Automated image focus detecting algorithm for low-cost handheld microscope
Machine Learning Project Final Report
Erika Chelales, Riley Deutsch, Roujia Wang

Introduction

Our lab has developed a Pocket Colposcope, a low-cost, portable, trans-vaginal digital imaging tool that provides real-time visualization of precancerous or cancerous lesions on the cervix [1]. The images captured from the device may be processed through an algorithm for prediction of the cervix as precancerous, cancerous, or non-cancerous. While the Pocket Colposcope currently provides high quality images which can be used for identification of cervical lesions, one feature that we would like to develop is a real-time feedback system for the user. The Pocket Colposcope is easy to use, and may be used by the physician or the patient themselves. However, because not all physicians or patients are experts in image quality, sometimes the images captured and provided to the algorithm for diagnosis are not in focus. Out of focus images do not perform well in the algorithm, nor are they easy for a physician to interpret. By implementing a real-time feedback system for the user, we can force our device to only capture in-focus images, thus eliminating the possibility of capturing out-of-focus images which are uninterpretable, or misinterpreted in the algorithm.

Related Work

Obtaining high-quality in-focus images from hardware is an issue for many users and manufacturers, even when autofocus systems are used. A deep learning network developed by Google tackles the classification of microscope images into absolute levels of defocus [2]. The authors used images patches and allowed for prediction uncertainty when no cells or objects were present. The authors used half of their image set for training only, and applied Poisson noise to synthetically defocus the images. Their network consisted of two 5x5 SAME convolutions, followed by RELU nonlinearities, and 2x2 SAME max pooling, with two fully connected layers with a RELU and dropout of 0.5 in between the fully connected layers. On the binary classification task the model achieved an F-score of 0.89, which is improved from the state of the art 0.84 F-score. For the task of identifying absolute defocus, the model is able to predict within one level of true defocus in 95% of the images.

In order to ensure that usable images are acquired, there has been considerable effort to develop deep neural networks for image quality assessment. In 2016, Bosse et al developed a no-reference image quality assessment network that blindly estimates the quality of images without assuming anything about the type of image distortion [4]. This network analyzed images in a patch wise manner with 10 3x3 convolutional layers, 5 maxpool layers, and two fully connected layers. Using a mean absolute error loss function and image quality labels, the network achieved a prediction accuracy of 0.989.

In/out of focus images detection also have useful implementations in field of digital pathology. Slide scanner based digital pathology often produce out-of-focus scans due to unevenness of tissue samples. People use deep learning methods to detect and then correct out-of-focus area in digital scans of pathology slides. Senaras et al group developed a deep-learning based software, DeepFocus, which can detect and segment out of focus areas in pathology scans automatically [5]. DeepFocus is composed of five convolution layers, three max-pooling layers (after the 3rd, 4th and 5th

convolution layers) and two fully connected layers to achieve classification and segmentation on 216,000 images. It achieves 93.2% accuracy, which is 23.8% higher than existing detection methods.

Methods

We have identified a dataset consisting of in-focus and out-of-focus images of cells for use in our project. While these images are not representative of the cervix, they will serve the purpose of identifying an algorithm for our machine learning network which can be modified for use on cervical images when a larger cohort of in-focus and out-of-focus cervical images has been compiled. The dataset we will use is the BBBC006 Human U2OS cells, as shown in ***Figure 1***, (out of focus) images set (https://data.broadinstitute.org/bbbc/BBBC006/). This dataset if defined so that images are labels according to degree of focus. Layers corresponding to 15, 16, and 17 are in-focus, while 0-14 are below the sample focus plane, and 18-30 are above the sample plane and each layer is separated by 2 microns.

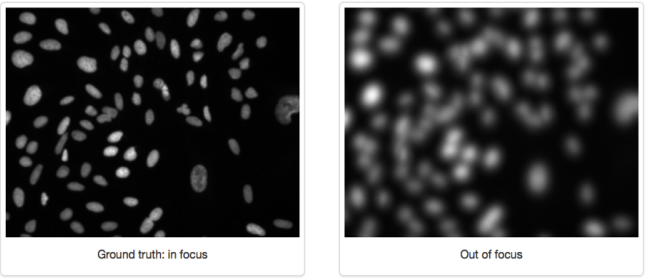


Figure 1: Representative images of Broad Institute dataset. Bottom panel is BBBC006 in/out of focus images of human U2OS cells.

Next, to simulate the physical layer, we would like to modify the image data set by applying a Gaussian blurring filter to a subset of the images which will correspond to the degree that the image is out of focus. We will then apply our algorithm again to classify the images by degree of blurriness. The idea is that we could implement this algorithm real-time so that the device would only capture images when they are in focus. We will quantitatively analyze the accuracy of the classifications of the in-focus and out-of-focus images. We expect to use the U-Net as our CNN for both focus classification and fluorescence segmentation algorithms.

Results

First, we attempted to see if our data had any underlying feature differences that did not require convolutions to separate, similar to the MNIST dataset which can be separated with quite good success by using symmetry and intensity values. However, our data is not symmetrical to we explored a variety of options to classify features of the cell images, with low success, as depicted in ***Figure 2*** for all of the images (1,536 in each set) for the 10 and 15 plane datasets. As anticipated, a convolutional network would be necessary to classify our dataset because the features which distinguish in and out of focus images are best depicted in the Fourier space, which the convolutional network will use.

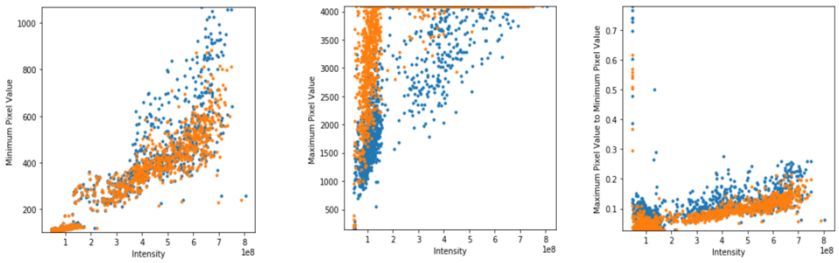


Figure 3: Attempts at separating in focus and out of focus images based on features calculated form the image pixels. From left to right: intensity vs. maximum pixel value, intensity vs. minimum pixel value, and intensity vs. maximum/minimum pixel value ratio.

We adapted the Keras-based CNN from Simonyan, Zisserman et al. 2015, specializing on large image dataset recognition [3]. We tested two types of activation filters: ReLU and Tanh on our dataset for the same network structure, shown in ***Figure 3,*** and both used a learning rate of 0.0001. The results are depicted in ***Figure 4*** for the 10 and 16 image layers. The activation of ReLU seems to perform better more consistently, but both lose accuracy with increasing epochs. When the same network is run on the image layers 05, 10, 16, and 17, but with 500, 400, 300, 200, or 100 randomly selected images from each set, the results are shown in ***Supplemental Figure 1***. It appears that with a larger dataset, the ReLU starts to out-perform the Tanh activation. However, the network appears to still perform quite well with classifying the data regardless of how many images are used from each of the four layers of focus chosen.

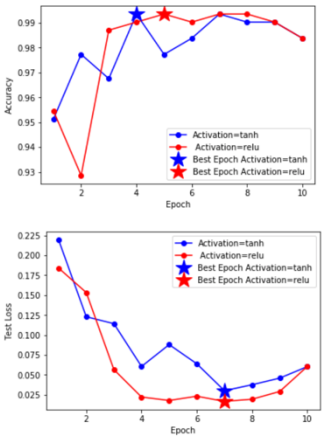


Figure 2: Training accuracy (top) and testing loss (bottom) as a function of epochs for the CNN shown in Figure 3, run on 1000 images from the in-focus dataset (16) and the dataset that is 10 microns out of focus (10)

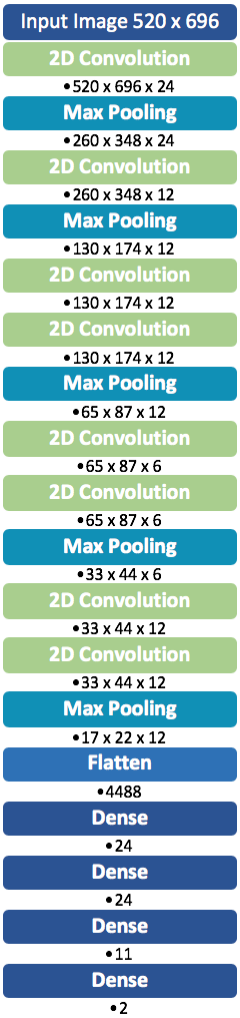


Figure 5: CNN outline for our classification algorithm

To optimize the architecture of the network, we ran a series of experiments to determine how kernel size affects classification accuracy. Training was done on networks with filters of size 3, 5, 7, 9, or 11. The validation accuracy for each of these scenarios can be seen in ***Figure 5.*** Based on the high performance of the 5x5 kernel, we elected to use this size for further experiments.

We then further optimize the computational time and CNN by testing how the testing and training accuracy changed over numbers of parameters with ReLU and Tanh activation (***Figure 6***). A range of 9 different parameters from 5,000 to 1,600,000 are tested with our image dataset for ReLU activation and a range of 50,000 to 3,000,000

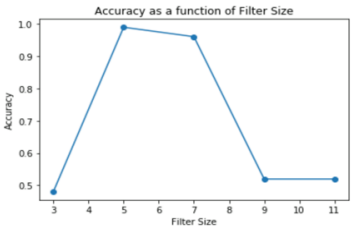


Figure 4: Testing accuracy as a function of kernel size for the in-focus plane (16) and the images from the plane 10 microns out of focus

parameters are testing with Tanh activation. For both activation types, the accuracy of CNN first increased and then sharply decreased as the number of parameters reach above their threshold (1,500,000 for ReLU, 3,000,000 for Tanh). Similar trends of training error and testing error is observed in these runs. Based on our results, ReLU achieve better accuracy than Tanh less parameters needed.

We also wanted to optimize the batch size used in the CNN. When using 1000 images from each image layer (05, 10, 16, and 17) the batch size was varied to optimize the results. The ReLU and Tanh activations with various batch sizes are depicted in ***Figure 7***. Increasing the batch size appears to improve the accuracy of the Tanh activation network, however, the ReLU activation network does not seem to be greatly affected by the changes in batch size.

We also attempted to classify our data based on degree of out of focus of the image. We did this with three types of images, an in-focus group, and two out of focus groups, one more out of focus than the other (10 microns out of focus, and 20 microns out of focus). We explored a variety of options to classify features of the cell images, with low success, as depicted in ***Figure 8*** for 1000 images from each of the 05, 10, 16, and 17 plane datasets. As anticipated, a convolutional network would be necessary to classify our dataset because the features which distinguish in and out of focus images are best depicted in the Fourier space, which the convolutional network will use.

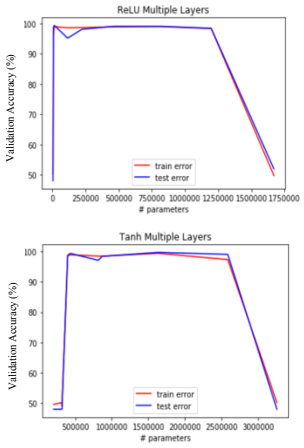


Figure 6: Validation accuracy as a function of the number of parameters used in the CNN for both ReLU and Tanh activations.

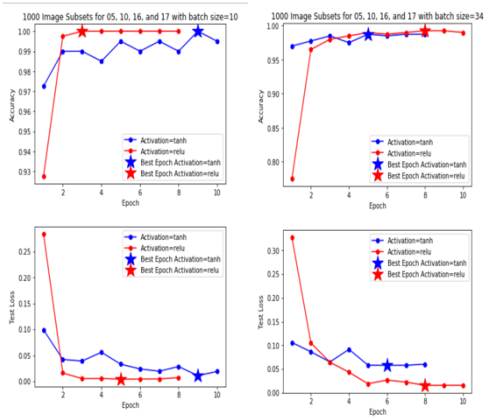


Figure 7: Training accuracy and validation loss as a function of epochs for both Tanh and ReLU activations for a batch size of 10 (left) and a batch size of 34 (right).

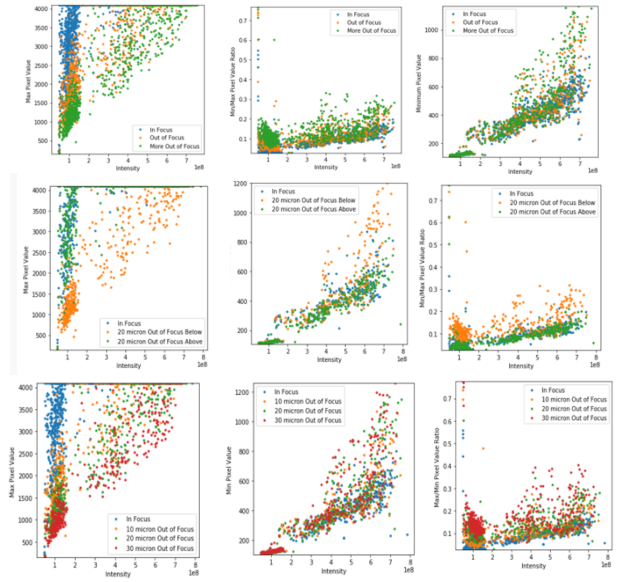


Figure 8: Attempts at separating in focus and out of focus images based on features calculated form the image pixels. From left to right: intensity vs. maximum pixel value, intensity vs. minimum pixel value, and intensity vs. maximum/minimum pixel value ratio. Top row: classifying in focus (plane 16), 10 microns out of focus (plane 10) and 20 microns out of focus (plane 05). Middle row: classifying in focus (plane 16), 20 microns out of focus above (plane 25) and 20 microns out of focus below (plane 10).

Next, we ran the same network as used for the binary classification, but for the three degrees of focus. The task was much more difficult, with accuracies decreased for the Tanh activation more obviously. The network also appears to be slower to train. The ReLU best accuracy was 97.5% and the Tanh best accuracy was 86.5% when run on 500 image subsets from the larger dataset. When run using large subsets (1000 images from each plane), the accuracies improve to 95.3% for the Tanh and 97.5% for the ReLU using the same network. The process was repeated using 4 planes from the dataset: plane 00 (30 microns out of focus), 05(20 microns out of focus), plane 10 (10 microns out of focus), and planes 16/17(in focus). The dataset includes images which are out of focus from above and below the plane of focus. Using the dataset of 16 (in focus), 05 (out of focus, 20 microns below), and 25 (out of focus, 20 microns above). The results are depicted in ***Figure 9.***

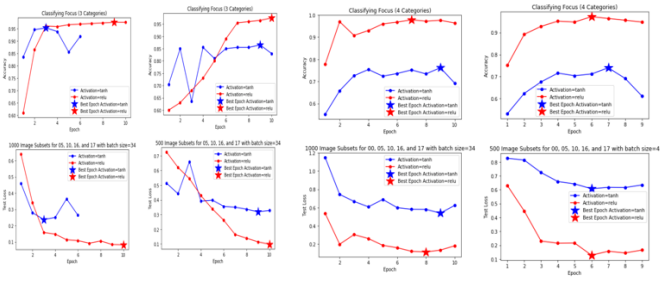


Figure 9: Training accuracy and validation loss as a function of epochs for classifying in focus (plane 16), 10 microns out of focus (plane 10) and 20 micron out of focus (plane 05) with 1000 image subsets (left), and 500 image subsets (middle left), for classifying in focus (plane 16), 10 micron out of focus (plane 10), 20 micron out of focus (plane 05) and 30 micron our of focus (plane 00) for 1000 image subsets (right middle) and 500 image subsets (right).

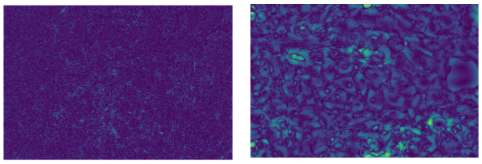


Figure 10: Aperture radius optimization results: Left (r=100) with accuracy of 65%; Right (r=10) with accuracy of 96%

Since the network performed so well with the dataset, the goal of optimizing a physical layer was to reduce the number of convolutional layers while maintaining accuracy. For these experiments, the first three convolutional layers and first two max-pooling layers were removed. It was decided that an aperture acting as a high pass filter would be of interest because the difference in blurry and focused images would be caused by high frequency content. Though the optimization algorithm was not as successful as anticipated, it was found that a high pass filter with a radius of 100 pixels could achieve an accuracy of 96% while a filter with a radius of 10 pixels could only achieve an accuracy of 65% in ***Figure 10***. This allowed us to further optimize the aperture phase using radius of 100 pixels. The initial aperture phase and optimized aperture phase is shown in ***Figure 11***. With increased number of training loops, we are able to achieve a classification accuracy of 99%.

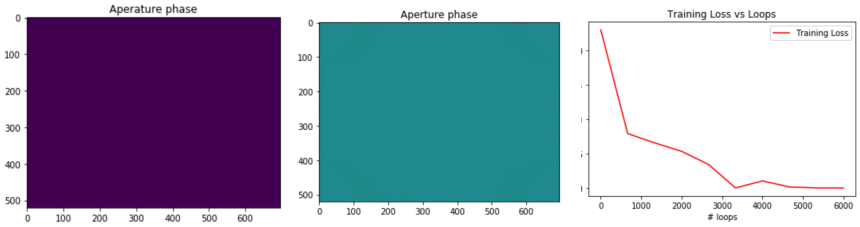


Figure 11: Aperture phase optimization results with radius of 100: Left before optimization; middle: after optimization; right: training loss vs number of loops graph. Accuracy of classification is 99%

Discussion

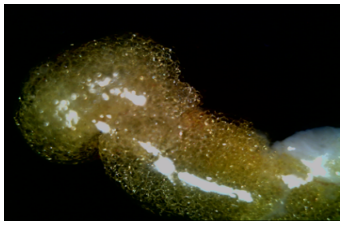


Figure 12: Representative image of anatomical imaging of a clinical breast biopsy using Pocket Mammoscope. The center region is out-of-focus and edges are in-focus due to unevenness of tissue sample.

One limitation of our algorithm is that we used a dataset acquired from a confocal microscope, as a result, the entire image is in the same focal plane because the sample is placed on a slide. However, we would like our system to be used with both microscopic and macroscopic samples. Macroscopic samples could include tissue samples, such as biopsies or cervical images. Thus, while this CNN algorithm performed well, and typically had above 90% accuracy, we should test this on data that does not have as uniform characteristics throughout the image. For example, a biopsy sample as shown in ***Figure 12*** which is an example of the which have a size of 2592x1944 and consists of only a quarter of the biopsy, may have portions of the image that are in focus and out of focus. The entire dataset of these images contains much larger images, which are more complex and have heterogeneous focus across the image. While we wanted to classify the images in a binary manner of in focus or out of focus, this poses an issue. Instead, we may need to consider designing our system to capture images when 90% or more of the image is in focus because it may not be possible to have the entire sample in focus if the surface of the sample is uneven. A future direction may involve optimizing a more sophisticated physical layer that allows in focus images to be acquired for samples that traditionally display heterogeneous focus.

Additionally, our compute engine used NVIDIA Tesla K80, which may not provide sufficient computing power for larger image datasets. When we attempted to run more complex CNNs we often had memory error issues, or found that the CNN was going to take much longer to run than we could afford to allow due to the necessity of iterating our network within the deadlines of the project. We believe with a more powerful GPU we could not only run our network more quickly, but also test our algorithm on more complicated and larger datasets.

In the future, we would like to implement a reinforcement learning technique with our images to have a live-feedback system for the user. We would like to design the algorithm to not only recognize out of focus images and classify them by degree out of focus, as demonstrated by our currently algorithm, but to then adjust the components in our handheld microscope, such as the aperture, to create an in-focus image. The algorithm would also prevent the user from capturing and saving an out-of-focus image. We would of course have a manual override for false-positive cases. If the user believes a photo is not in focus, and the algorithm is classifying it as in focus, we would allow the user to manually adjust the microscope or by adjusting the aperture, to create an in-focus image. The algorithm would still check this image before allowing the user to capture it. Ideally, because we would implement a reinforcement learning algorithm, the feedback system would be live-stream and continuous.

Overall, we have demonstrated the ability to classify images as in or out of focus, and by the degree out of focus that they are and have optimized an aprtue and aperture phase for this purpose. We believe that this CNN could be the basis of a reinforcement learning algorithm that would allow our device to only capture in focus images.

References

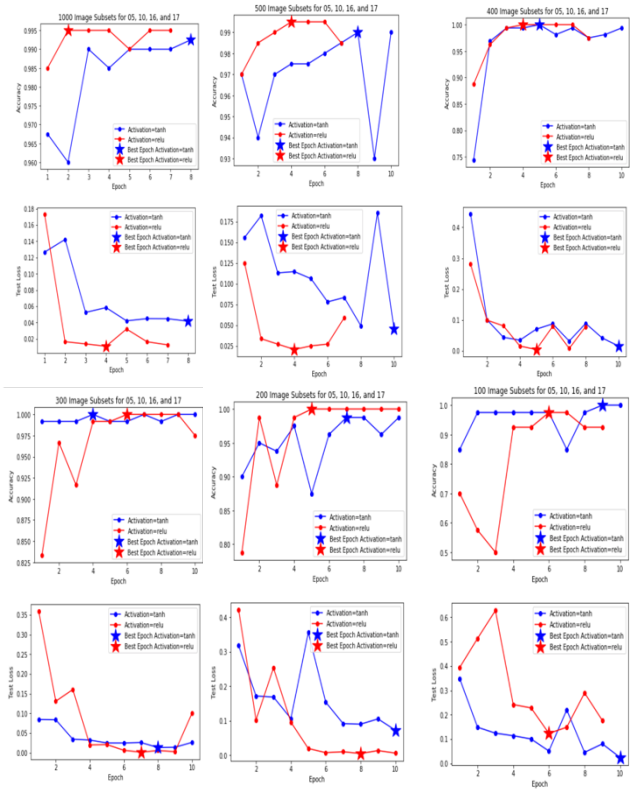
[1] Lam CT, Mueller J, Asma B, et al. An integrated strategy for improving contrast, durability, and portability of a Pocket Colposcope for cervical cancer screening and diagnosis. *PLoS One*. 2018;13(2):e0192530. Published 2018 Feb 9. doi:10.1371/journal.pone.0192530

[2] Yang, Samuel J et al. “Assessing microscope image focus quality with deep learning.” BMC bioinformatics vol. 19,1 77. 15 Mar. 2018, doi:10.1186/s12859-018-2087-4

[3] Simonyan, Zisserman et al. “Very Deep Convolutional Networks for Large-Scale Image Recognition.” *Computer Vision and Pattern Recognition*, 10 Apr. 2015.

[4] Bosse, S., Maniry, D., Wiegand, T., & Samek, W. (2016, September). A deep neural network for image quality assessment. In *2016 IEEE International Conference on Image Processing (ICIP)* (pp. 3773-3777). IEEE.

[5] Senaras C, Niazi MKK, Lozanski G, Gurcan MN. DeepFocus: Detection of out-of-focus regions in whole slide digital images using deep learning. *PLoS One*. 2018;13(10):e0205387. Published 2018 Oct 25. doi:10.1371/journal.pone.0205387



Supplemental Figure 1:Training accuracy and testing loss for Tanh and relu activations for varying sizes of image subsets from each distance of out of focus images and in focus images. In focus images are 16 and 17, out of focus are 05 and 10 where 05 is 20 microns out of focus and 10 is 10 microns out of focus.