DeepVID: A Self-supervised Deep Learning Framework for Two-photon Voltage Imaging Denoising

Chang Liu[1,2], **Jelena Platisa**[3,4,5], **Xin Ye**[2,6], **Allison M. Ahrens**[7], **Ichun Anderson Chen**[6], **Ian G. Davison**[6,7,8], **Vincent A. Pieribone**[3,4,5], **Jerry L. Chen**[2,6,7,8], **Lei Tian**[1,2,6]

[1]*Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215, USA*
[2]*Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA*
[3]*Department of Cellular and Molecular Physiology, Yale University, New Haven CT 06520, USA*
[4]*Department of Neuroscience, Yale University, New Haven CT 06520, USA*
[5]*The John B. Pierce Laboratory, New Haven CT 06520, USA*
[6]*Center for Neurophotonics, Boston University, Boston MA 02215, USA*
[7]*Department of Biology, Boston University, Boston MA 02215, USA*
[8]*Center for Systems Neuroscience, Boston University, Boston MA 02215, USA*
cl6@bu.edu, leitian@bu.edu

Abstract: High-speed population-level voltage imaging is suffered from the shot noise limit. We developed a self-supervised deep learning framework for voltage imaging denoising (DeepVID) without the need for any ground-truth high-SNR data. © 2022 The Author(s)

1. Introduction

Voltage imaging is an evolving tool to continuously image neuronal activities for large number of neurons. Recently, a high-speed low-light two-photon voltage imaging framework was developed, which enabled kilohertz-scanning on population-level neurons in the awake behaving animal [1]. However, with a high frame rate and a large field-of-view (FOV), shot noise dominates pixel-wise measurements and the neuronal signals are difficult to be identified in the single-frame raw measurement. Another issue is that although deep-learning-based methods has exhibited promising results in image denoising [2], the traditional supervised learning is not applicable to this problem as the lack of ground-truth "clean" (high SNR) measurements.

To address these issues, we developed a self-supervised deep learning framework for voltage imaging denoising (DeepVID) without the need for any ground-truth data. Inspired by previous self-supervised algorithms [3,4], DeepVID infers the underlying fluorescence signal based on the independent temporal and spatial statistics of the measurement that is attribute to shot noise. DeepVID achieved a 15-fold improvement in SNR when comparing denoised and raw image data.

2. Methods

DeepVID combines self-supervised frameworks implemented in DeepInterpolation [3] and Noise2Void [4]. The network was designed to denoise a single frame from each sub-area at a time. It was trained to predict the central frame N_0 using an input image time series, consisting of N_{pre} frames before and N_{post} frames after the central frame, in addition to a degraded central frame with several "blind" pixels. A random set of pixels (p_{blind}) in the central frame were set as blind pixels using a binary mask, whose intensities were replaced by a random value sampled from randomly selected pixels in the frame.

The network architecture of DeepVID was based on the DnCNN [2], a fully convolutional network with residual blocks **(Figure 1A)**. This architecture was chosen to better accommodate the 8:1 aspect ratio in the sub-image scanned by each beamlet. The network was constructed with 2D convolution layers (Conv), batch normalization (BN) layers and Parametric Rectified Linear Unit (PReLU) activation layers, with 16 repeated residual blocks in the middle. Each residual block contained two 3x3 Conv layers with BN layers followed, and an PReLU activation layer was appended after the first BN layer. The skip connection was added to link low dimensional and high dimensional features by adding the feature map of the input and the output for each residual block.

The hyperparameters ($N_{\mathrm{pre}} = N_{\mathrm{post}} = 3$, $\mathrm{p}_{\mathrm{blind}} = 10\%$) were optimized to maintain the temporal dynamic of voltage signal spikes while recovering a high single-frame spatial resolution. The loss function was the mean squared error (i.e. L2 loss) between the original and denoised central frame and was calculated only on the blind pixels. The training was performed using the Adam optimizer with 360 steps per epoch and a batch size of 4. The training stopped after going over all samples in the data set one time to avoid overfitting. The learning rate was initialized at 5x10^{-6} and reduced to 1x10^{-6} when the loss on the validation set did not decrease in the past 288,000 samples.

Once DeepVID was trained, inference denoising of subsequent image data using the trained model was performed frame-by-frame by feeding each corresponding 7-frame image time series. It can be performed at approximately 200 frames per second on a single Nvidia P100 GPU.

3. Results

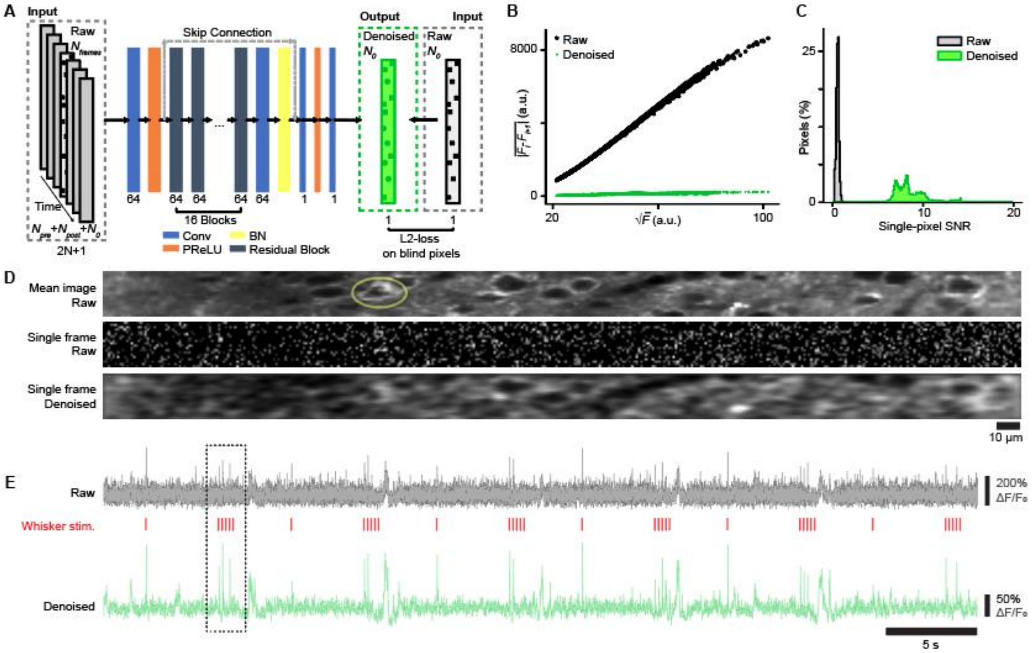

Figure 1 The framework, the network structure, and example results of DeepVID.

We first assessed the frame-to-frame variability in fluorescence signal in the raw data and confirmed that the fluctuation in each pixel is proportional to the square root of the mean fluorescence (**Figure 1B**), as expected for shot noise limited signals. DeepVID drastically reduced the frame-to-frame variability, resulting in a 15-fold improvement in SNR when comparing denoised and raw image data (SNR: 0.567±0.002, raw; 8.858±0.027, denoised, n = 8,000 pixels) (**Figure 1C**). By breaking this fundamental noise constraint, the underlying fluorescence signal can be more accurately inferred at individual time points (**Figure 1D**). The reduction in shot noise fluctuations in denoised traces readily allowed for the identification of potential sensory-evoked and non-evoked spiking events (**Figure 1E**).

4. References

[1] Platisa, J., Ye, X., Ahrens, A. M., Liu, C., Chen, I. A., Davison, I. G., Tian, L., Pieribone, V. A., & Chen, J. L. (2021). High-Speed Low-Light In Vivo Two-Photon Voltage Imaging of Large Neuronal Populations. *BioRxiv*, 2021.12.07.471668.

[2] Zhang, K., Zuo, W., Chen, Y., Meng, D., & Zhang, L. (2016). Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. *IEEE Transactions on Image Processing*, *26*(7), 3142–3155.

[3] Lecoq, J., Oliver, M., Siegle, J. H., Orlova, N., Ledochowitsch, P., & Koch, C. (2021). Removing independent noise in systems neuroscience data using DeepInterpolation. *Nature Methods 2021*, 1–8.

[4] Krull, A., Buchholz, T.-O., & Jug, F. (2018). Noise2Void - Learning Denoising from Single Noisy Images. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, *2019-June*, 2124–2132.