (all within 400-800 s) compared to other methods. Error bars in training curve correspond to standard deviation across 10 instances of training.

G) Accuracy vs training data size trade-off for optimized architectures. Each dot corresponds to mean RMSE accuracy on 600 test images for one instance of trained network. In total, 10 instances were trained for each condition with random condition specific subset of total data used for training each instance. RMSE accuracy plateaus above 500 images for both architectures. Data comes from strain ZIM504. Box indicates 25th and 75th percentile, whiskers indicate 5th and 95th percentile.

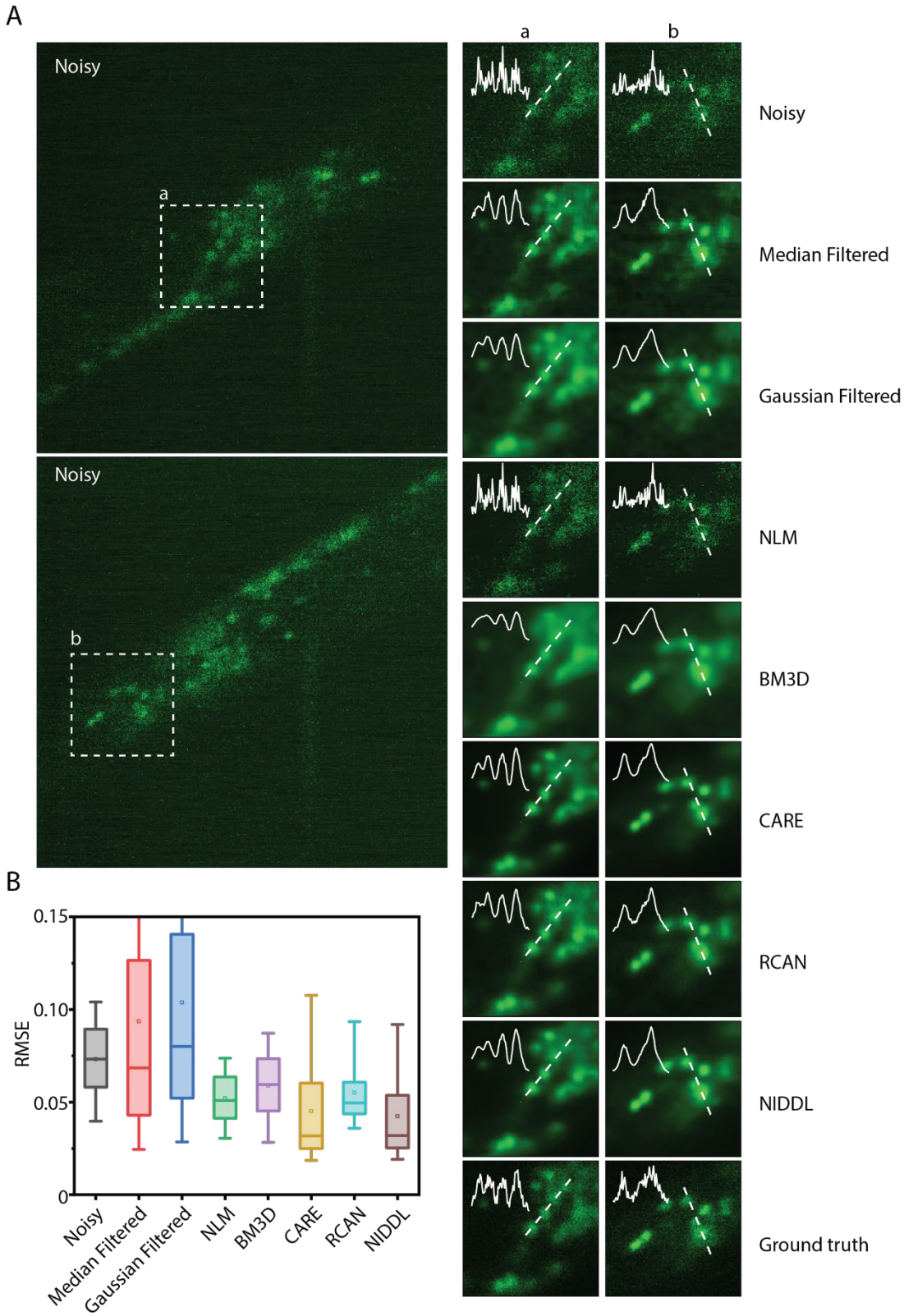


Figure 2. NIDDL denoises whole-brain images in C. elegans.

A) Left – Two example noisy images (single z planes) from noisy whole-brain image stacks (acquired at low laser power). Right – corresponding denoised output generated by different methods. Cell nuclei are labelled with nuclear localized GCaMP5K. Data comes from strain ZIM504. Inset shows intensity profile along the dotted line.

B) Comparison of RMSE to ground-truth high SNR image across noisy images, and denoised images output by various methods including Median filtered, Gaussian filtered, NLM, BM3D, CARE, RCAN and NIDDL (n = 600 images). Boxes indicate 25th and 75th percentile, whiskers indicate 5th and 95th percentile of data.

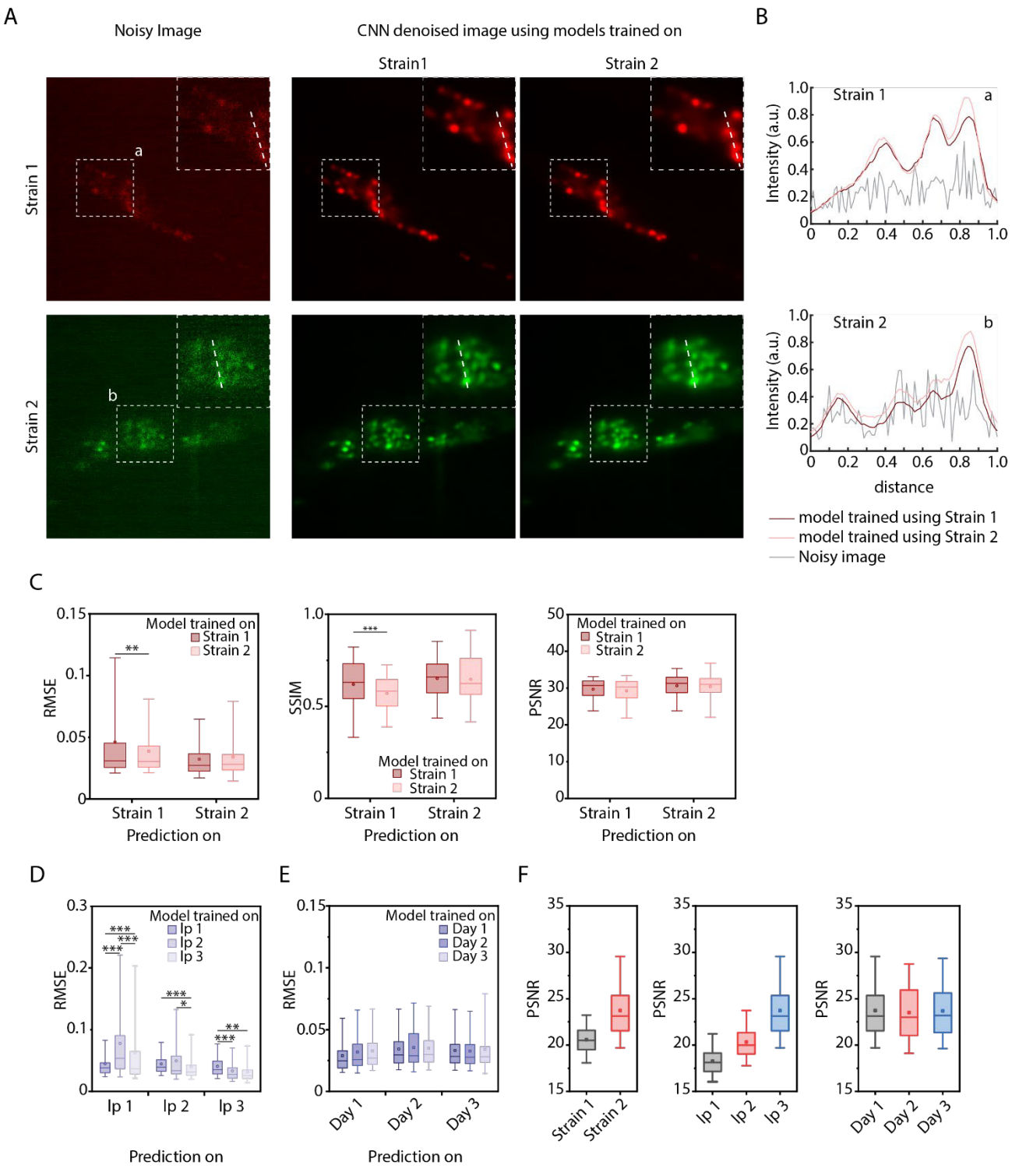


Figure 3. NIDDL performance generalizes across strains and experiments

A) Denoising accuracy quantification on images from 2 different strains with different cell labelling markers when model is trained with specific strain's data only. Data comes from OH16230 (Strain – 1), and ZIM504 (Strain – 2). A) Left – example noisy images from 2 strains, right – corresponding denoised image outputs by 2 different networks trained on specific strain's data.

B) Intensity profiles along the dotted lines shown in insets in A for noisy images and denoised images output by 3 different networks.

*C) Comparison of within strain denoising performance and across strain denoising performance. Left – RMSE accuracy, Middle - SSIM accuracy, and Right – PSNR accuracy on noisy images from 2 strains when networks are trained on specific strain's data. (n = 100, 3,006, 1,403, 50 images for 4 conditions, ***p<0.001, Bonferroni paired comparison test). Boxes indicate 25th and 75th percentile. Whiskers indicate 5th and 95th percentile of data.*

*D) Denoising accuracy on images acquired with 3 different laser power settings (lp1 – extremely low laser power, lp2 – very low laser power, lp3 – low laser power) when model is trained with specific laser power setting's data only. Data comes from strain ZIM504. (n = 88, 2,728, 2,728, 2,262, 100, 2,262, 2,806, 2,806, 226 images for 9 conditions, ***p<0.001, **p<0.01, *p<0.05 Bonferroni paired comparison test). Boxes indicate 25th and 75th percentile. Whiskers indicate 5th and 95th percentile of data.*

E) Denoising accuracy on images from 3 different days' imaging sessions when model is trained with specific day's data only. Data comes from ZIM504. (n = 40, 1,043, 1,043, 1,532, 50, 1,532, 1,403, 1,403, 50 images for 9 conditions). Boxes indicate 25th and 75th percentile. Whiskers indicate 5th and 95th percentile of data.

F) Comparisons of SNR levels in noisy images across conditions. Left – across different strains (n = 3,006, 1,403 images), middle – across different laser powers (n = 1,403, 1,364, 1,101 images), right – across different days (n = 1,403, 1,532, 1,043 images). Boxes indicate 25th and 75th percentile. Whiskers indicate 5th and 95th percentile of data.

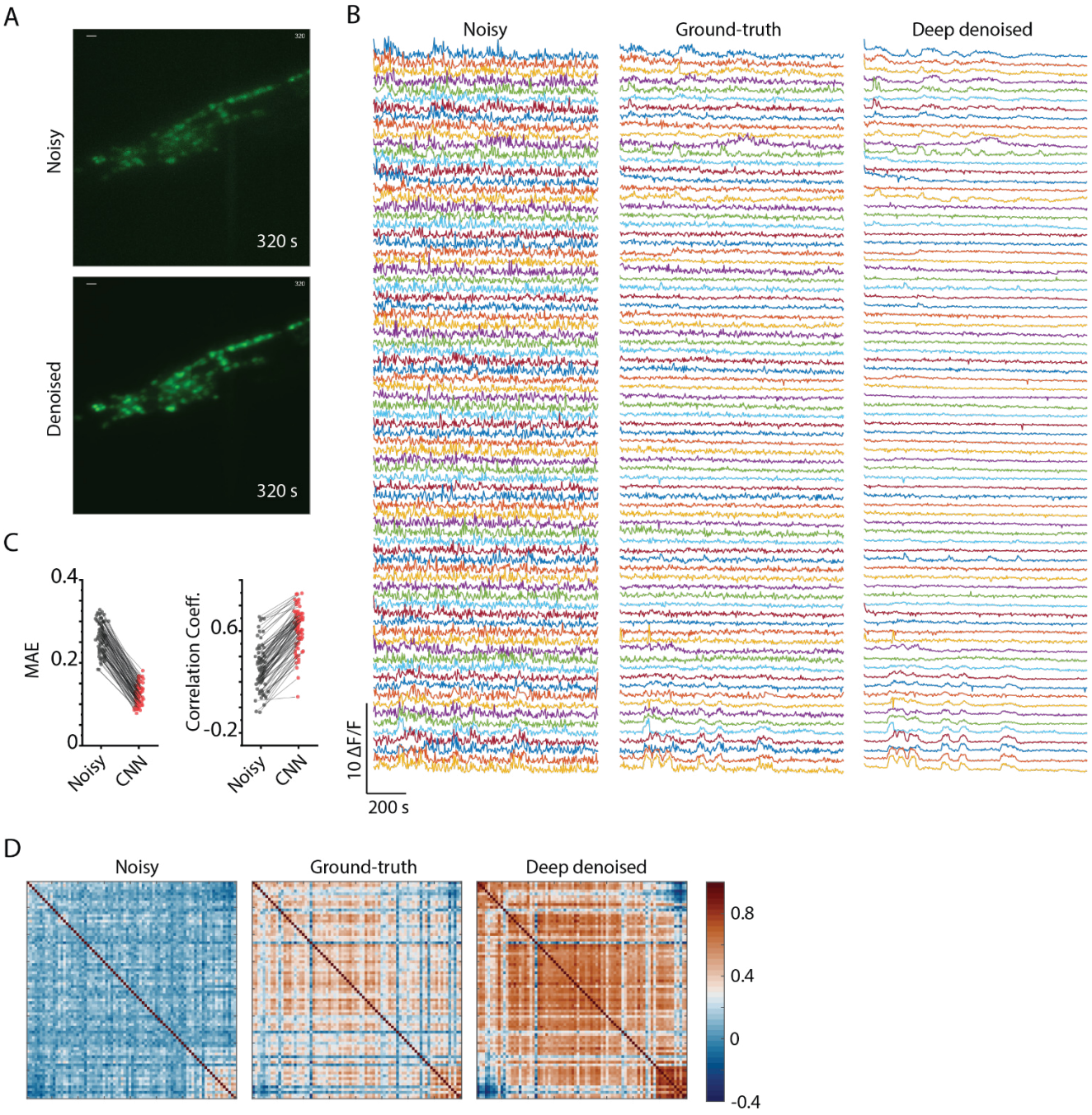


Figure 4. High quality whole-brain Ca trace recovery using NIDDL.

A) Top – max projection of an example image stack from a noisy whole-brain video recording (acquired at low laser power). Bottom – corresponding deep denoised output. Cell nuclei are labelled with nuclear localized GCaMP5K. Data comes from strain ZIM504.

B) Neuron activity traces extracted from the noisy video (shown in A), high SNR ground-truth video for the same recording (acquired at high laser power), and deep denoised video output by network trained only on separate image data.

C) Cell-wise comparison of mean absolute errors (MAE) (Left) and Pearson correlation coefficients (Right) of traces extracted from noisy video and denoised video to corresponding traces extracted ground-truth video.

D) Pairwise Pearson correlation among neuron activity traces extracted from noisy video, ground-truth video, and deep denoised video.

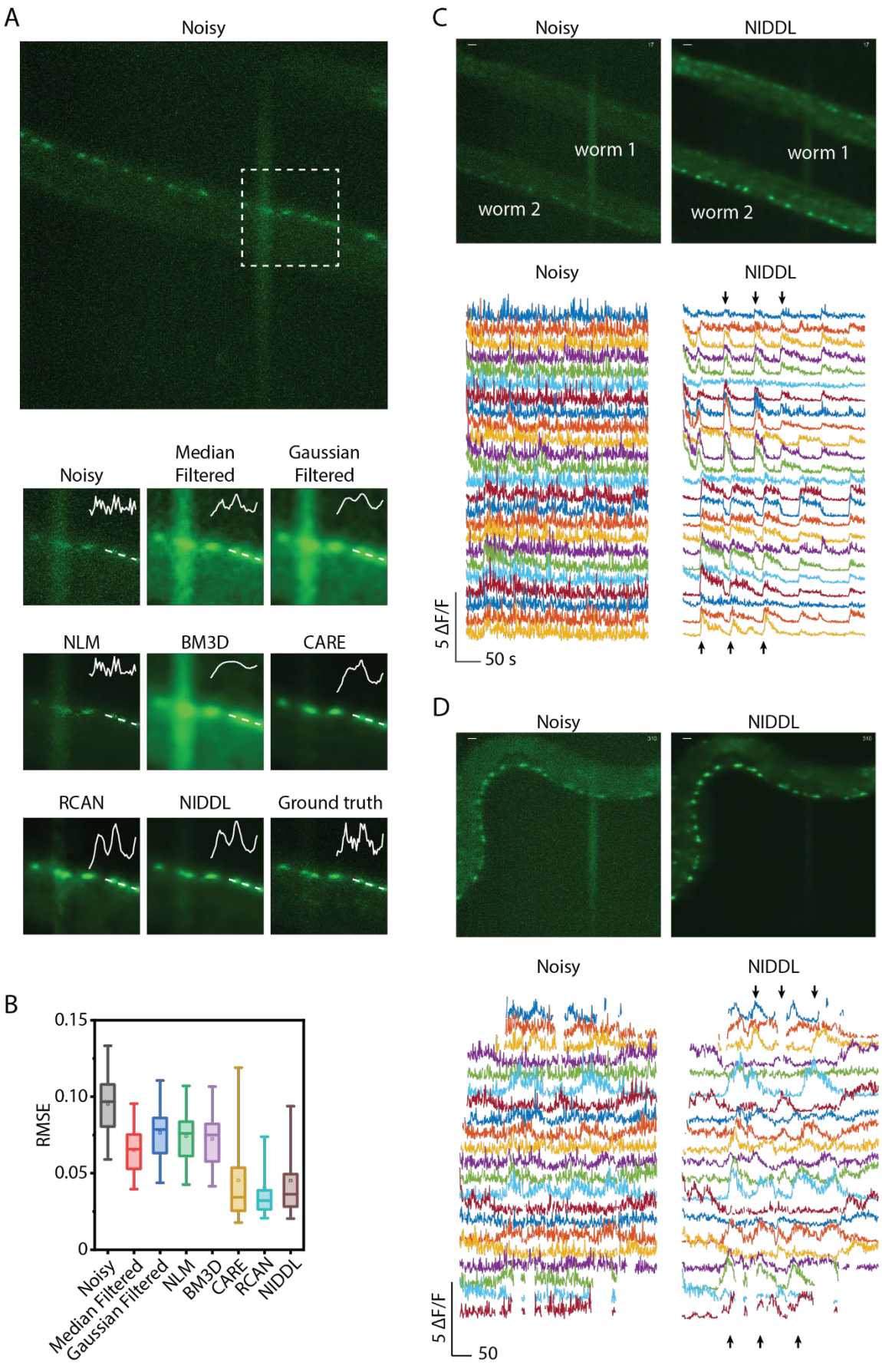


Figure 5. High quality Ca trace recovery in large field-of-view imaging of spatially distributed motor neurons.

A) Top - max projection of a large FOV noisy image stack (acquired at low laser power, 20X magnification) showing motor neurons in ventral nerve cord. Bottom – denoised outputs generated by various methods shown for square box marked in the top image. Inset shows intensity profile along the dotted line. Cell nuclei are labelled with nuclear localized GCaMP6s. Data comes from strain OH16230.

B) B) Comparison of RMSE to ground-truth high SNR image across noisy images, and denoised images output by various methods including Median filtered, Gaussian filtered, NLM, BM3D, CARE, RCAN and NIDDL (n = 60-217 images). Images were collected in-device using strain OH16230. Box indicates 25^{th} and 75^{th} percentile. Whiskers indicate 5^{th} and 95^{th} percentile.

C) Top left – maximum projection of an image stack from a noisy large FOV video recording (acquired at low laser power, 20X magnification) showing motor neurons in ventral cord of two animals restrained in microfluidic device. Top right – corresponding denoised output by NIDDL. Bottom left – single pixel neuron activity traces extracted from the noisy video for worm 2 labeled in images above. Bottom right – corresponding single pixel neuron activity traces extracted from the deep denoised video (arrows indicate coordinated activities). Data comes from OH16230 strain.

D) Top left – max projection of an image stack from a noisy large FOV video recording (acquired at low laser power, 20X magnification) showing motor neurons in ventral cord of a freely moving animal. Top right – corresponding deep denoised output. Bottom left – single pixel neuron activity traces extracted from the noisy video. Bottom right – corresponding single pixel neuron activity traces extracted from the deep denoised video (arrows indicate coordinated activities). Data comes from OH16230 strain.

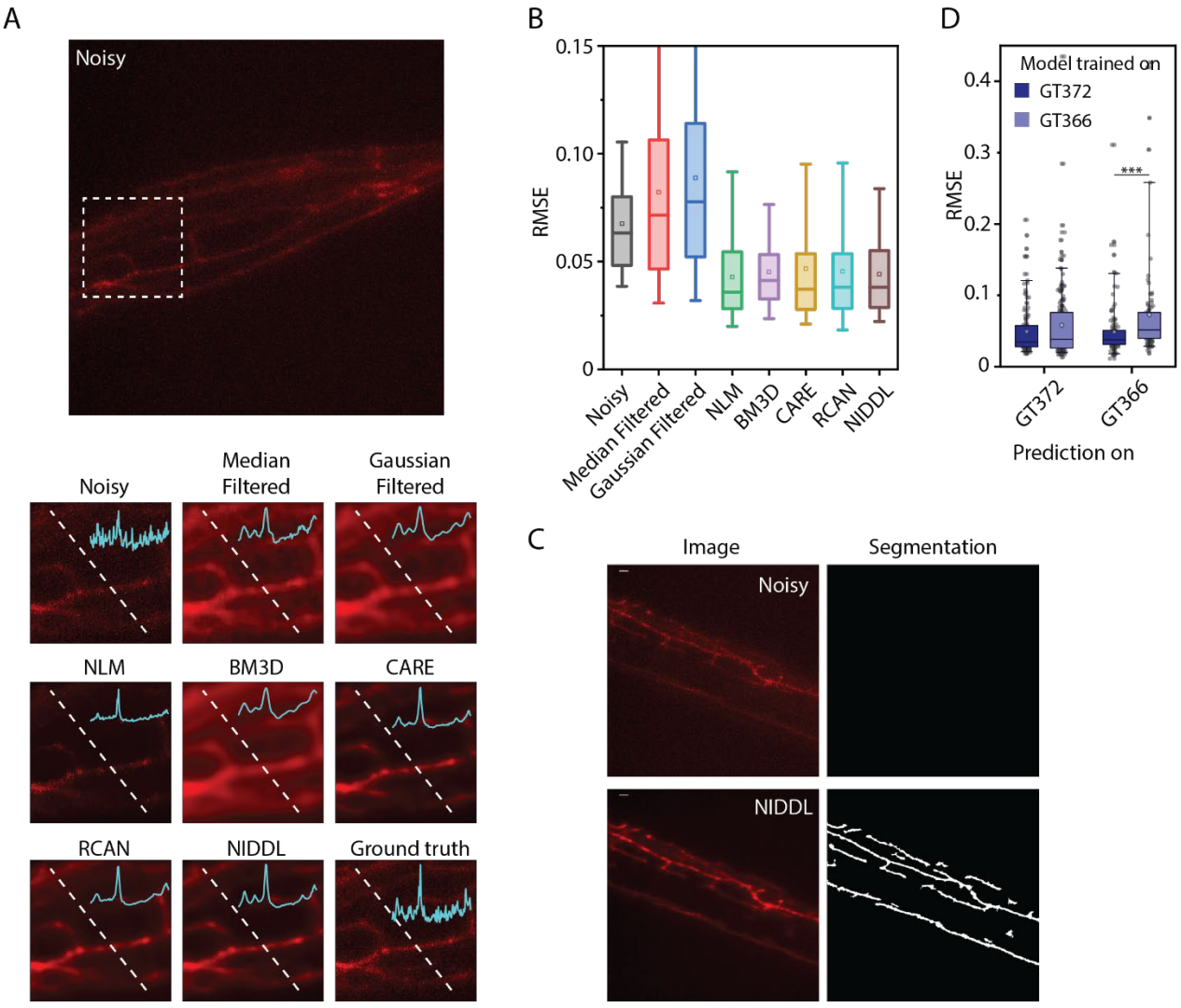


Figure 6. NIDDL recovers complex neurite morphology of mechanosensory neurons

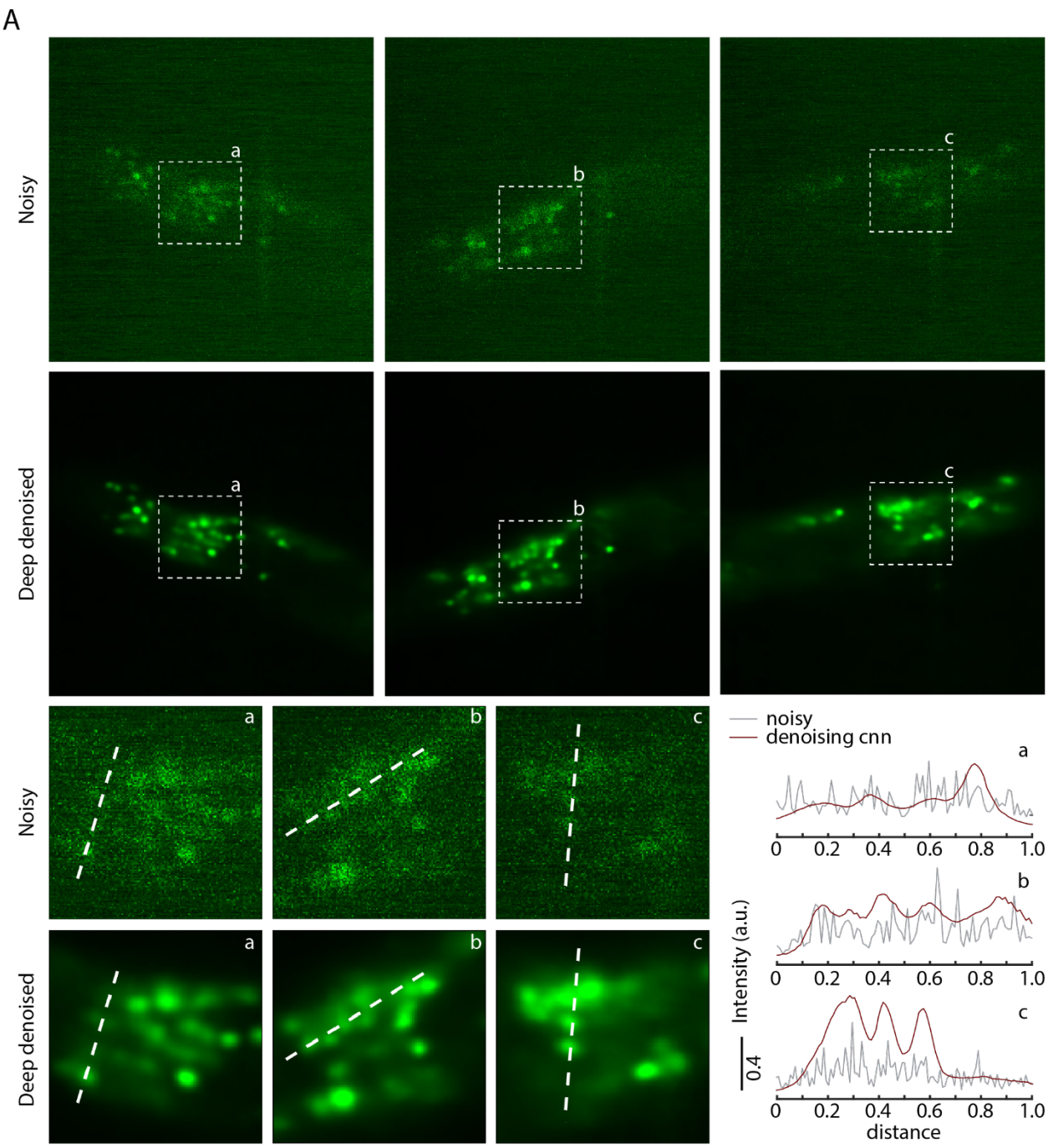
A) Top – max projection of a noisy image stack (acquired at low laser power) showing neurites of harsh touch mechanosensory neuron PVD labeled with mScarlet. Data comes from strain GT366. Bottom – denoised outputs generated by various methods shown for dotted box in top image. Cyan trace in inset denotes pixel intensities along the dotted line.

B) Comparison of RMSE accuracy across noisy images, and denoised images output by various methods. Data comes from strains GT372 and GT366. (n = 86-443 images). Boxes indicate 25th and 75th percentile, whiskers indicate 5th and 95th percentile of the data.

C) NIDDL denoising of images facilitate neurite segmentation. Top - example noisy image showing harsh touch mechanosensory neuron PVD's neurites, no regions are detected in noisy images with simple morphological operations (see Methods – Neurite segmentation). Bottom - corresponding NIDDL denoised output and segmented neurites in denoised image. Data comes from strain GT366.

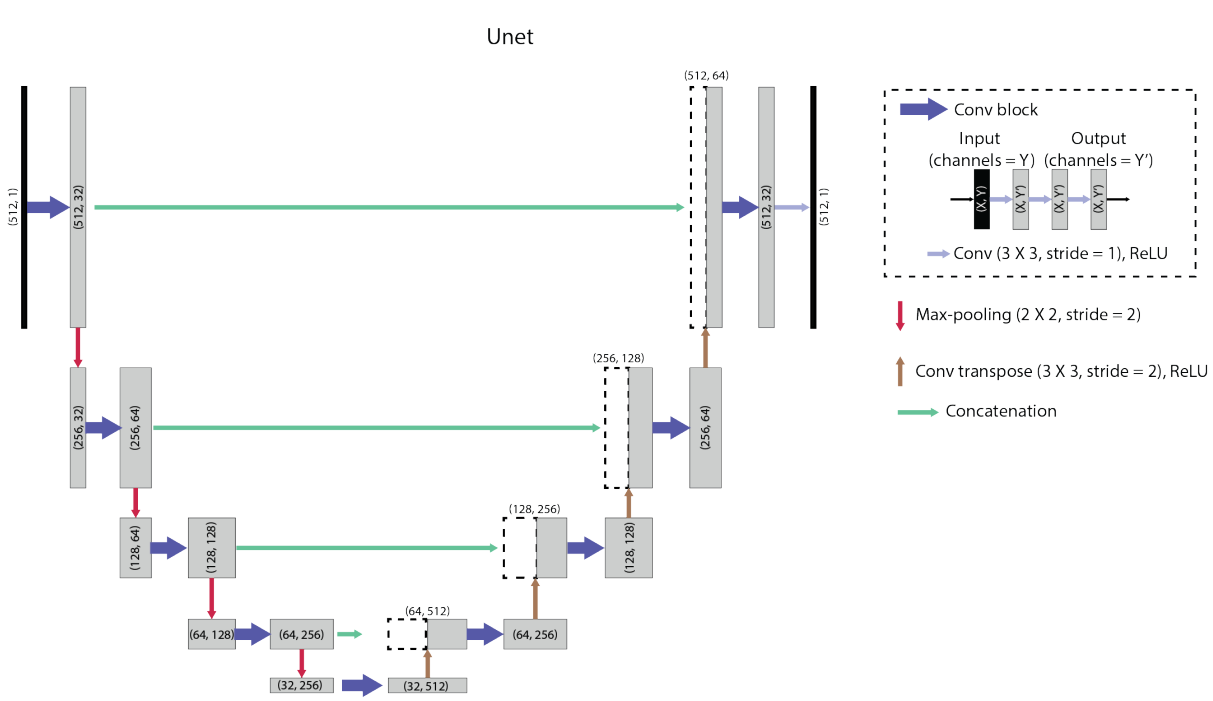
*D) Deep denoising RMSE accuracy comparison on noisy images from 2 different strains, GT372 and GT366 that label neurites of gentle touch and harsh touch mechanosensory neurons respectively, when models are trained on specific strain's data. (n = 129, 203, 118, 97 images for 4 conditions, ***p<0.001,*

**p<0.05, Bonferroni paired comparison. Boxes indicate 25th and 75th percentile, whiskers indicate 5th and 95th percentile of the data.*

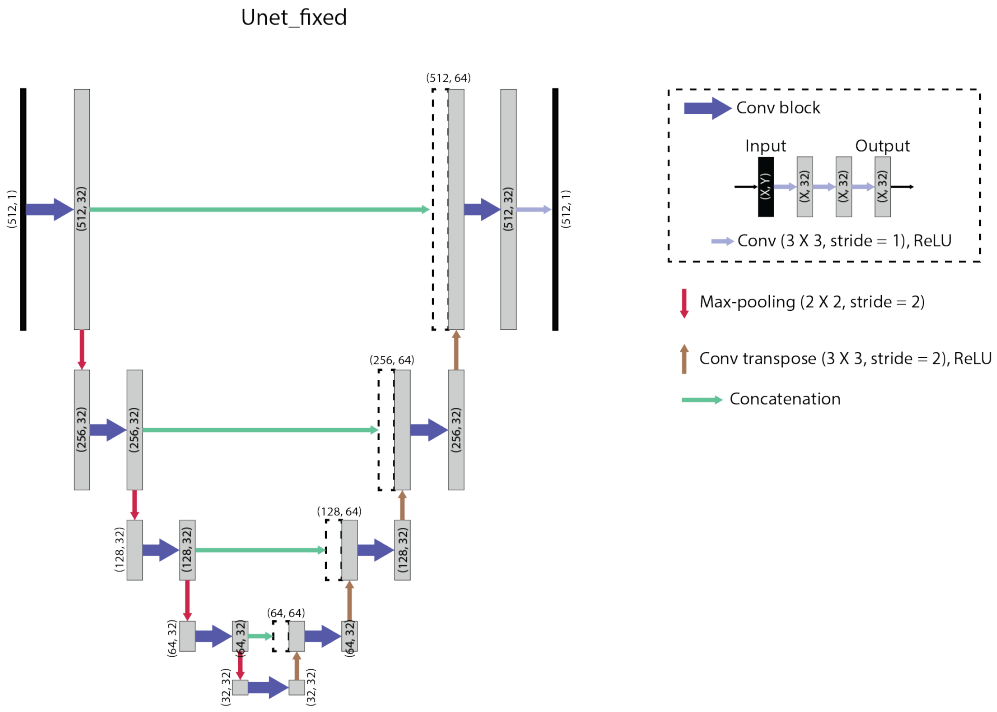


Supplementary Figure 1. Additional examples of whole-brain noisy images denoised by trained network. A) Top – random example noisy images (acquired at low laser power) from different animals (1 z-plane from 3D image stack) and corresponding denoised output, bottom left – zoom ins of image portions highlighted with dotted box in top panel, bottom right – pixel intensities extracted from noisy images and deep denoised images along the dotted lines shown in insets. Clear peaks and valleys in intensity profiles in denoised images correspond to individual nuclei. Data comes from strain ZIM504.

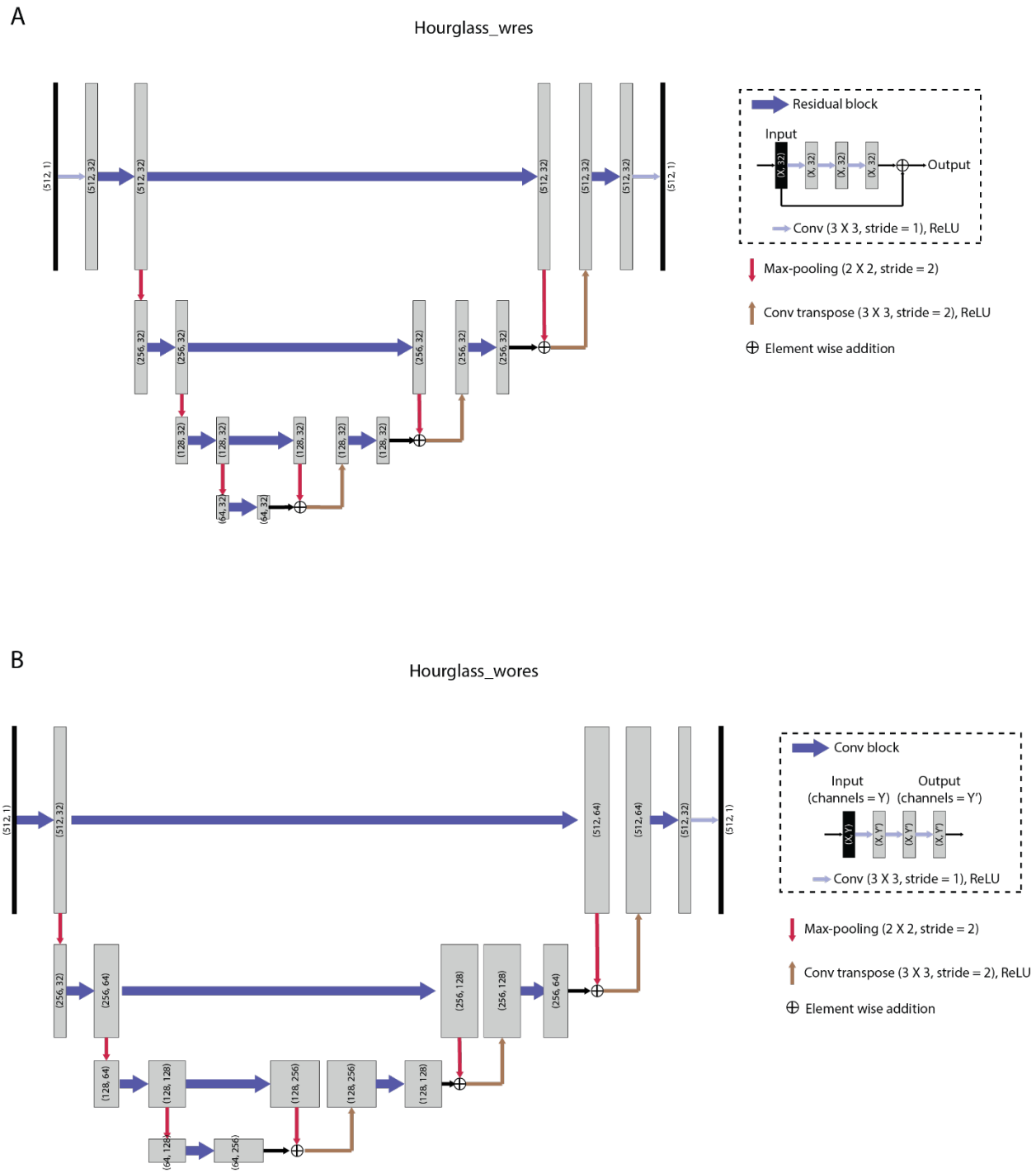
A



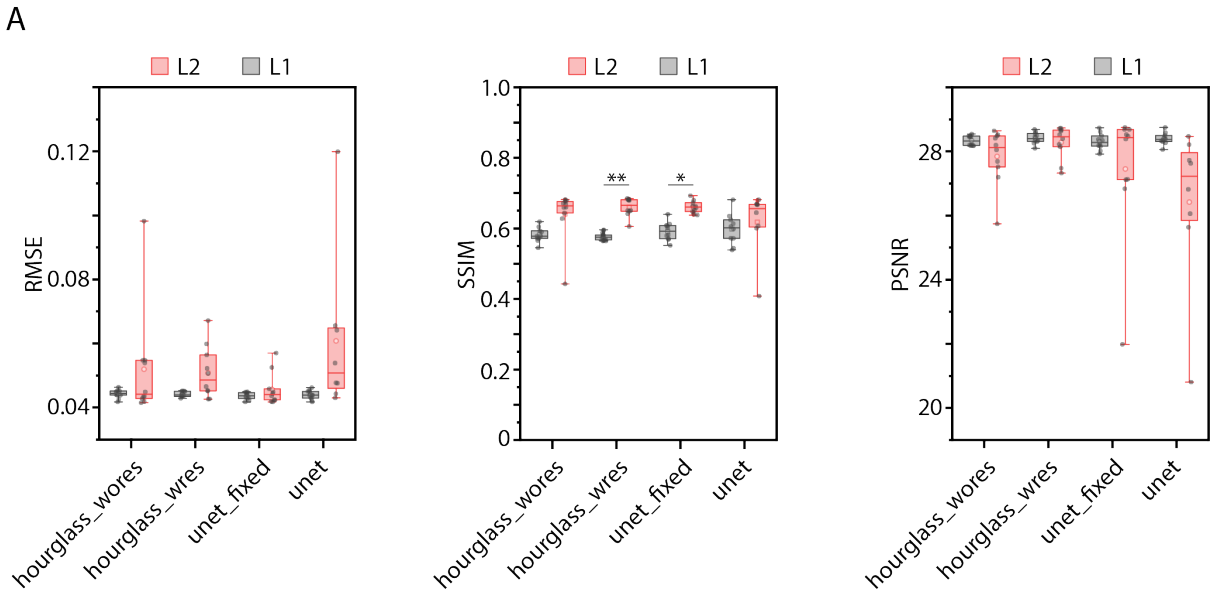
B



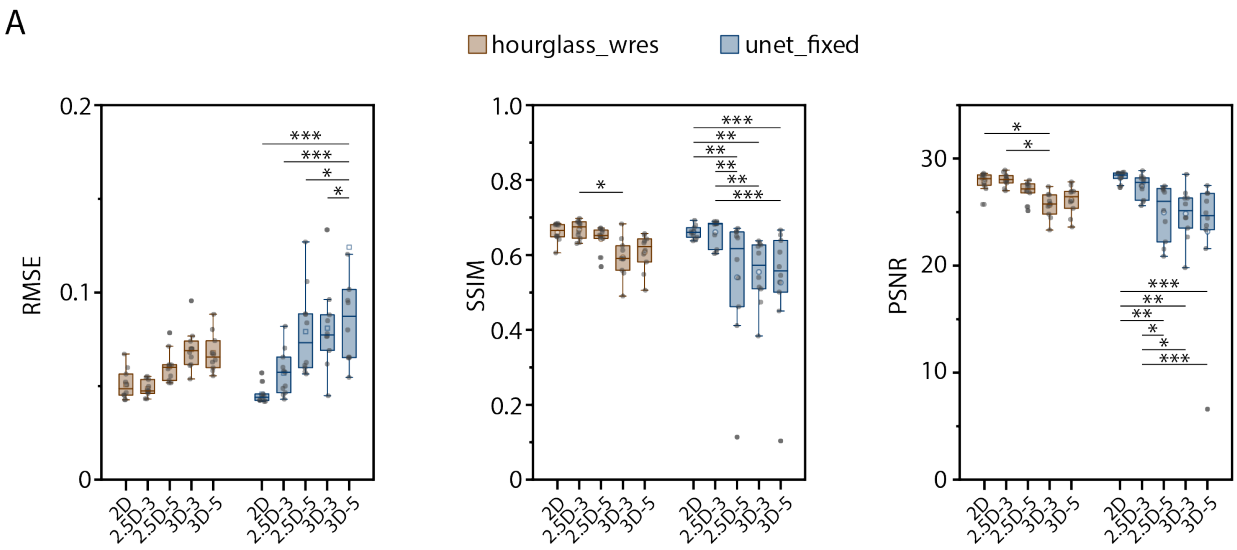
Supplementary Figure 2: Architecture details of variants of UNets tried for denoising images. A) Traditional UNet architecture. B) Optimized UNet architecture with fixed channel depth.



Supplementary Figure 3: Architecture details of variants of Hourglass architecture tried for denoising images. A) Optimized hourglass architecture with fixed channel depth and residual connections in convolution block (conv block). B) Hourglass architecture with channel depth doubling after max-pooling operations and no residual connections in conv block.

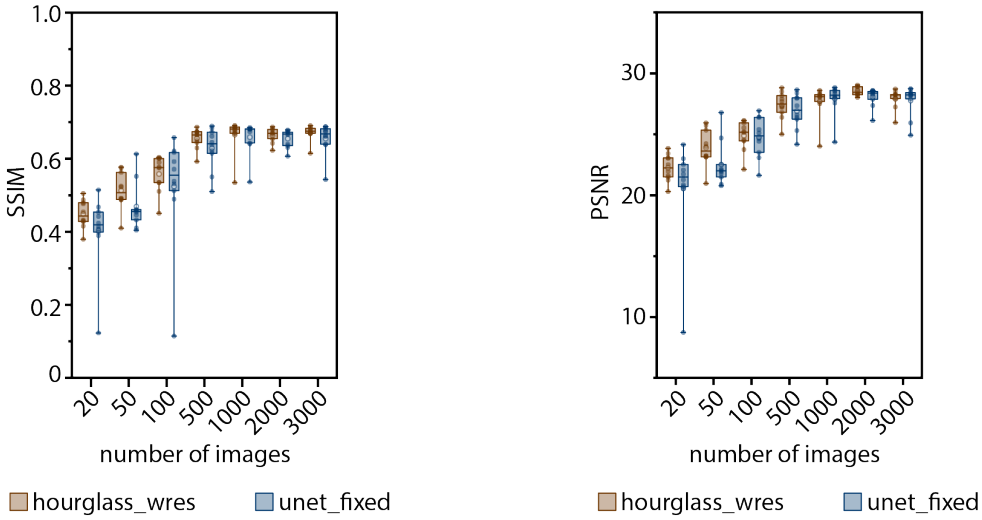


*Supplementary Figure 4. Denoising accuracy comparison across neural network architecture variants and loss functions. A) Left - RMSE accuracy, middle – SSIM accuracy and right - PSNR of denoised images across architectures trained with specific loss functions (L1 and L2 loss). Each dot in every panel corresponds to mean accuracy on 600 test images for one instance of trained network. In total, 10 instances were trained for each condition with random subset of total data used for training each instance. Data comes from strain ZIM504. (n = 10, **p<0.01, *p<0.05, Bonferroni paired comparison test). Boxes indicate 25th and 75th percentile, whiskers indicate 5th and 95th percentile of data.*

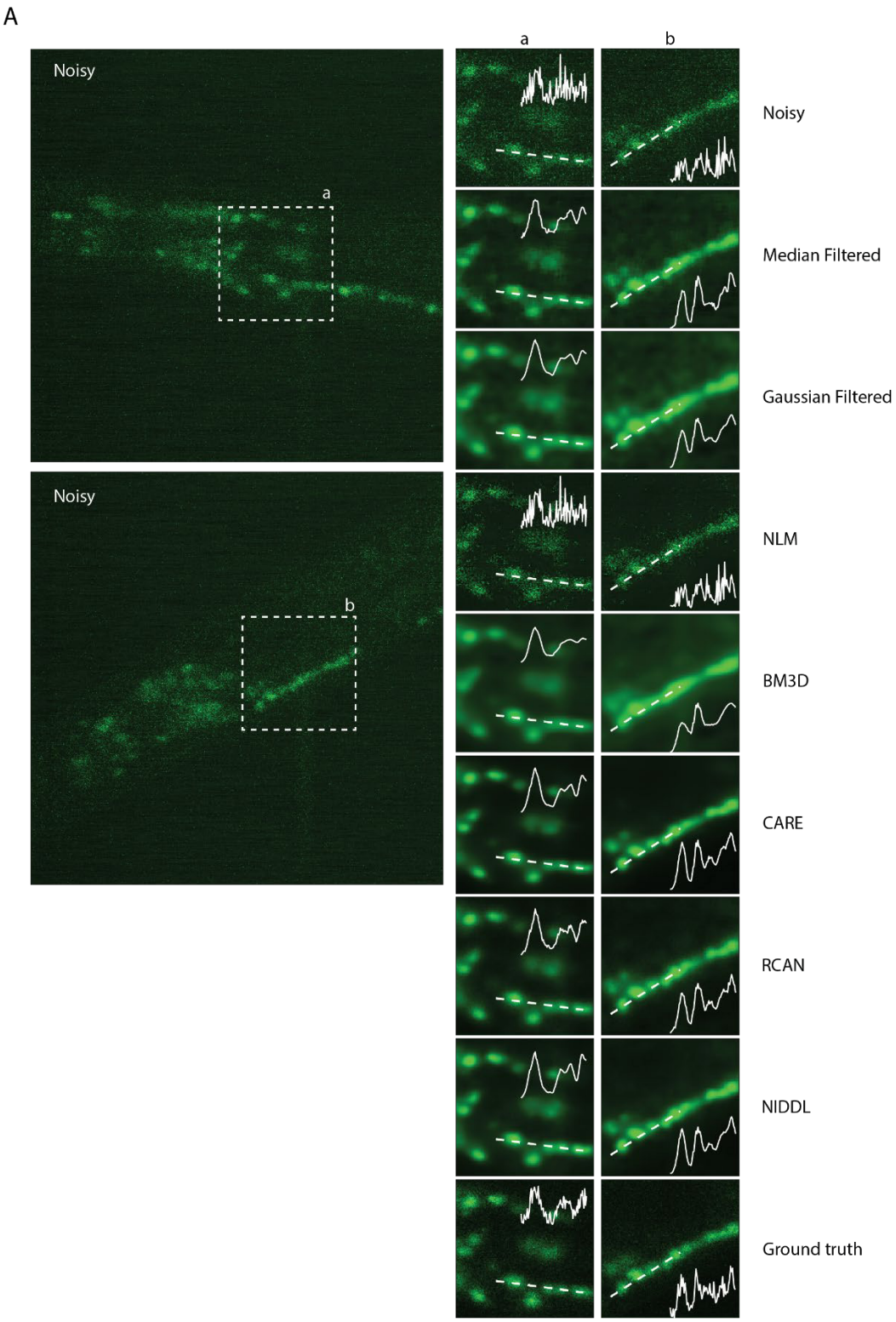


*Supplementary Figure 5. Denoising accuracy comparison across training modes for the optimized architectures 'unet_fixed' and 'hourglass_wres'. See Methods – Network Optimization 3 for description of training modes. A) Left - RMSE accuracy, middle – SSIM accuracy and right - PSNR of denoised images across training modes for 'hourglass_wres' and 'unet_fixed' architectures. Each dot in every panel corresponds to mean accuracy on 600 test images for one instance of trained network. In total, 10 instances were trained for each condition with random subset of total data used for training each instance. Data comes from strain ZIM504. (n = 10, ***p<0.001, **p<0.01, *p<0.05, Bonferroni paired comparison test). Boxes indicate 25th and 75th percentile, whiskers indicate 5th and 95th percentile of data.*

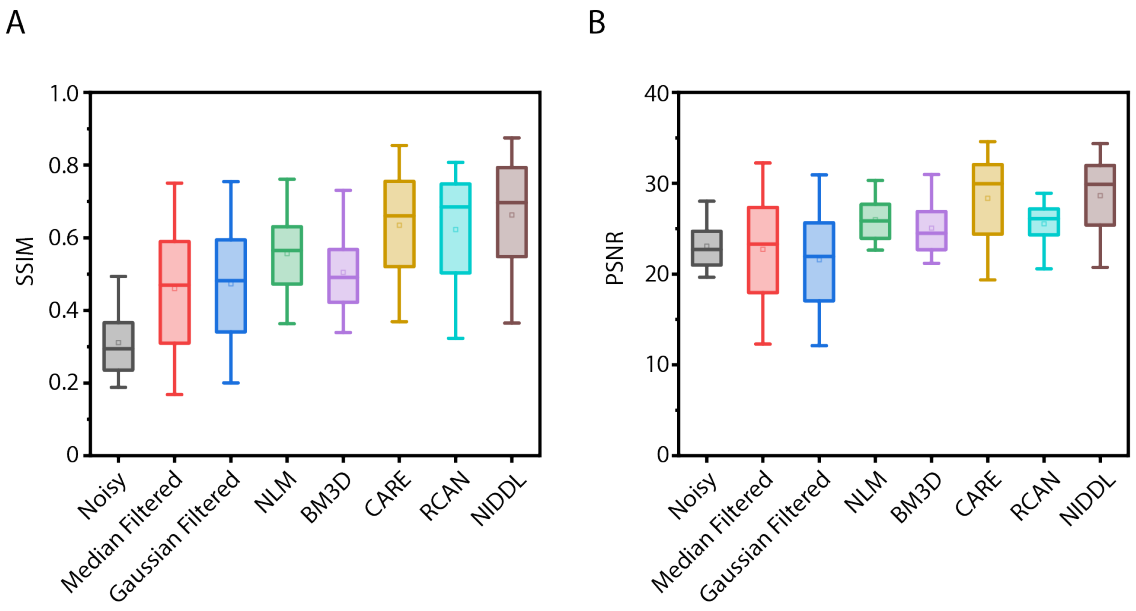
A



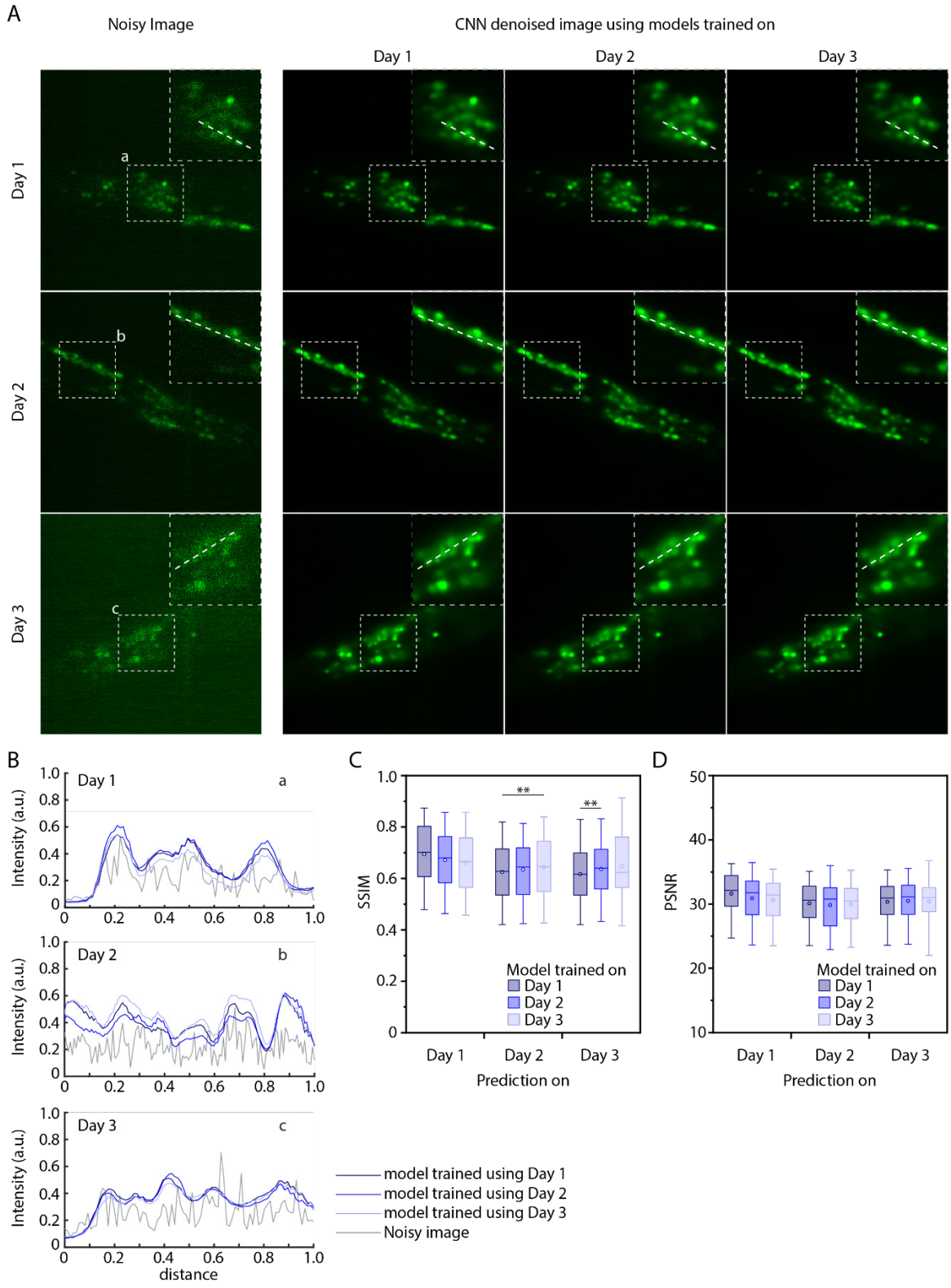
Supplementary Figure 6. Accuracy vs training data size trade-off for optimized architectures. A) Left – SSIM accuracy, right – PSNR of denoised images vs number of images used for training the networks. Each dot corresponds to mean accuracy on 600 test images for one instance of trained network. In total, 10 instances were trained for each condition with random condition specific subset of total data used for each training instance. Data come from strain ZIM504. Boxes indicate 25th and 75th percentile, whiskers indicate 5th and 95th percentile of data.



Supplementary Figure 7. Additional qualitative comparison examples of denoising methods on whole-brain images. Left - example noisy images (single z planes) from noisy whole-brain image stacks (acquired at low laser power). Right – corresponding denoised output generated by different methods. Cell nuclei are labelled with nuclear localized GCaMP5K. Data comes from strain ZIM504. Inset shows intensity profile along the dotted line.

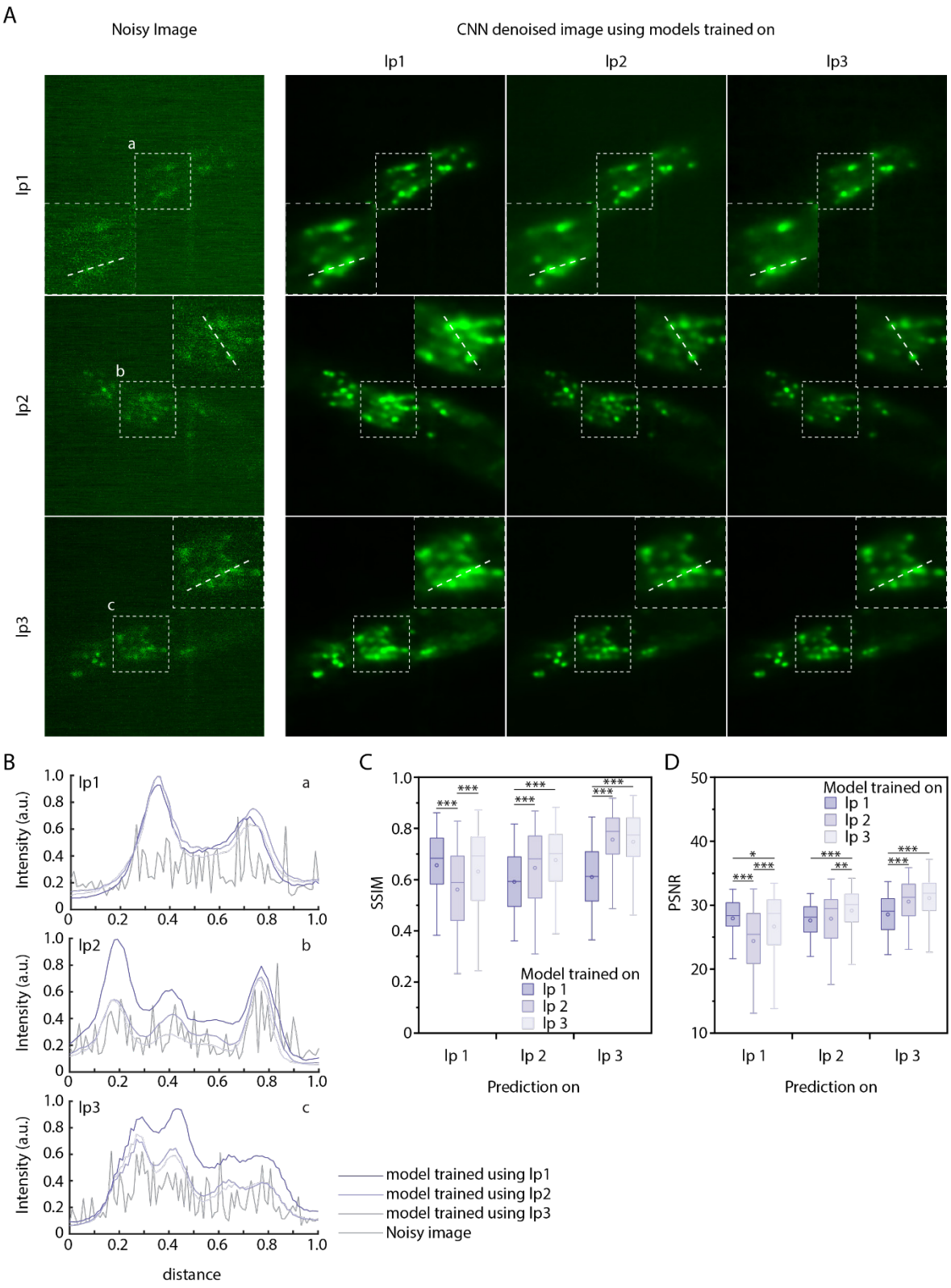


Supplementary Figure 8. Additional accuracy comparison of various denoising methods for whole-brain images. A) Left - SSIM accuracy, right – PSNR across noisy images, median filtered images, Gaussian filtered images, and deep denoised images (n = 600 images). Data comes from strain ZIM504. Boxes indicate 25th and 75th percentile, whiskers indicate 5th and 95th percentile of data.



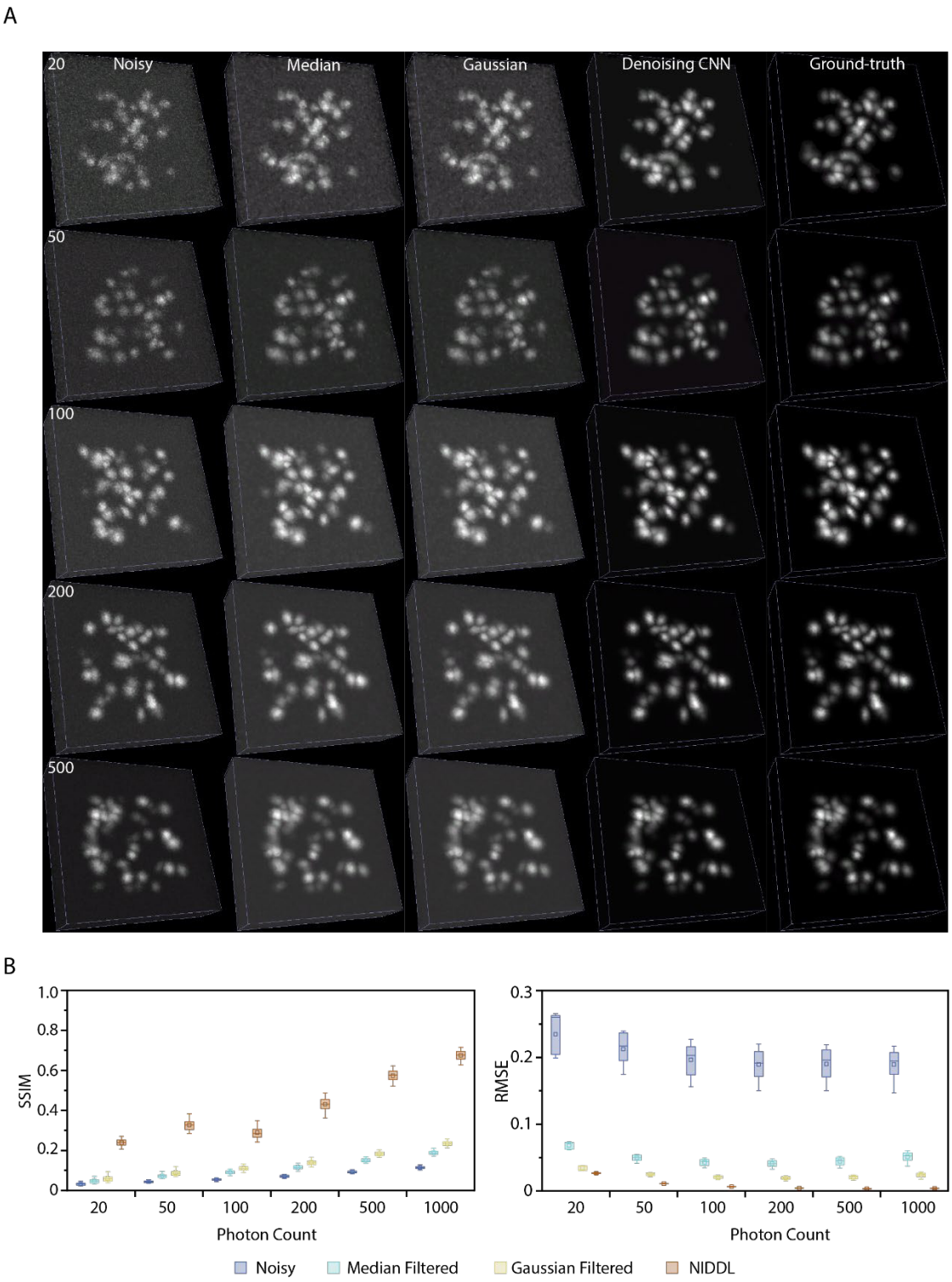
Supplementary Figure 9. Additional deep denoising accuracy quantification on images from 3 different imaging sessions (different days) when model is trained with specific day's data only. Data comes from ZIM504 strain. A) Left – example noisy images from 3 imaging sessions, right – corresponding denoised image outputs by 3 different networks trained on specific day's data. B) Intensity profiles along the dotted

*lines shown in insets in A for noisy images and denoised images output by 3 different networks. C) SSIM accuracy and D) PSNR achieved by deep denoising on noisy images from 3 imaging sessions when networks are trained on specific strain's data. (n = 40, 1,043, 1,043, 1,532, 50, 1,532, 1,403, 1,403, 50 images for 9 conditions, **p<0.01, Bonferroni paired comparison test). Boxes indicate 25th and 75th percentile, whiskers indicate 5th and 95th percentile of data.*

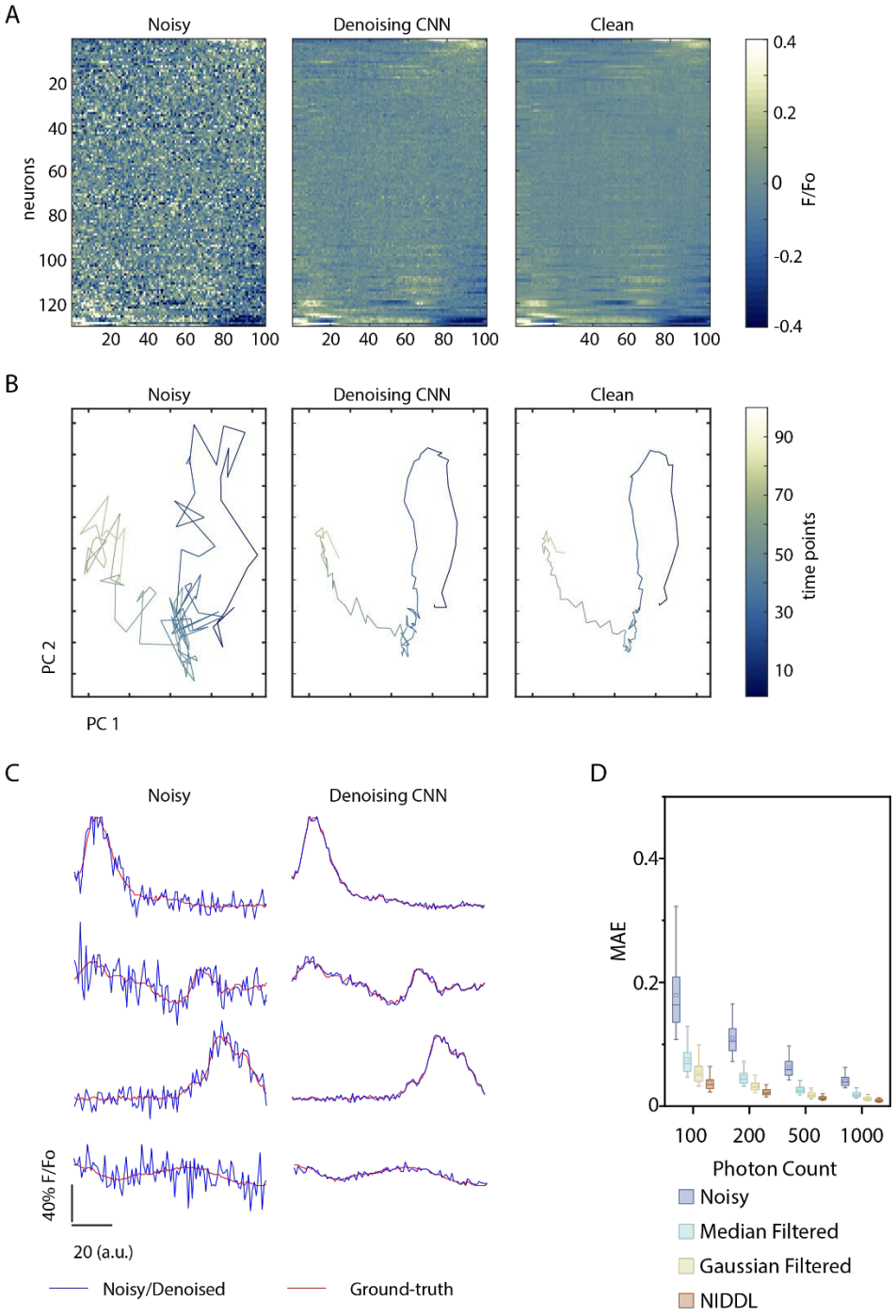


Supplementary Figure 10. Additional deep denoising accuracy quantification on images acquired with 3 different low laser power settings when model is trained with specific setting's data only. Data comes from ZIM504 strain. A) Left – example noisy images acquired across 3 laser power settings, right –

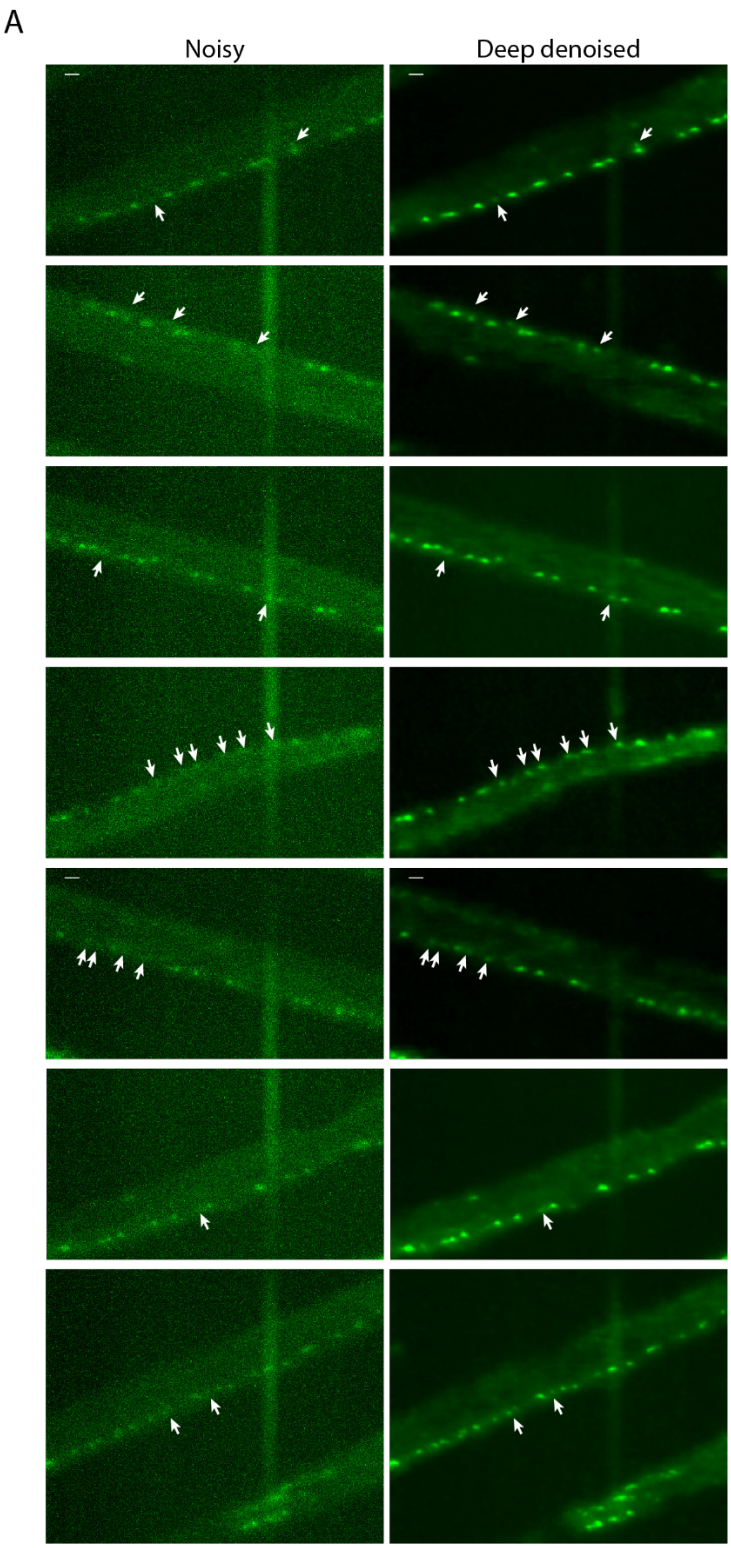
*corresponding denoised image outputs by 3 different networks trained on specific settings' data. B) Intensity profiles along the dotted lines shown in insets in A for noisy images and denoised images output by 3 different networks. C) SSIM accuracy and D) PSNR achieved by deep denoising on noisy images across 3 laser power settings when networks are trained on specific setting's data. ((n = 88, 2,728, 2,728, 2,262, 100, 2,262, 2,806, 2,806, 226 images for 9 conditions, ***p<0.001, **p<0.01, *p<0.05, Bonferroni paired comparison test). Boxes indicate 25th and 75th percentile, whiskers indicate 5th and 95th percentile of data.*



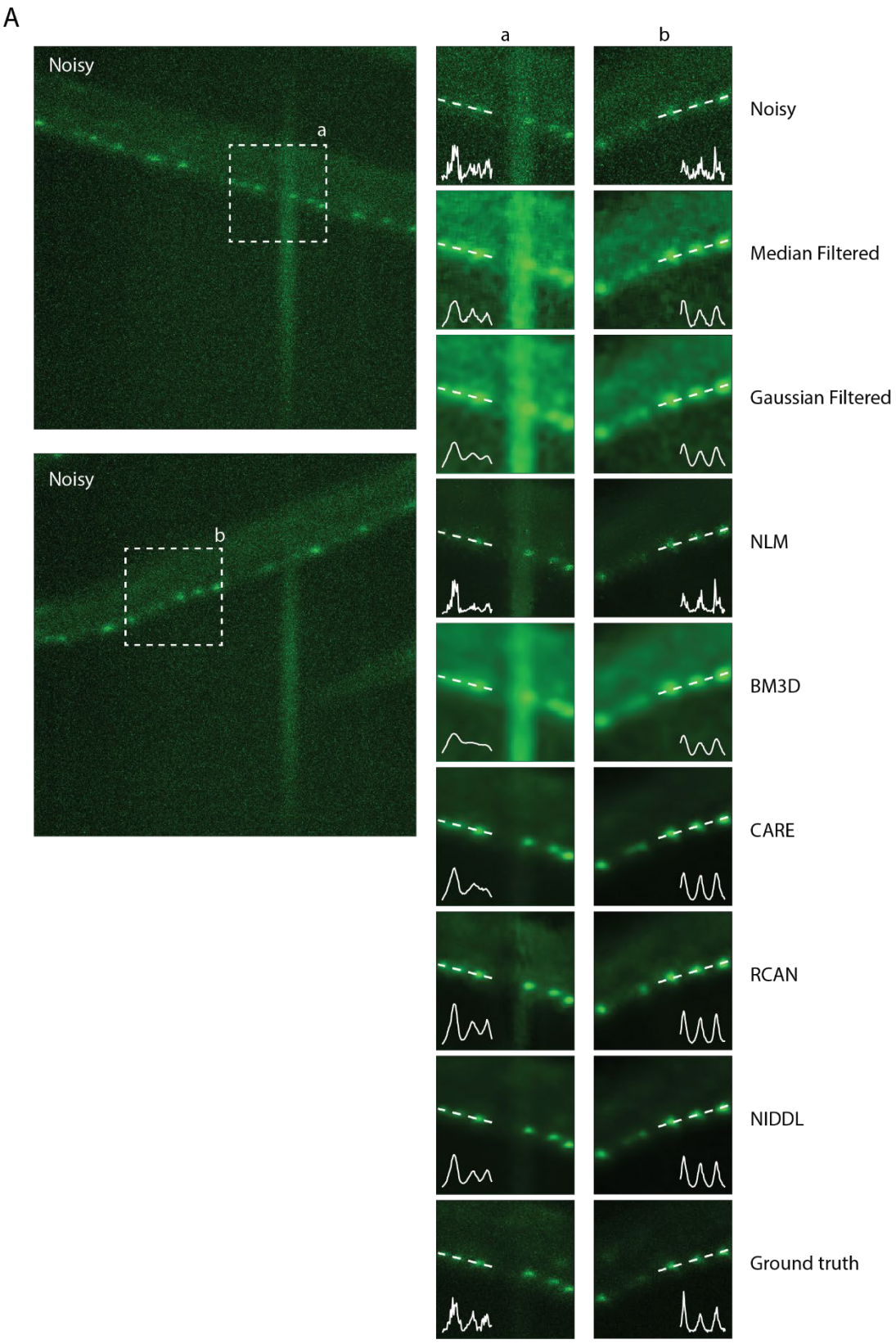
Supplementary Figure 11. Additional denoising accuracy quantification on synthetic images. A) Examples of synthetic noisy and ground-truth image stacks across range of signal levels in images (photon count levels, see Methods - synthetic image data generation for details), and corresponding denoised images generated by median filtering, Gaussian filtering, deep denoising. B) Left – SSIM, Right - RMSE comparison across methods on synthetic images across range of photon count levels (n = 100 image stacks). Boxes indicate 25th and 75th percentile, whiskers indicate 5th and 95th percentile of data.



Supplementary Figure 12. Accuracy quantification of traces extracted from deep denoised videos using semi-synthetic whole-brain video datasets (see Methods – Semi synthetic video data generation). A) Heatmap showing neuron activity traces of cells in a semi-synthetic noisy video, corresponding ground-truth video, and traces extracted from video denoised with deep network. (Video corresponds to 200 photon count level). B) Neuron activities from noisy video, ground-truth video and deep denoised video in A projected on to first 2 PC's. Deep denoising recovers low-dimensional neuron activity dynamics. C) Comparison of traces in noisy video to traces in ground-truth video for few example cells before and after deep denoising (same video as in A). D) MAE of traces to ground-truth traces across denoising methods and signal levels on videos (photon count levels). (N = 1 video, n = 130 cells in each video for each condition). Boxes indicate 25th and 75th percentile, whiskers indicate 5th and 95th percentile of data.

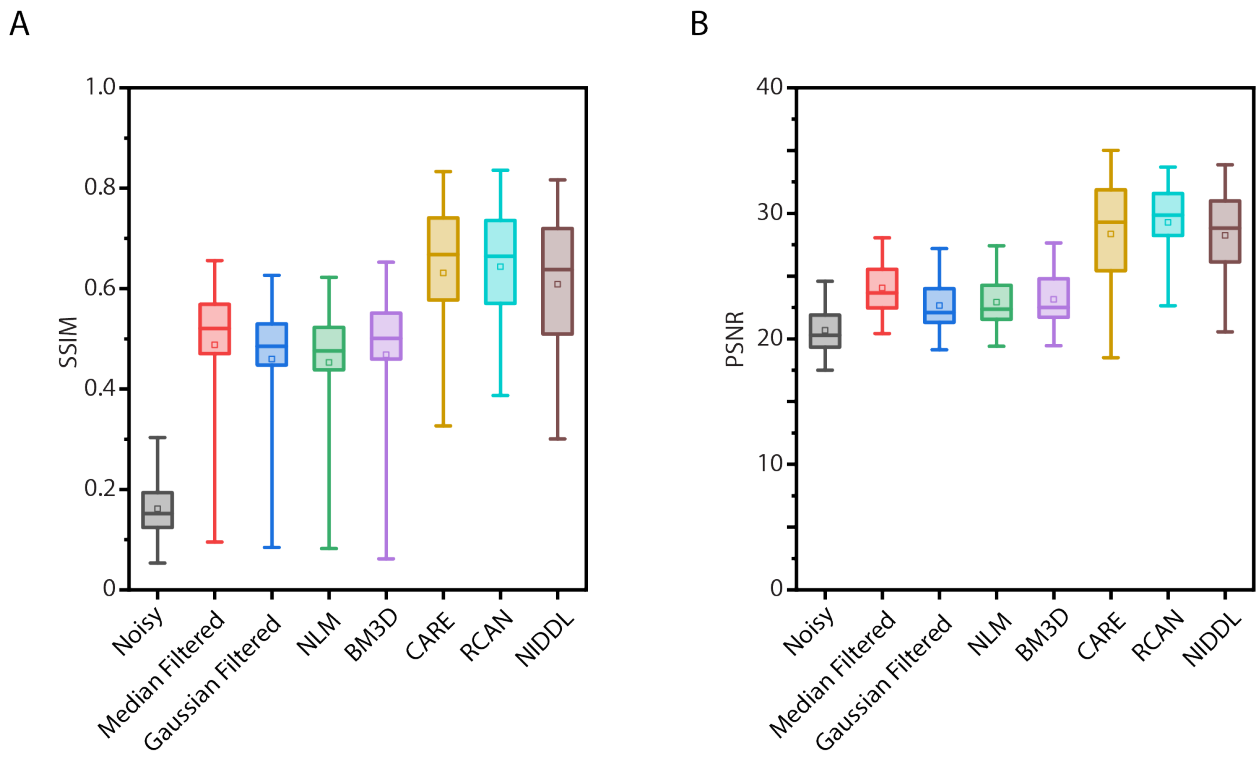


Supplementary Figure 13. Additional examples of deep denoising of large field of view (FOV) noisy image stacks. Left - max projections of noisy images (acquired at low laser power, 20X magnification) showing motor neurons in ventral nerve cord. Right – corresponding deep denoised outputs. Data comes from strain OH16230. Arrows indicate example cells that are difficult to identity in noisy images but can be easily identified in denoised images.

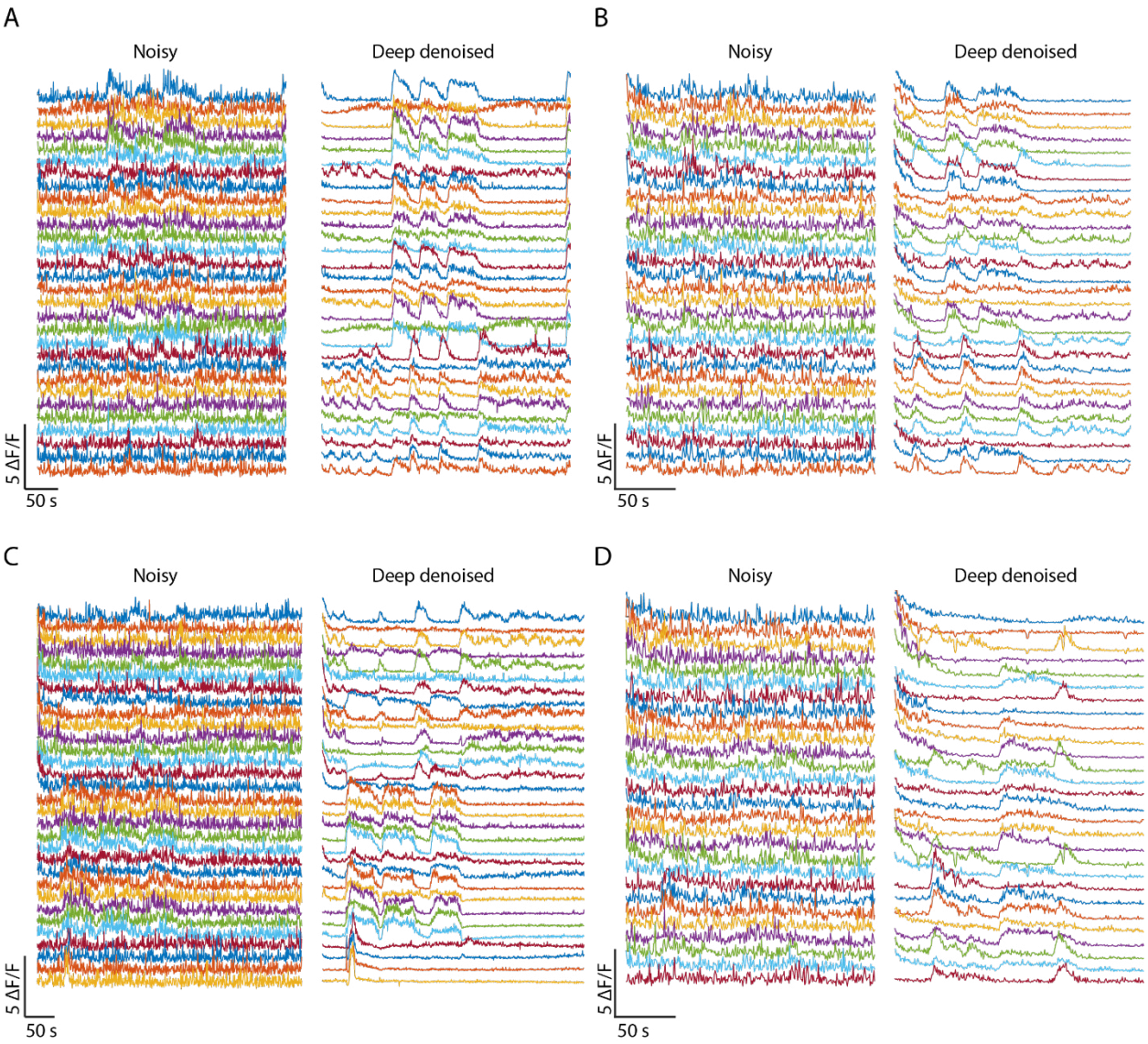


Supplementary Figure 14. Additional qualitative comparison examples of denoising methods on large FOV (20X) motor neuron images. Left - max projections of noisy images (acquired at low laser power,

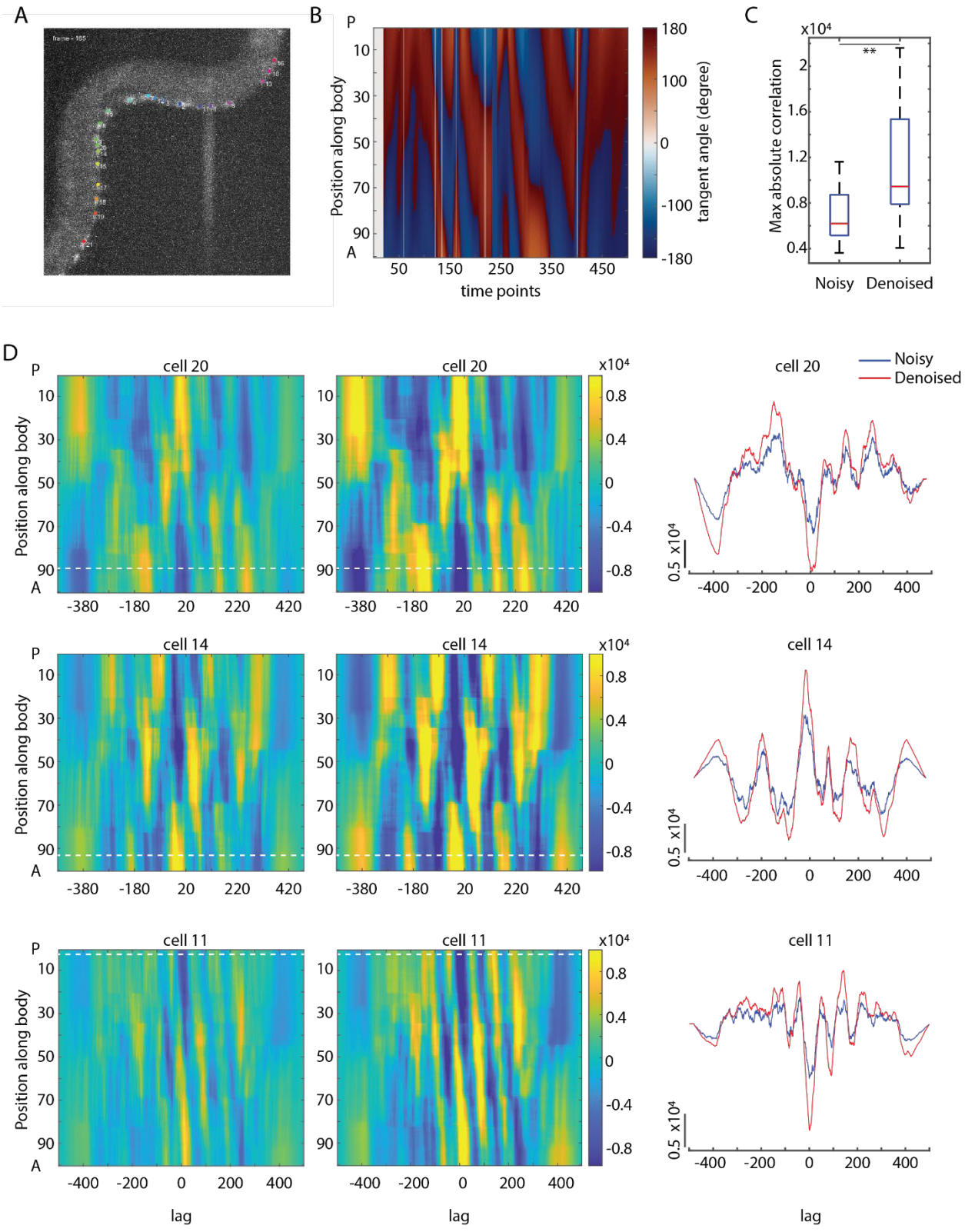
20X magnification) showing motor neurons in ventral nerve cord. Right – corresponding denoised output generated by different methods shown for dotted box in noisy images. Inset shows intensity profile along the dotted line. Data comes from strain OH16230.



Supplementary Figure 15. Accuracy comparison of NIDDL with traditional denoising methods and deep learning based methods on large FOV (20X magnification) ventral cord data as shown in Fig. 2F. A) Left – SSIM accuracy, right – PSNR comparison across various methods. (n = 60-217 images). Boxes indicate 25th and 75th percentile, whiskers indicate 5th and 95th percentile of data.

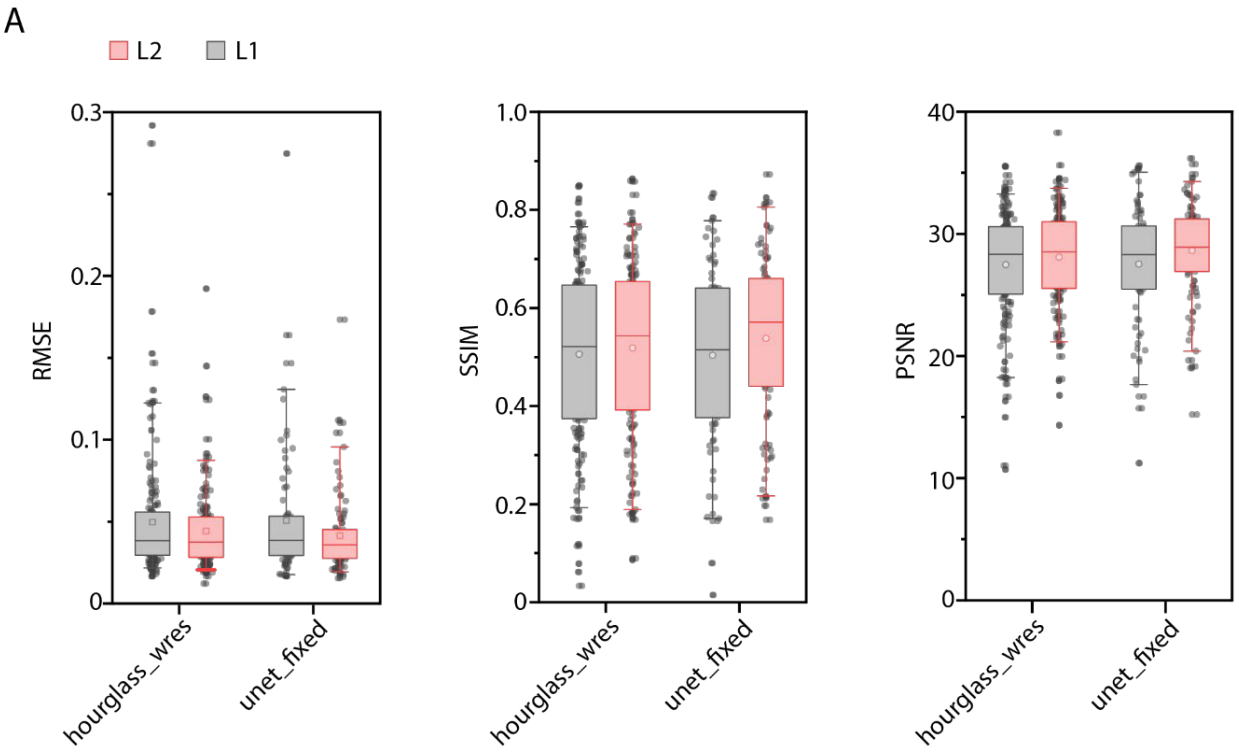


Supplementary Figure 16. Additional examples of ventral cord neurons' activity traces extracted from in-device recordings. Left panel – single pixel neuron activity traces from noisy videos, right – corresponding traces from deep denoised videos for A) worm 1 in Figure 2H, B) worm 2, C) worm 3, D) worm 4.

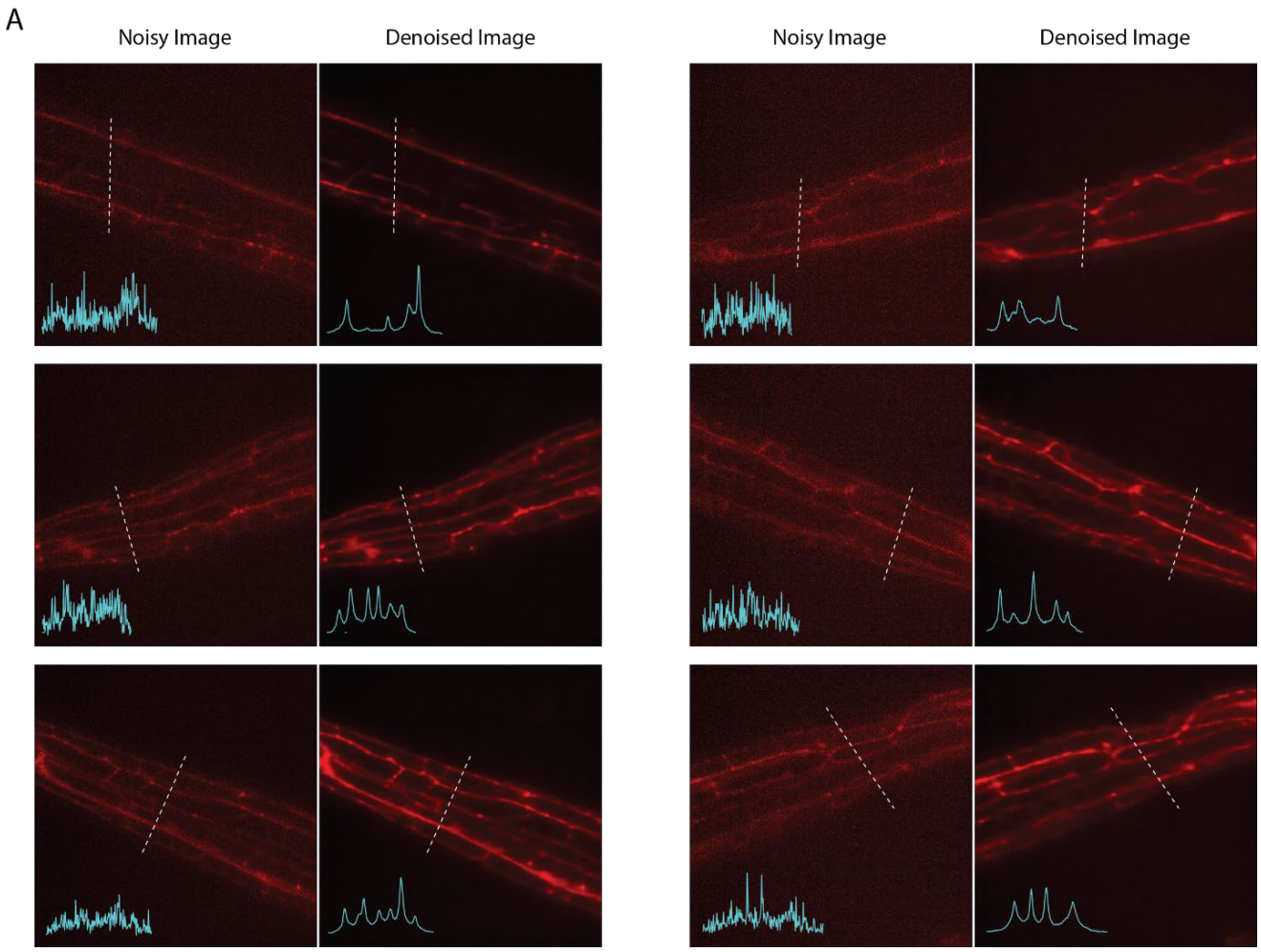


Supplementary Figure 17. Deep denoising enables correlating neuron activity to behavior in freely moving animal. A) Example frame from video recording of freely-moving C. elegans showing tracked ventral cord neurons. B) Tangent angle along the body as it changes with time. 'A' denotes anterior side and 'P'

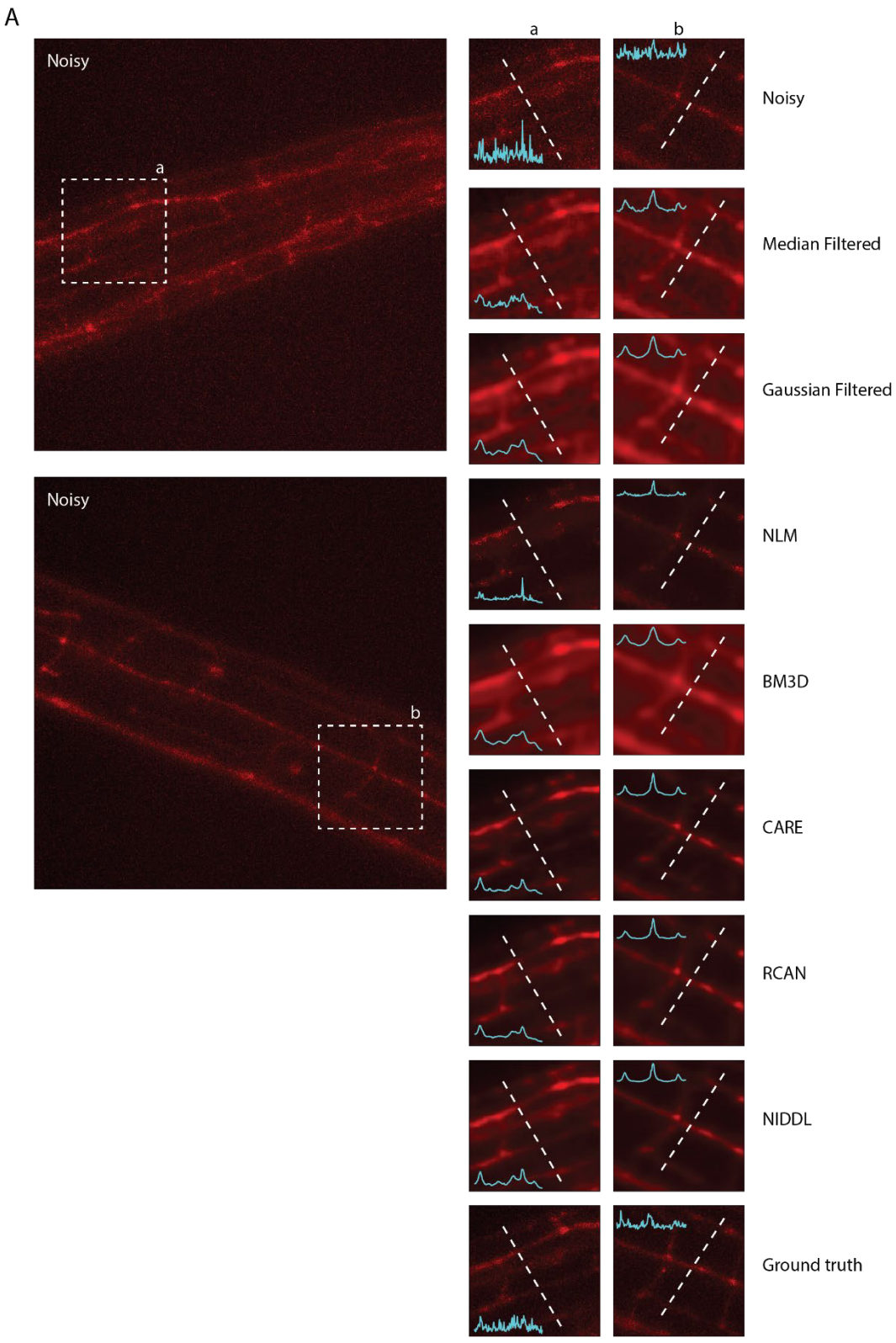
*denotes posterior side. C) Comparison of maximum absolute value of correlation between neuron activities and curvature (n = 15 cells, **p<0.01, Bonferroni paired comparison test). Boxes indicate 25th and 75th percentile, whiskers indicate extreme data points D) Left and middle panels – heat map plots of cross correlation of neuron activities to animal curvature. Body portion was discretized into 100 segments. Rows in each heat map indicate cross correlation of neuron activity to curvature at the specific body segment. 'A' denotes anterior side and 'P' denotes posterior side. White dotted line indicates the position of the cell on the body portion. Right panels – cross correlation along the dotted white line in left and middle panels i.e. cross correlation of neuron activity to local curvature of the body where the cell is located.*



Supplementary Figure 18. Network performance comparisons across two different architectures and loss functions for denoising neurite images. A) RMSE accuracy, B) SSIM accuracy and C) PSNR of deep denoised images from test data set across optimized neural network architectures ('unet_fixed' and 'hourglass_wres') and loss functions (L1 and L2). Data comes from strains GT372 and GT366. (n = 174, 159, 79, 99 images for 4 conditions). Boxes indicate 25th and 75th percentile, whiskers indicate 5th and 95th percentile of data.

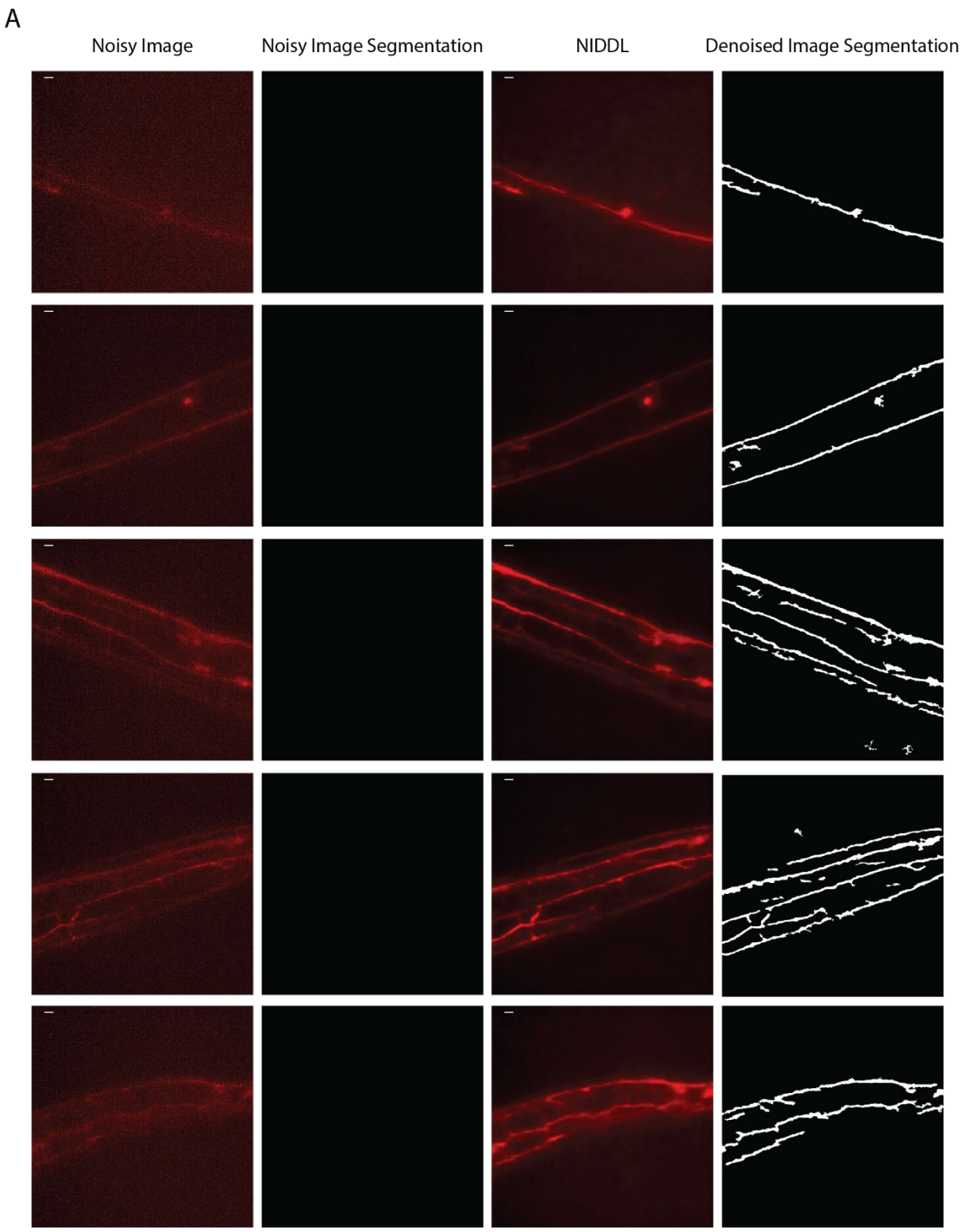


Supplementary Figure 19. Additional examples of deep denoising of harsh touch mechanosensory neuron PVD's neurites. A) Left panels – noisy images (acquired at low laser power), right panels – corresponding deep denoised output. Cyan traces in inset indicate pixel intensities along dotted lines. Data comes from strain GT366.

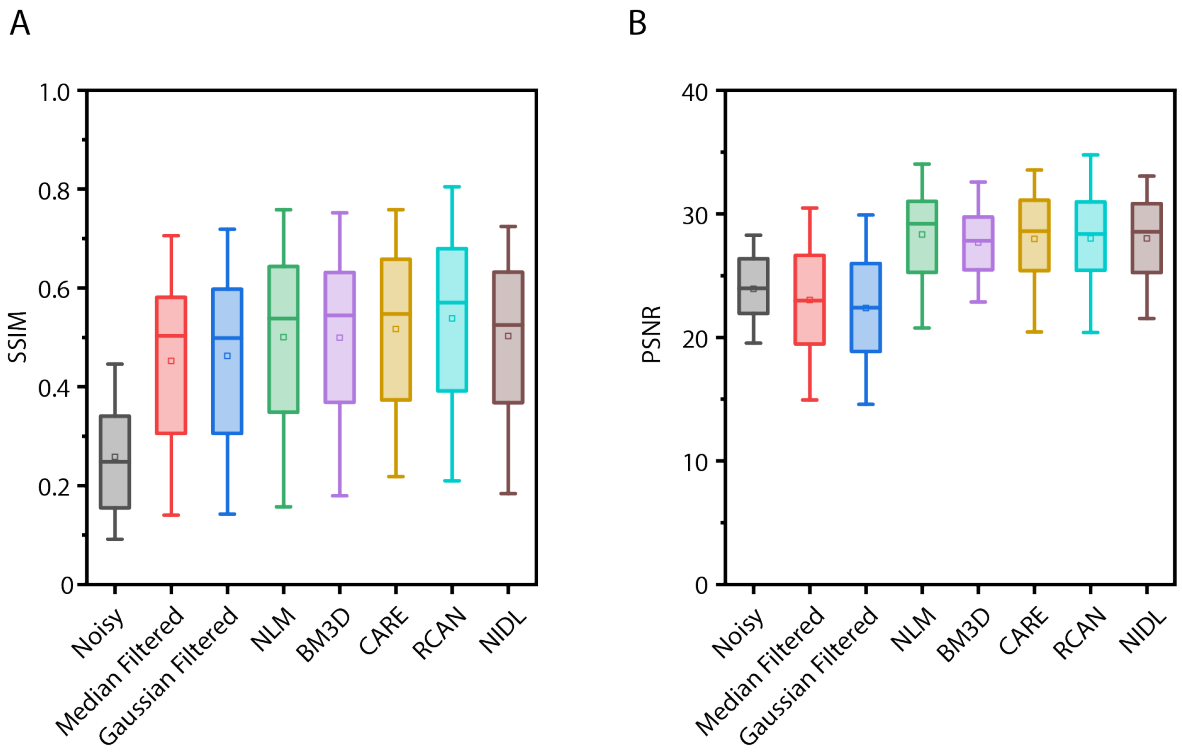


Supplementary Figure 20. Additional qualitative comparison examples of different denoising methods on harsh touch mechanosensory neuron PVD's neurites. Left - max projections of noisy images (acquired at low laser power, 20X magnification) showing motor neurons in ventral nerve cord. Right – corresponding

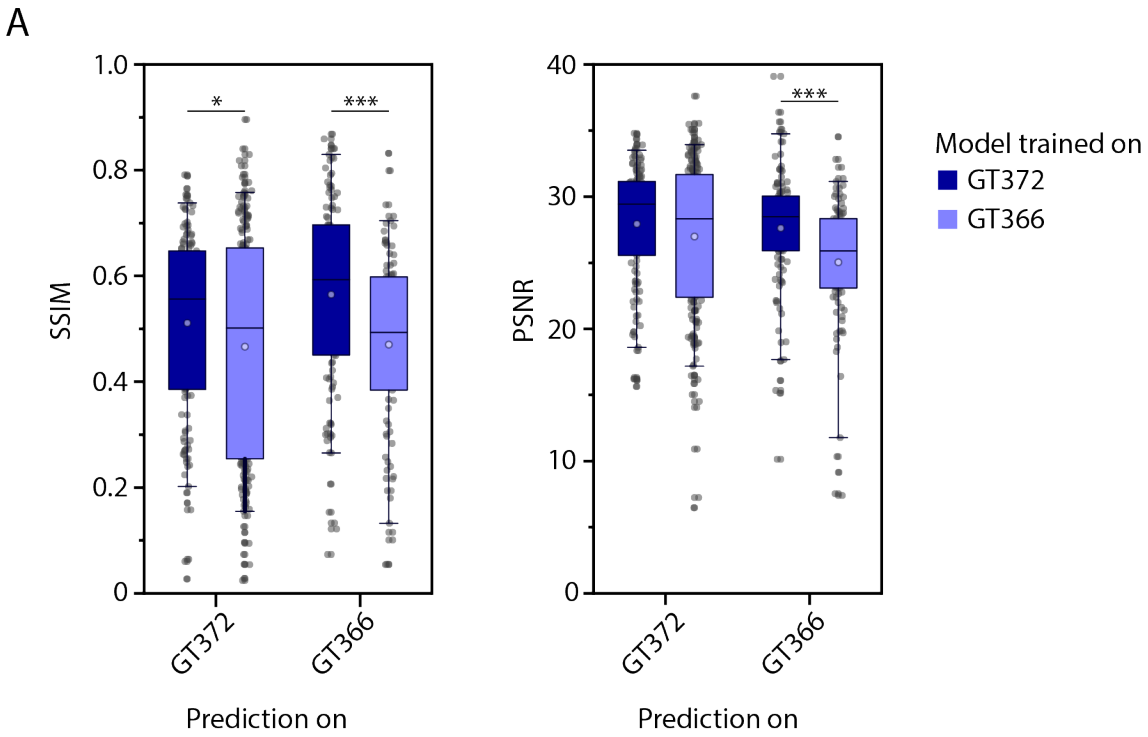
denoised output generated by different methods shown for dotted box in noisy images. Inset shows intensity profile along the dotted line. Data comes from strain OH16230.



Supplementary Figure 21. Deep denoising of images facilitate neurite segmentation. A) Examples of noisy images showing harsh touch mechanosensory neuron PVD's neurites, no regions are detected in noisy images with simple morphological operations (see Methods – Neurite segmentation), corresponding deep denoised outputs and segmented neurites in denoised images. Data comes from strain GT366.



Supplementary Figure 22. Additional accuracy comparison of NIDDL with other non-deep-learning and deep-learning based methods for neurites of gentle touch and harsh touch neurons. A) SSIM accuracy and B) PSNR of noisy images, median filtered images, Gaussian filtered images, and deep denoised images. Data comes from strains GT372 and GT366. (n = 86-443 images). Boxes indicate 25th and 75th percentile, whiskers indicate 5th and 75th percentile.



*Supplementary Figure 23. Accuracy comparison on noisy images from 2 different strains, GT372 and GT366 that label neurites of gentle touch and harsh touch mechanosensory neurons respectively, when models are trained on specific strain's data. A) Left – SSIM accuracy and right – PSNR achieved by deep denoising when networks are trained on specific strain's data. (n = 129, 203, 118, 97 images for 4 conditions, ***p<0.001, *p<0.05, Bonferroni paired comparison test). Boxes indicate 25th and 75th percentile, whiskers indicate 5th and 95th percentile of data.*

Supplementary Video Captions

Supplementary Video 1 – Deep denoising examples of low SNR whole-brain images. Video shows 3 low SNR images from different animals (1 z-plane from 3D stack) acquired at low laser power and corresponding deep denoised image generated by trained network. Data comes from strain ZIM504 and OH16230. Scale bar corresponds to 5 µm.

Supplementary Video 2 – Deep denoising of whole-brain recording. Max-projection images for low SNR video acquired at low laser power, high SNR video acquired at high laser power, and corresponding deep denoised video generated by network. Cell nuclei are labelled with nuclear localized GCaMP5K. Data comes from strain ZIM504. Scale bar corresponds to 5 µm.

Supplementary Video 3 – Deep denoising of ventral cord neurons for animals restrained in microfluidic device. Max projection images of low SNR video acquired at low laser power and corresponding deep denoised video. Video shows two animals in microfluidic device. Data comes from strain OH16230. Scale bar corresponds to 10.4 µm.

Supplementary Video 4 – Deep denoising of ventral cord neurons for freely moving animal. Max projection images of low SNR video acquired at low laser power and corresponding deep denoised video. Data comes from strain OH16230. Scale bar corresponds to 10.4 µm.

Supplementary Video 5 – Deep denoising of neurites of harsh and gentle touch mechanosensory neurons. Video shows 3 low SNR images (max-projections of 3D stack) from different animals acquired at low laser power and corresponding deep denoised image generated by trained network. Data comes from strain GT372 and GT366. Scale bar corresponds to 5 µm.