SuperMIBI: A convolutional neural network for prediction of upsampled mutliplexed imaging data from short acquisition times

Category: *Computer Vision*

Authors: Noah Greenwald (nfgreen@stanford.edu) & Erin McCaffrey (erinmcc@stanford.edu)

Abstract

Biological imaging provides a rich source of information to evaluate hypotheses across a range of disciplines. However, acquisition time is a limiting factor to most imaging platforms, as there is a fundamental tradeoff between data quality and the amount of time required for data collection. Here we present a proof-of-concept for SuperMIBI, which takes advantage of advances in deep neural networks—specifically convolutional neural networks (CNN)—to tackle this problem by predicting upsampled, denoised multiplexed imaging data from noisy, low quality data collected at short acquisition time. We use a simple CNN architecture to predict signal intensity in multichannel data from input collected at one-fourth the acquisition time. We compared our network's performance with an analogous denoising method. After several iterations of modifying the network architecture, we were able to obtain predicted outputs that are starting to look similar to the ground truth output, though underperforming when compared with the baseline method.

Introduction

Imaging is a one of the most critical tools used in experimental biology. Traditional imaging platforms, such as fluorescence microscopy have enabled important and novel insights into basic biological mechanisms. However, imaging experiments often require non-trivial data collection times for data with sufficient quality and resolution. This process would be more efficient with a method to predict the output of this labor-intensive process using images collected over a shorter duration. Recently the ability to employ deep learning to restore image data quality was robustly demonstrated for fluorescence microscopy[1]. However, no analogous solution for mass spectrometry-based imaging has been produced. Here, we will be leveraging imaging data generated by Multiplexed Ion Beam Imaging (MIBI), a high dimensional mass spectrometry imaging platform for biological tissues, developed by our lab at Stanford. Unlike traditional optical imaging platforms, MIBI employs secondary ion mass spectrometry to avoid the spectral overlap of fluorophores enabling much higher dimensional imaging of cells in tissues.

Similar to the relationship between fluorescence data quality and image exposure time, MIBI data quality is limited by a parameter called the dwell time, which dictates how long the imaging platform spends collecting data in each pixel of a region of interest (ROI) in tissue. Here we present SuperMIBI, a CNN- based approach to tackle this problem by predicting. This network takes as input multichannel MIBI images collected at short acquisition time and outputs predicted images with reduced noise and signal intensity collected at a four-fold increase in acquisition time. Implementing SuperMIBI into both routine MIBI experiments as well as more specialized imaging problems described previously, will significantly reduce the throughput required for high quality data generation. This will ultimately increase the pace at which new insights in various fields of tissue biology can be generated.

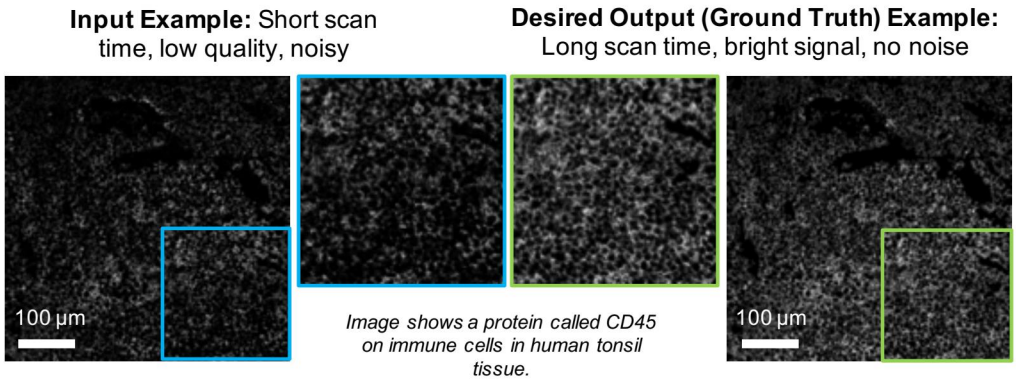
Related Work

Deep learning has been applied to several problems in the field of biological imaging[4,5]. One recent application similar to our problem was the introduction of the content-aware image restoration CNN (CARE-CNN), which employed a CNN architecture to denoise and restore fluorescent signal in biological images. Interestingly the authors of this network were able to overcome a lack of input data by creating low-quality images *in silico*. While not pursued here, access to training data is also an obstacle for our problem so methods such as those employed by the CARE authors could be pursued in future work. Another strength of this paper was the utilization of non-local means denoising as a traditional approach to image improvement, which we decided to pursue as our baseline to compare against our model's performance. In addition to specific biological applications, the super-resolution CNN (SCRNN) published by Dong et al demonstrated the ability to use a simple network architecture to upsample low-resolution and quality images[6]. Given the simplicity of this network's architecture—its biggest strength—and reasonable performance, we opted to build the SuperMIBI framework from this model.

Dataset and Features

Our dataset comprised three X and Y pairs of images, all of which had a size 1024 x 1024 x 36. However, for our proof-of-concept SuperMIBI network we opted to train and test over two out of the 36 channels collected. The input data (X) was collected at one-fourth the image acquisition time of the paired output (Y). Moreover, the output images were denoised with our lab's intensity-based k-nearest neighbor denoising algorithm[3]. An example of a paired X and Y example for one channel measured, CD45, a protein on the membrane of immune cells, is shown in **Figure 1**. Each of the input images was cropped into images of size 128 x 128 x 2 with a stride of 43. Next, we assigned 95% of our data set for training and the remaining 5% for testing. During training we performed data augmentation with the Keras ImageDataGenerator class by implementing 180^{o} image rotations between epochs of training, as well as vertical and horizontal flipping[7]. As a baseline to evaluate our model against we employed non-local means (NLM) denoising, which at a high-level works by coloring pixels based on the appearance of related pixels[8]. We employed the 'fast-mode' implementation of this noise processing approach provided by scikit-image[9]. We qualitatively compared the NLM denoising output with the input, ground truth, and our model's predicted output.

Figure 1: Training example of short and long acquisition data for one channel.



Methods and Model Training

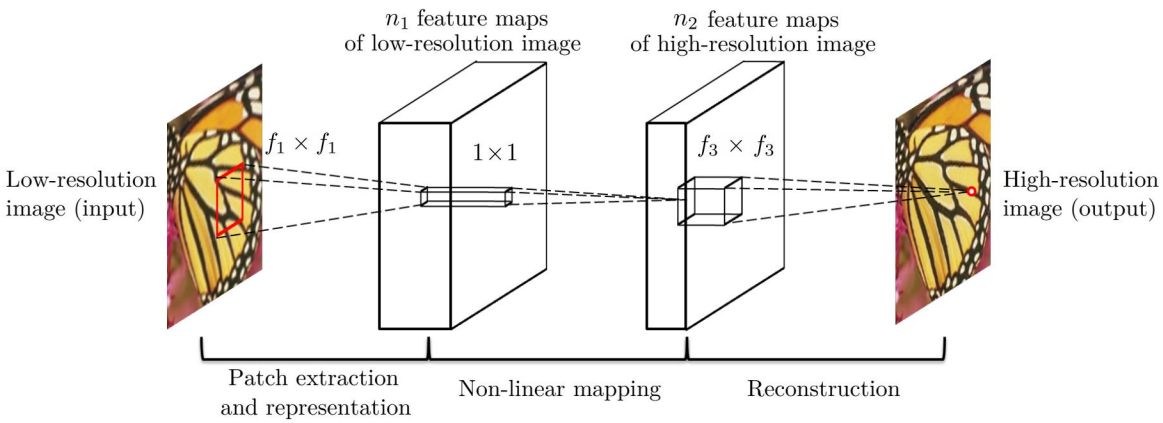


Figure 2 (adapted from Dong et al): SRCNN basic network architecture.

We built the SuperMIBI model architecture from the basic architecture of the SRCNN (**Figure 2**), which has three layers to first extract patch features of the image, perform non-linear mapping, and reconstruct the predicted upsampled RGB image. This network uses convolutional layers with a stride of 1 and 'same' padding followed by batch normalization, and ReLU activation functions. Weights were initialized with Xavier initialization. The loss function used for training the model was mean-squared error (MSE):

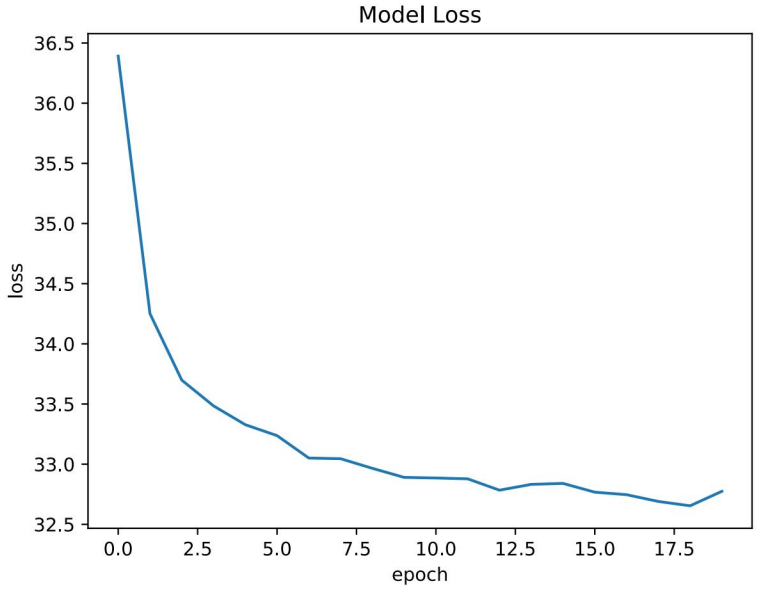
$$L(\Theta) = \frac{1}{n}\sum_{i=1}^{n} ||F(\mathbf{Y}_i;\Theta) - \mathbf{X}_i||^2,$$

The loss was minimized over using standard backpropagation with a learning rate of 0.001 and batch size of 32 with an L2 penalty for weight regularization of 0.01:

$$w_i^{k+1} = w_i^k - \underbrace{2\alpha\lambda w_i}_{L_2 \text{ penalty}} - \alpha\frac{\partial J}{\partial w_i}$$

In the SRCNN paper the authors use filter sizes of 9-1-5 for the three layers, respectively, and recommend using 64 filters for the first layer and 32 filters for the second layer. However, when we used these parameters, we observed poor performance that was improved by 1) Increasing the filter size of layer 1 to 11 x 11 to convolve over larger features in the image and 2) Increasing the total number of filters in layers 1 and 2 to 128 and 64, respectively, which improved the accuracy, though at the cost of increased time for training. One important adjustment we made was to standardize our data by z-scoring on a per channel basis to have a mean of 0 and unit variance. Since the dynamic range of signal varies across channel and between input/output pairs, this step was critical for improving model performance. **Figure 3** displays the loss over training for the final iteration of our model on the dual-channel dataset.

Figure 3: SuperMIBI loss over the course of training.



Results and Discussion

Unlike in traditional imaging applications, where different RGB channels represent the same underlying features, the different channels in our image represent completely different cell types, as many cells would be expected to be positive for certain markers and negative for others. Thus, we started by optimizing our network on a single channel to avoid this confounding behavior (**Figure 4a**). We found that the filter size of the first feature had a dramatic impact on the output of the network, with small filter sizes not being able to understand the spatial layout of the cells, whereas large filter sizes produced blurry images. We found that our loss plateaued dramatically after 5 epochs, and further training produced little benefit.

We next switched to a more challenging marker, CD45 (**Figure 4b**), which marks only the border of our cells. We found that the network struggled to correctly identify the circular shape of these cells. Adjusting the filter size produced different outputs, but none was able to recapitulate the true underlying shape of the cell. We identified a similar plateau in the loss function.

We then integrated these two markers together (**Figure 4c**), in order to see if the network could learn from the easier marker how to interpret the more challenging marker. We hypothesized that the additional information from the cells that were positive for both markers would allow the network to achieve better performance. However, in general we found that incorporating the more challenging marker not only produced poor results on that marker, as in the original single channel evaluation, but that it resulted in poor performance spillover into the channel that had previously been achieving decent performance. This will likely require further optimization in the future in order to circumvent this problem.

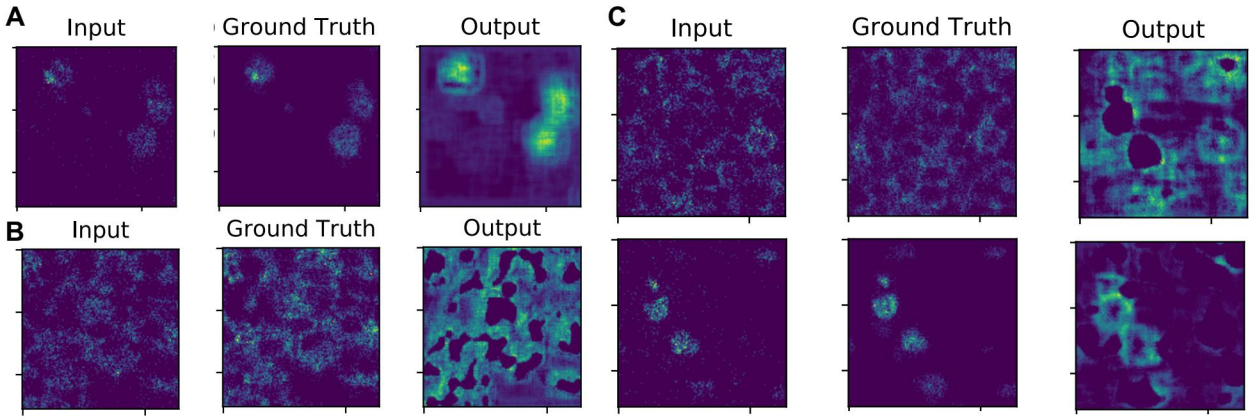


Figure 4: Model experiment output examples. A) Ki67, nuclear protein B) CD45, membrane protein C) Model trained on multichannel (Ki67 + CD45) image input.

Conclusion and Future Directions

Overall our experiments demonstrate the potential to apply CNNs for upsampling MIBI data, but further development is necessary to produce a model that performs this task accurately across both single channels and, ultimately, in highly multiplexed datasets. One of the biggest challenges with our dataset is the sparsity and non-Gaussian distribution of signal in MIBI images. SuperMIBI currently struggles to distinguish signal, which is characterized by higher intensity and spatial density, from noise, which is sparser and more pixelated. This might be improved with a more complex network architecture that can utilizes filters with kernel sizes that can resolve both the granular and global structure of signal in the image. We plan to explore architectures, such as those used in the CARE-CNN. With further tuning of SuperMIBI we plan to implement it into our data collection pipeline in order to improve the efficiency and throughput of biological imaging experiments peformed with MIBI.

Code

Scripts & documentation are available at: https://github.com/ngreenwald/SuperMIBI.

Contributions

Noah worked on implementing the data importation, cropping, and augmentation as well as implementing the model architecture and running experiments. Erin worked on data collection and processing, performing the NLM denoising baseline comparison, implementing visualization scripts, and data normalization/standardization. Both teammates evaluated results of experiments and discussed next steps for implementation and tuning (i.e. modifying network architecture, adjusting filter size or numbers, etc.), and worked on preparing figures and text for the final report and poster.

References

1. Weigert, M., et al. (2018). Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nature methods*, *15*(12), 1090.
2. Angelo, M., et al. (2014). Multiplexed ion beam imaging of human breast tumors. *Nature medicine*, *20*(4), 436.
3. Keren, L, et al. (2018). A structured tumor-immune microenvironment in triple negative breast cancer revealed by multiplexed ion beam imaging. *Cell*, *174*(6), 1373-1387
4. Nehme, Elias, et al. "Deep-STORM: super-resolution single-molecule microscopy by deep learning." *Optica* 5.4 (2018): 458-464.

5. Christiansen, E. M., Yang, S. J., Ando, D. M., Javaherian, A., Skibinski, G., Lipnick, S., ... & Goyal, P. (2018). In silico labeling: Predicting fluorescent labels in unlabeled images. *Cell*, *173*(3), 792-803.
6. Dong, C., Loy, C. C., He, K., & Tang, X. (2016). Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, *38*(2), 295-307.
7. https://keras.io/preprocessing/image/
8. Buades, Antoni, Bartomeu Coll, and J-M. Morel. "A non-local algorithm for image denoising." *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. Vol. 2. IEEE, 2005.
9. http://scikit-image.org/docs/dev/auto_examples/filters/plot_nonlocal_means.html