Deep Learning of Optical Diffraction

Ismail Degani
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology, Cambridge MA
idegani@mit.edu

I. Introduction

Deep convolutional neural networks (CNN's) have been shown to learn highly complex patterns relative to previous approaches [1]. In this project, we show that a CNN can be taught to correctly classify objects that are enshrouded by optical diffraction. This line of inquiry is motivated by the field of digital holography [2,3] where images are captured by a bare CCD sensor without the aid of a lens. Instead of focused images, holographic imaging systems capture diffraction patterns. They then fully "reconstruct" an estimate of the true image by iterative spatial deconvolution methods [4]. This technique can be thought of as effectively replacing a physical lens with a flexible "software" lens. While this yields a myriad of advantages in microscopy [5,6], a major drawback of full reconstruction is that it can be computationally expensive and unnecessary. Therefore it makes sense to wonder whether this reconstruction step can be supplanted by a lightweight CNN classification architecture.

The "D3 platform" [3] is an example of a holographic imaging system where full reconstruction is unnecessary and costly. Figure 1 illustrates the working principle of the technology:

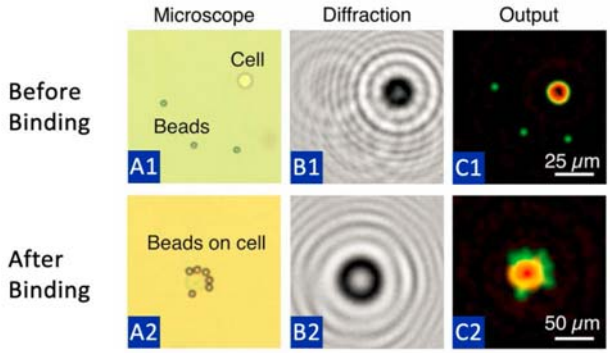


Figure 1: *D3 Diffraction Binding Assay and Reconstruction* [3]

First, antibody-coated beads are constructed to selectively bind to specific surface markers on cells of interest. The sequence A1→ A2 in Fig. 1 illustrates this binding. By identifying and counting the "bead-cell complexes" (A2), one can obtain an accurate quantitative readout for screening purposes. The caveat is that diffraction patterns B1→ B2 are captured, instead of easily discernable brightfield images. The full reconstruction is shown in C1/C2, where the cells and beads now substantially match their brightfield counterparts. This computation requires a dedicated GPU server and it makes the system difficult to use in field settings where WiFi connectivity may be unreliable. In this study, we attempt to mitigate this problem by leveraging deep neural networks.

II. Datasets and strategy

The end goal of this research is to observe the diffraction pattern B2 of Fig. 1, and automatically be able to classify it as a cell with exactly three bound micro-beads. A neural network that performs this task may be computationally expensive to train; however it is hoped that once trained, the network will potentially be capable of fast feed-forward decision-making that far exceeds what is possible with the current state-of-the-art.

To this end, a labeled B-cell lymphoma dataset has been provided by the Center for Systems Biology at the Mass. General Hospital. This dataset consists of eight 5 megapixel images, each containing up to 50,000 cells each. Labeling includes the position and morphology of the cells and micro-beads in the assay. Additionally, the phase transmission information is also captured and used for classification purposes. (Beads have a different index of refraction than cells, which is why they are highlighted in green in C2 of Fig.1).

When designing this research plan, it was deemed advantageous to postpone working with the actual dataset until several simulated datasets were yielding success. The intuition was that having less complexity and a fully controllable characteristics guaranteed some measure of incremental success that could built on. For example, the problem of diffraction classification can be broken down into two independent tasks: (1) image segmentation, and (2) individual cell classification. Each of these has different challenges that may be explored more effectively with tailored, simplified datasets.

Therefore, this project is structured as a series of experiments that attempted to leverage best practices in neural network architectures to solve small challenges on the path toward diffraction reconstruction. We employ canonical 2 and 3 layer convolutional neural networks with multiple fully connected layers, and also experiment with a "specialist" architecture U-Net that is designed to solve specific problems relevant to image segmentation.

III. Repurposed MNIST Classifier

It is worth questioning whether or not a CNN can effectively classify a cell-bead binding event. This requires learning rotational invariance, which unlike translational invariance is not a "natural" capability of a convolutional layer. To confirm this, we attempted to package the problem as an MNIST classification task. We generated 60,000 training examples and 10,000 test examples of random binding events to match the structure and size of MNIST. Cells and beads were constructed to have matched pixel diameters to actual cells (5-7 um), and the distribution of binding occurrences was also made identical to empirical data. About 50% of the cells were left "unbound," and the remaining were constructed with 1-5 beads at random non-overlapping orientations. This resulted in a total of six possible classification labels. We then trained an MNIST classifier on this dataset. Fig. 2 illustrates this dataset and the intuition behind this strategy: the MNIST digits bear resemblance to the "hieroglyphs" of different cell-bead combinations. We hoped to leverage a stable and performant CNN to confirm basic operation before addressing the more challenging aspects of diffraction reconstruction. .

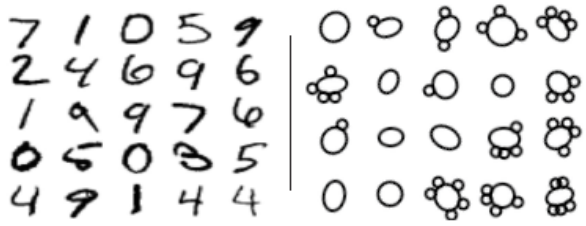


Figure 2: *Cell-binding in comparison to MNIST*

Results: A 2-convolutional layer network with 2 fully connected layers was trained using the Adam Optimizer with a learning rate of 0.001. This architecture was shown to achieve 99.2% test accuracy on MNIST. The network was run for 20 epochs on the simulated cell-binding training set, and managed to achieve **97.61%** accuracy on the corresponding test set.

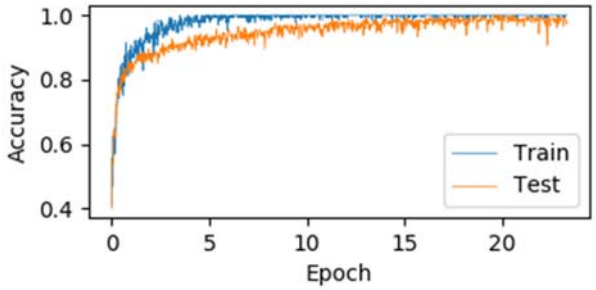


Figure 3 *Two-layer CNN Performance: 97.61% Accuracy*

Because convolutional layers "sweep across" an image to produce activations, they do not have any intrinsic mechanism of learning rotations. Therefore, it is worthwhile to speculate on how this network might be learning rotational invariance. To explore this, we visualize the outputs of the first convolutional layer in Fig. 4. It appears that this convolution output layer has "thickened" the cell walls of the input. Perhaps the network is attempting to homogenize all rotational variants into a single thickness parameter. It could then make a decision solely based on this and avoid the rotational issue altogether.

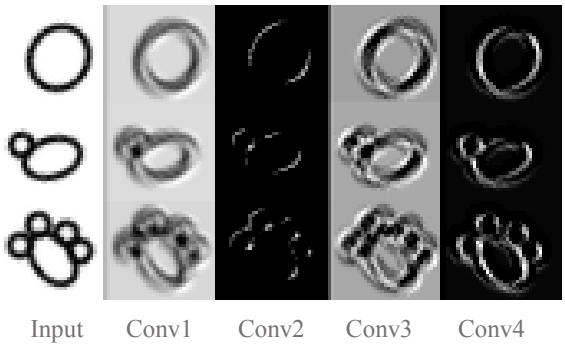


Figure 4 *Output of first convolutional layer*

With the success of this basic classifier, we attempt to discover whether the same convolutional architecture is capable of learning diffraction.

IV. Isolated Diffraction Simulation

Optical diffraction can be expressed as a convolution integral, and several methods exist to compute this transfer function. [4] We implemented the Fresnel propagator, which simulates the effect of diffraction on each simulated cell of our initial dataset. We process the diffraction using simulated light at wavelength λ=405nm, at a z-distance of 0.2mm, and with a pixel resolution of 1.2um. These conditions are substantially similar to what is actually generated by the holographic hardware (D3) whose images we ultimately aim to process. Fig. 5 is an illustration of this transfer function applied to a sample of simulated cells.

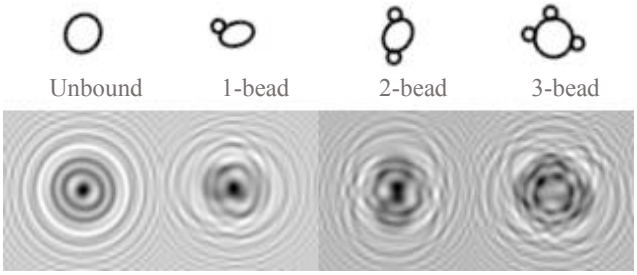


Figure 5 *Isolated Diffraction Dataset*

With this diffracted dataset in hand, we are ready to retrain our network to classify simplified, isolated cells. As shown in Figure 5, a human observer can easily discern the difference between unbound and bound states. The unbound cell exhibits a distortion-free diffraction pattern which is clearly distinct from the other patterns. However, as the number of bound beads increases, it becomes more difficult to classify the cell.

Results: Using the MNIST network, we were unable to achieve better than 75-80% test accuracy on the diffracted datasets. Convolutional filters with greater dimensions (10, 15, 20 pixels) were tried to give each filter a greater receptive field. This was done in an attempt to counteract the spatial "spreading" effect of diffraction. When this did not succeed, additional convolutional and fully connected layers were also added. Still, this did not change the outcome other than vastly increasing the training time. Fig. 6 represents the best performing network, a 3-convolutional layer architecture with increased filter sizes (20x20px) and one additional fully connected layer. It can be observed that the training steps are more volatile relative to the non-diffracted dataset. Additionally, there are discontinuities in the training data at the 12th epoch which were due to the optimizer returning a cost of NaN in a certain batch. This

suggests that there may have been numerical instabilities present in the problem formulation.

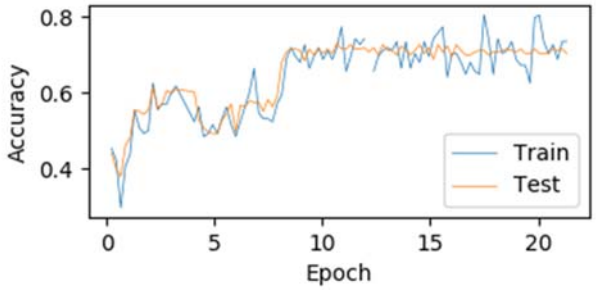


Figure 6*: Performance of Diffraction Classifier*

V. U-Net Segmentation (Undiffracted)

After disappointing performance on diffraction, we proceeded with a new strategy for image segmentation. A specialized architecture was chosen named U-Net [7]. U-Net won the 2015 ISBI challenge for neuronal structure segmentation by a large margin due to its unique layer design. It has no fully connected layers, and instead down-samples and up-samples images through successive shrinking and growing convolutional layers. To implement U-Net segmentation, a new simulated dataset consisting of 2,000 un-diffracted cells-bead "ensembles" was constructed along with a ground truth "mask" representing the perfect segmentation. A sample of this simulated data is shown in Fig. 7. Note that the mask is only activated for cells, and not for the smaller stray bead objects. This means the network needs to learn to ignore stray beads, but activate when they are bound to a cell.

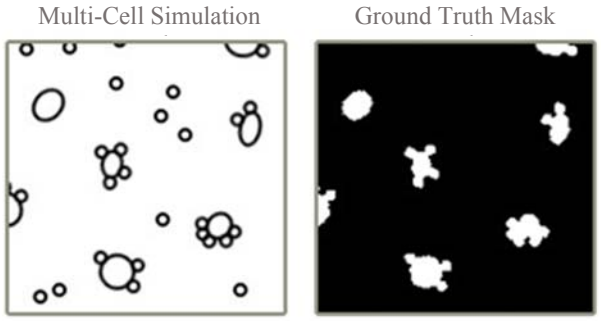


Figure 7 *Simulated "Ensemble" Dataset*

The U-Net architecture was customized to the cell ensemble dataset, where desired output is the mask of Fig. 7: a 200x200 boolean map. Fig. 8, illustrates the chosen 3-layer configuration, and highlights the characteristic "U" that gives the architecture its name. The heart of U-Net is a pipeline of upsampling and downsampling layers. Each downsampling layer consists of two 3x3 convolutional sub-layers ('VALID' padding) that are ReLU activated and maxpooled with a window size and stride of length 2. Each downsampling layer reduces input height/width dimensions by $\left(\frac{n-4}{2}\right)$. It also doubles output layer depth by a factor of two. Once the input has passed through all downsampling layers, it traverses the upsampling layers. Here it is deconvolved (using a transpose gradient) and concatenated with the output of its symmetrical twin in the downsample layer. This concatenation step is the salient feature of U-Net: it allows the architecture to simultaneously incorporate both high and low resolution information into its output classification decision.

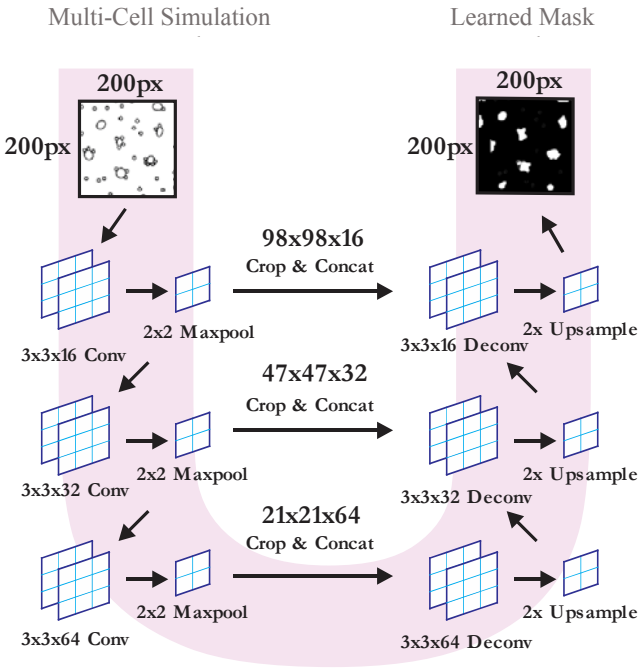


Figure 8: *Custom 3-Layer U-Net Architecture*

Results: The U-Net architecture performed extremely well on the above simplified segmentation problem, achieving 96.48% test set accuracy. Segmentation accuracy is determined pixel-by-pixel based on where the learned mask agrees with the ground truth. A time lapse is shown in Fig. 9 of U-Net learning the mask at each epoch. We can see the micro-bead activations slowly disappear and the learned segmentation start resembling the mask. The momentum optimizer that was chosen in this case is the reason for the step-like decay in learning rate at every epoch.

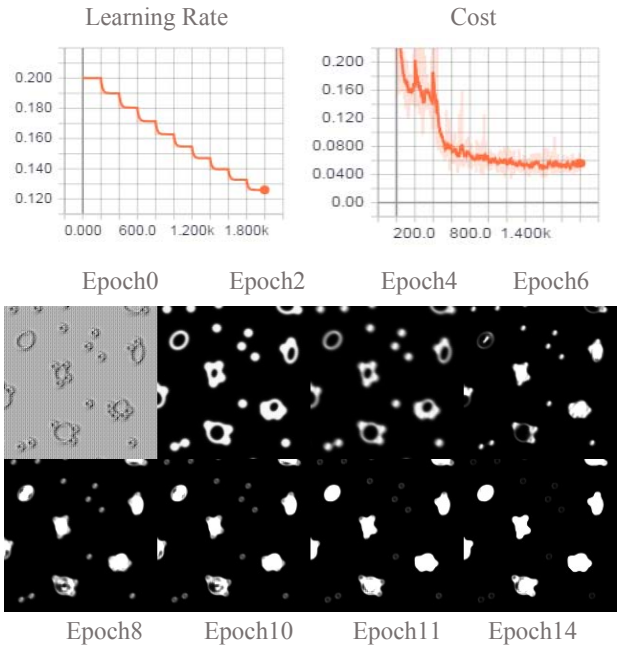


Figure 9 *U-Net Segmentation Map*

VI. Diffraction Segmentation (Preliminary)

The exciting performance of U-Net suggests an efficient, generalized approach to handling diffraction: skip learning the isolated diffraction patterns (as in Fig.5), and proceed directly to

segmentation of an ensemble of diffraction patterns. To accomplish this, we perform a diffraction operation on the dataset of Fig. 7, but keep the same un-diffracted ground-truth mask. This produces a new dataset with new training examples, but the same labels.

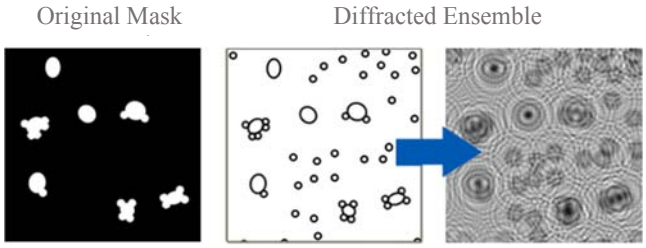
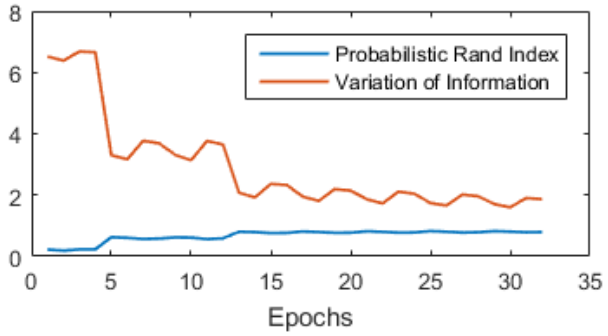


Figure 10: *Diffraction of Ensemble Dataset*

This problem formulation has numerous advantages. It requires minimal intervention on behalf of the designer, unlike the tiered approach explored through much of this project. It also provides a direct path to test actual data far sooner than expected, which is the most important advantage. Holographic imaging data is produced in a format nearly identical to the mask / diffraction input required by U-Net. Therefore it takes minimal processing to begin training on this data.

Results: An identical 3-layer U-Net architecture was trained using the diffraction dataset of Fig. 10. Preliminary results are promising. The accuracy of (~95%) is not particularly meaningful because of the high degree of black pixels in the mask. Due to this asymmetric distribution, a CNN could achieve high accuracy simply by choosing black very often. For this reason, we evaluate the segmentation using Probabilistic Rand Index, and Variation of Information. The PRI increases from 0.23 to 0.80, while VI decreases from 6.52 to 1.87 over the course of about 32 epochs. This suggests an improving classification, though further analysis is necessary to prove that the machine is actually learning to segment diffracted objects.



Qualitatively, it appears that the network is indeed learning the mask of Fig. 7/9/10. It is actively suppressing the diffuse waves of the diffraction, and starting to display high intensity at the center of the diffraction pattern.

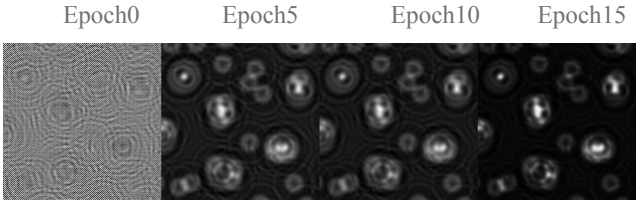


Figure 11: *Preliminary diffraction reconstruction by U-Net*

VII. Conclusion

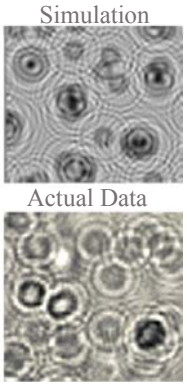
We have analyzed how several CNN architectures handle the challenging problem of diffraction reconstruction. Among all, U-Net provides a compelling architecture that might facilitate both automated segmentation and classification.

While it is too soon to determine whether U-Net segmentation might be a compelling replacement for baseline holographic reconstruction algorithms, it is certainly a viable direction that holds promise.

A valuable lesson learned during this research is that image segmentation is very straightforward using a CNN. Redefining classification in terms of a binary or multi-channel mask is a powerful way to widen the applicability of CNN's to a new class of problems.

VIII. Future Steps

The next logical steps in this project would be to complete a thorough analysis of U-Net's performance on the Mass General Lymphoma dataset. Enhancements could be made to the convolutional architecture itself to add rotational invariance where needed. For example, particular convolutional filters could potentially rotate as they scan across a training set. This might result in more "native" support for rotations with the CNN.



Currently, the simulated data bears very close resemblance to actual holographic data, as seen in the comparison images to the right. Additional improvements could also be made to the simulations so that they even more closely resemble real data. For example, noise, CCD jitter, and other artifacts/debris that are often present in a real biological sample should also be simulated.

Acknowledgments

This work was made possible with generous assistance from the Center for Systems Biology at the Massachusetts General Hospital. In particular, PI's Hakho Lee and Hyungsoon Im provided relevant datasets and valuable guidance throughout this project. I would also like to thank the 6.874 staff for administering this wonderfully productive and enjoyable course.

Resources

Full source code for this project is available at the GitHub link below. Data generation routines for all simulated datasets are included, however the D3 Lymphoma dataset is not publicly available. Access can be requested by contacting the author.

https://github.com/deganii/6.874-Final-Project

References

[1] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classication with deep convolutional neural networks. In: NIPS. pp. 1106{1114 (2012)

[2] Xu W, Jericho MH, Meinertzhagen IA, Kreuzer HJ (2001) Digital in-line holography for biological applications. Proc Natl Acad Sci USA 98(20):11301–11305.

[3] Im, H. et al, Digital diffraction analysis enables low-cost molecular diagnostics on a smartphone., Proceedings of the National Academy of Sciences 2015 112 (18) 5613-5618; April 13, 2015, doi:10.1073/pnas.1501815112

[4] Kreis TM (2002) Frequency analysis of digital holography with reconstruction by convolution. Opt Eng 41(8):1829–1839

[5] L. P. Yaroslavskii and N. S. Merzlyakov, Methods of Digital Holography, Consultants Bureau, New York, London ~1980!.

[6] 5. L. Onural and P. D. Scott, ''Digital decoding of in-line holograms,'' Opt. Eng. 26~11!, 1124–1132 ~1987!.

[7] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation – MICCAI 2015, 234-241. doi:10.1007/978-3-319-24574-4_28