Deep Learning for Nuclei Detection and Classification of Colon Cancer Histology Images

(revised May 2019)

Guannan Shen and Yaxu Zhuang, Student Member, IEEE

**Abstract**—Nucleus recognition is an important yet challenging task due to cellular heterogeneity in tumor tissue. Most automated image analysis use a multi-stage processing pipeline to achieve nucleus recognition through several steps including nucleus segmentation, detection and classification. In this project, we learnt ideas from several published papers, including locality sensitive deep learning approach by Sirinukunwattana et al. This is a two-stage approach consists of a nucleus detection and classification. With nucleus detection, the foreground of the image was detected by pixel-wise semantic segmentation using fully convolutional networks U-net. A weighted cross entropy loss function was used for this unbalanced binary classification problem. For the classification, a CNN classifier with a standard softmax function utilizing neighboring information was used to predict the class label of detected nuclei. The accuracy metric was used to evaluate the performance of the neural networks.

**Index Terms**— Convolutional neural network, deep learning, histology image analysis, nucleus recognition.

xxxx-xxxx/0x/$xx.00 © 200x IEEE Published by the IEEE Computer Society

————————————————

1. Guannan Shen is with the Department of Biostatistics and Informatics, University of Colorado Anschutz Medical Campus, Aurora, CO 80045. E-mail: guannan.shen@ ucdenver.edu.
2. Yaxu Zhuang is with the Department of Biostatistics and Informatics, University of Colorado Anschutz Medical Campus, Aurora, CO 80045. E-mail: yaxu.zhuang@ucdenver.edu

—————————— ◆ ——————————

# 1 Introduction

C

ancer is one of the leading causes of death worldwide. According to American Cancer Society (ACS), estimated numbers of new cancer cases and deaths in 2019 (In 2019, there will be an estimated 1,762,450 new cancer cases diagnosed and 606,880 cancer deaths in the United States. Thus, fighting against the cancers is a significant challenge faced by both research scientists and clinic doctors.

Colon cancer is cancer of the large intestine (colon), which is the final part of your digestive tract. Most cases of colon cancer begin as small, noncancerous (benign) clumps of cells called adenomatous polyps. Over time some of these polyps can become colon cancers.

One of the biggest challenges in cancer detection is analyzing different types of tumors at cellular level. There are two common ways for this analysis. The first approach is to use multiple protein markers to mark different cells in cancer tissue. Another way is to use morphological clues in local neighborhoods to develop automated cellular recognition via image analysis.

In this project, we are going to adopt the framework from SC-CNN, a locality sensitive deep learning approach to detect and classify nuclei with two CNN based neural networks, respectively. This method is robust to heterogeneity in nuclei shapes and sizes, without handcrafted features.

There are several differences and advantages over other traditional CNN methods. First of all, this method is based on two properties: (a) The calculation of probability map incorporates distance from nucleus for detecting that object. (b) More accurate labeling of an object is achieved by using weighted entity of local predictions for a class label. Secondly, this approaches only including nucleus detection and classification and it does not require the step of nucleus segmentation which can be challenging due to possible low image quality and complex tissue architecture. Moreover, this approach utilizes more training data than traditional CNN method. It can also localize analysis to small nuclei in image. This is achieved by using a sliding window to train the networks on small patches instead of the entire image (which is usually the case for traditional CNN methods).

After evaluating the difficulty and time reqirement of this project, we decided to use fully convolutional networks Unet to achieve the ideas of this paper.

## Related Works

Adaptive nucleus localization and classification has been a hot topic in computer vison and image analysis. Previously, object detection, semantic segmentation, or regression are normal approaches for this topic. For example, regression random forests were used to predict nucleus centers (P. Kainz et al, 2015). Object counting regression methods have also been applied to nucleus detection.

Traditionally, approaches are heavily based on hand crafted morphological features such as thresholding, region growing, level sets, clustering and graph cuts (Veta, M., et al,2014). For example, one of the famous approaches utilized the difference of Gaussian (DoG) filter for cell detection followed by Hough transform to detect the peaks (Cosatto et al, 2008). In addition, a graph-cut based method initialized by seeds extracted from Laplacian of Gaussian (LoG) filter was employed by Al-Kofahi et al, 2010. After that, some significant techniques were innovated to further this approach, such as:

• Local isotropic phase symmetry for detection of beta cells in pancreas.

• Marker controlled watershed with seeds detected by thresholding.

• Fast radial symmetry transforms to identify nuclei centers.

• Maximally stable extremal regions for detection of nuclei.

• Active contours for cell detection and segmentation.

Modernly, deep learning methods have become the most popular approach for nucleus localization and classification when dealing with large data sets. In particular, CNNs were used for supervised nucleus segmentation, detection and classification in microscopy and histopathology images. One particular advantage of CNNs is: it can learn features that are generalizable to new tasks. At the time of 2016, image representations derived from pretrained CNNs have become the state of the art in computer vision tasks.

The history (from what I learned and google searched) of using CNN to nucleus localization and classification began at 2013, when Cires et al. (2013) have used CNN classification for the detection of mitoses. They used a 12-layer and a 10-layer deep network and achieves processing speeds between 0.01 and 0.03 megapixels per second at prediction time. Janowczyk and Madabhushi (2016) used an 8-layer deep AlexNet (Krizhevsky, 2010) CNN classification to detect lymphocytes in breast cancer images and a f1-measureof 0.900 was achieved in the experiment. Wang et al. (2016) applied a 7-layer deep LeNet classification network to detect nuclei for a subsequent cell subtype classification. Although LeNet is one of the most traditional CNN networks, the investigator can still surprisingly achieve an f1-measure of 0.822.

Moreover, leave out the extraction of local maxima from the PMap is also a common approach for this. For example, Xie et al. (2016) have integrated the PMap over an image region to estimate the nuclei count in that region. They have applied a 9-layer deep network. Xing et al. (2016) have used the connected regions as initialization for nuclei segmentation, in addition to apply a threshold to the PMap.

A 5-layer deep CNN classification was used by Khoshdeli et al. (2017) for the detection of nuclei in Hematoxylin and Eosin stained images of various tissue types. They have proposed to preprocess the input images by extracting the Hematoxylin channel using color deconvolution and applying a Laplacian of Gaussian filter. An f1-measure of 0.722 has been reported. Jacobs et al. (2017) have used a 14-layer deep regression network to detect nuclei in H&E stained prostate cancer subsequent nucleus type classification. The authors have evaluated transfer learning for the application with limited training data. They have trained on colon images and have fine-tuned their model with the prostate images. Depending on the amount of training data for the fine-tuning, reported f1-measures are between 0.849 and 0.864.

# Project Detail

## 2.1 Proposed Method

We planed to follow the framework from Korsuk’s paper and decided to use a two-stage approach. A detection stage is used to detect nucleus in the image and followed by a classification stage which serves to classify detected nucleus.

For detection, we will adopt the idea from *Structured Regression for Robust Cell Detection Yuanpu Xie* and SC-CNN*.* This is a variant of CNN with a structured regression layer as the last layer. We hope to predict the probability of a pixel being the center of a nucleus by this CNN. Spatially constrained might also be added to increase the reliability of the result. Compare to other methods that do not enforce the pixels close to the center of a nucleus to have a higher probability value than those further away, taking spatial constraints into account can guarantee that high probability values are concentrated in the vicinity of the centers of nuclei. For an image patch d ✕ d ✕ c extracted from c-channel color images and centered at (u, v) of image I, the proximity mask M can be calculated by:

= 1/(1+αD(i,j)) if 1+αD(i,j) < R and 0 otherwise.

where D(i,j) represents the Euclidean distance from pixel (i, j) to the nearest human annotated cell center, r the distance threshold and ɑ the decay ratio need to be set explicitly. The activation function of the last regression layer is chosen as the sigmoid function. Thus, a loss function with a weight term to penalize zero area will be used to minimize the sum of squares between the output and the label.

For classification, we will adopt the idea of neighboring ensemble predictor (NEP) method from SC-CNN. This predictor based on spatial ensembling leverages all relevant patch-based predictions in the local neighborhood of the nucleus to be classified, which has been shown can produce more accurate classification results than its single-patch based counterpart. Basically, a set of neighboring patches centered within a radius dβ of the center (u, v) will be used. The objective function is the typical loss function of the softmax CNN.

## 2.2 DataSet

The dataset contains two types of data: first, 30 H&E stained histology images of colorectal adenocarcinomas, which all have a common size of 500 ✕ 500 pixels; second, 30 corresponding annotation mat files, which all contain coordinates and the labels of nuclei. Nuclei annotation was done manually by pathologist. Nuclei were assigned with a class label out of 4 categories: epithelial, inflammatory, fibroblast, and miscellaneous. This dataset is a subset of 100 H&E stained histology images. For the original dataset, there are in total 7 722 epithelial, 5 712 fibroblast, 6 971 inflammatory, and 2 039 miscellaneous nuclei SC-CNN [5]. Thus, the classes of nuclei are not balanced. Within this subset, 4 images: img068, img084, img092, img100 were selected as the validation set, others will be used for training.

Training data augmentation method, initiation and training of networks methods will be adopted from the SC-CNN paper. For data preprocess, all image patches will be rotated (0 degree, 90 ,180, 270) and flipped along vertical or horizontal axis, for both networks. To make softmax CNN robust to the variability of the color distribution, the training patches will be arbitrary perturbed in HSV space, where hue (H), saturation (S), and value (V) variables were separately multiplied by random numbers rH 0.95 1.05 and rS, rV 0.9 1.1, respectively. In addition, we will extract multiple patches of the same nucleus at different locations to account for location-variant and for over-sampling.

In general, all initial weights will be chosen from Gaussian random numbers with mean 0 and standard deviation 10-2 and all initial biases will be set to 0. Both networks will be trained by stochastic gradient descent with momentum 0.9 and weight decay 10-4 for 120 epochs. The learning rate will be 10-2 for the first 60 epochs, then 10-3 for the next 40 epochs, and 10-4 for the last 20 epochs.

## 2.3 Updated Method

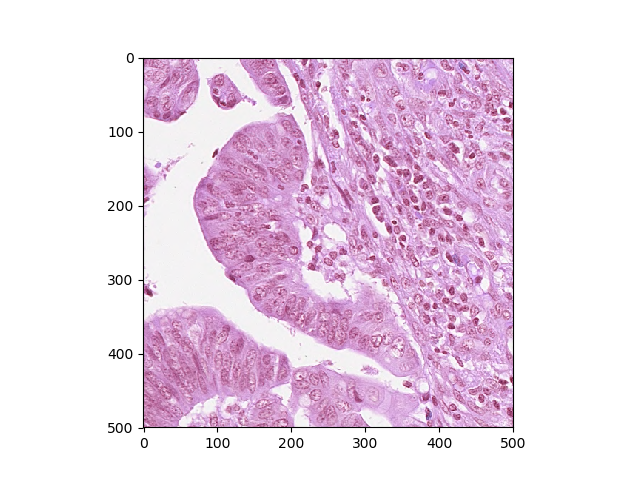
Since we have faced some difficulties in our project and due to the time and other limitations, we were not able to fully achieve methods that we planned to use in our proposal.

In general, the model should be able to detect and classify nuclei in the H&E staining microscopy image, as shown in Figure 1 A and B. Figure 1A is the raw image, and Figure 1B is the image with nuclei labeled as Red, green, blue and yellow for epithelial, fibroblast, inflammatory, and other nuclei, respectively.

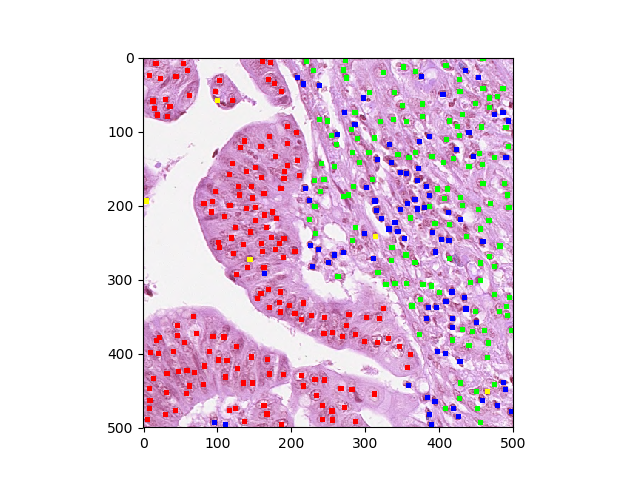
The whole process was divided into two stages and two independent convolutional neural networks were trained. At the first stage, U-net architecture was applied to conduct pixel-wise semantic segmentation. As a result, the foreground was detected from the background by the model trained with binary masks. At the second stage, a CNN classifier was trained by image patches of 4 types of nuclei, through transferring learning using pretrained weights of GooGleNet Inception V3.

In Detail, the dataset contains 30 images and corresponding annotation file including coordinates and labels of nuclei. 5 out of 30 images were chosen as test files, while others as training and validation dataset. The annotation file can be transferred as mask as shown in Figure 1C. Obviously, this is an unbalanced learning problem and the number of positive events is much less than the number of negative events. To tackle this problem, dilation operation was used to enlarge the area of nuclei. As shown in Figure 1 D and E, not only the center of nuclei was labeled but also the surrounding area of the center. The binary mask image shown in image E was used as the ground truth for the U-net architecture training at the stage 1. At this step, the U-net model was trained from the scratch by 10-fold cross validation using the training dataset. The input data was resized to 384\*384 and normalized to range 0-1 as shown in Figure 2. The images were rotated 45 degrees for data augmentation. The weighted logits cross entropy loss function was used for this unbalanced binary segmentation problem. The positive events were up-weighted by the factor of 8 in the loss function. Exponential Linear Unit (ELU) was used as the activation function. The learning rate decay method was applied. The learning rate was started with 0.001 and decreased by 25% for every 4 epochs. Batch size 16 and 20 epochs were used. Dropout method with dropout probability 0.33 was utilized as the regularization method to prevent overfitting. Batch normalization was applied to the last layer in the model. Adaptive moment estimation (Adam) was the optimization method used here. The loss during the training procedure was shown in Figure 3.

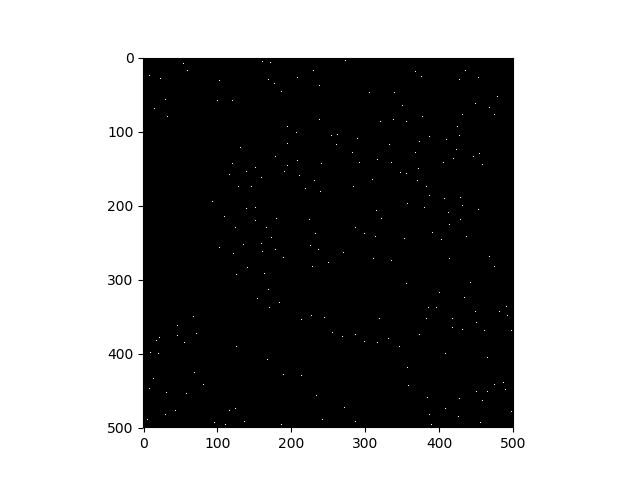
Figure 1. A-E. Given image 009 as the example. A is the original image. B is the example output of the model. C is the raw mask file. D is the up-sampled 4 classes mask. E is the binary mask.



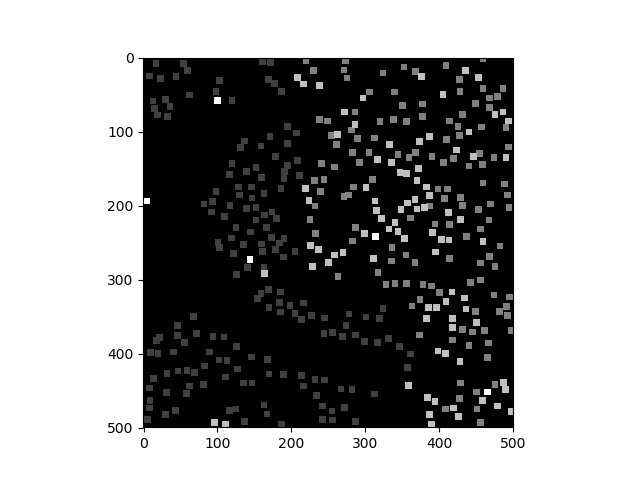
**A**



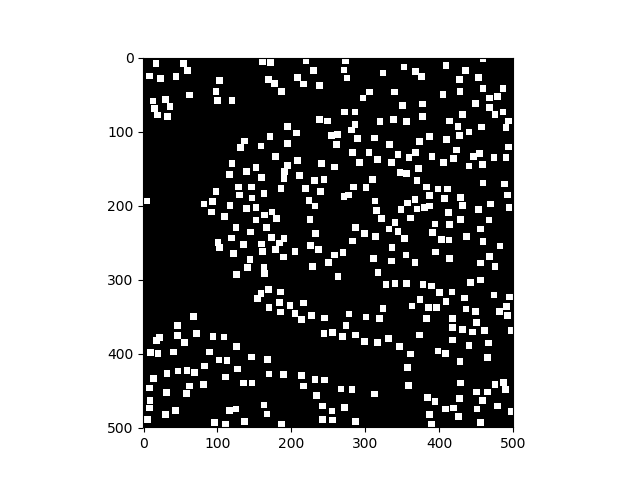
**B**

****

**C**



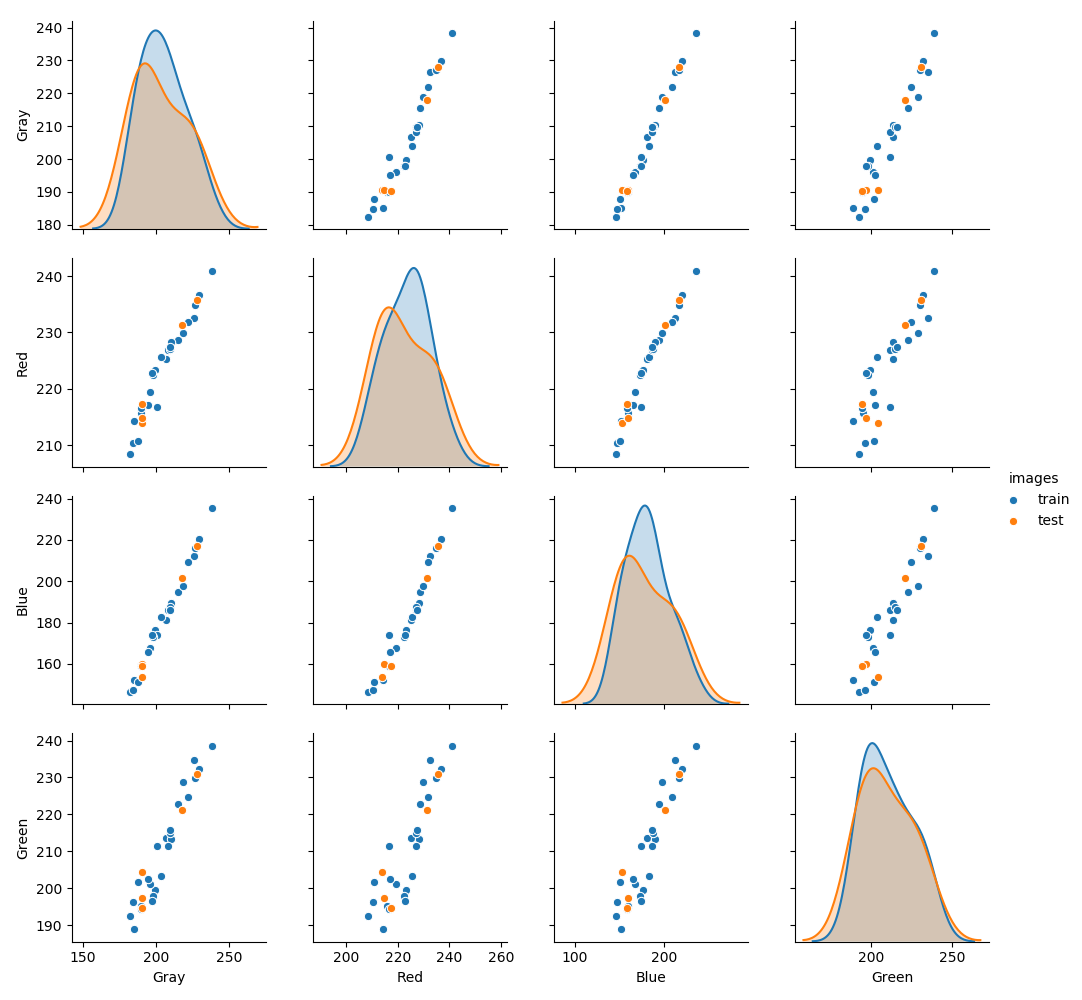
**D**

****

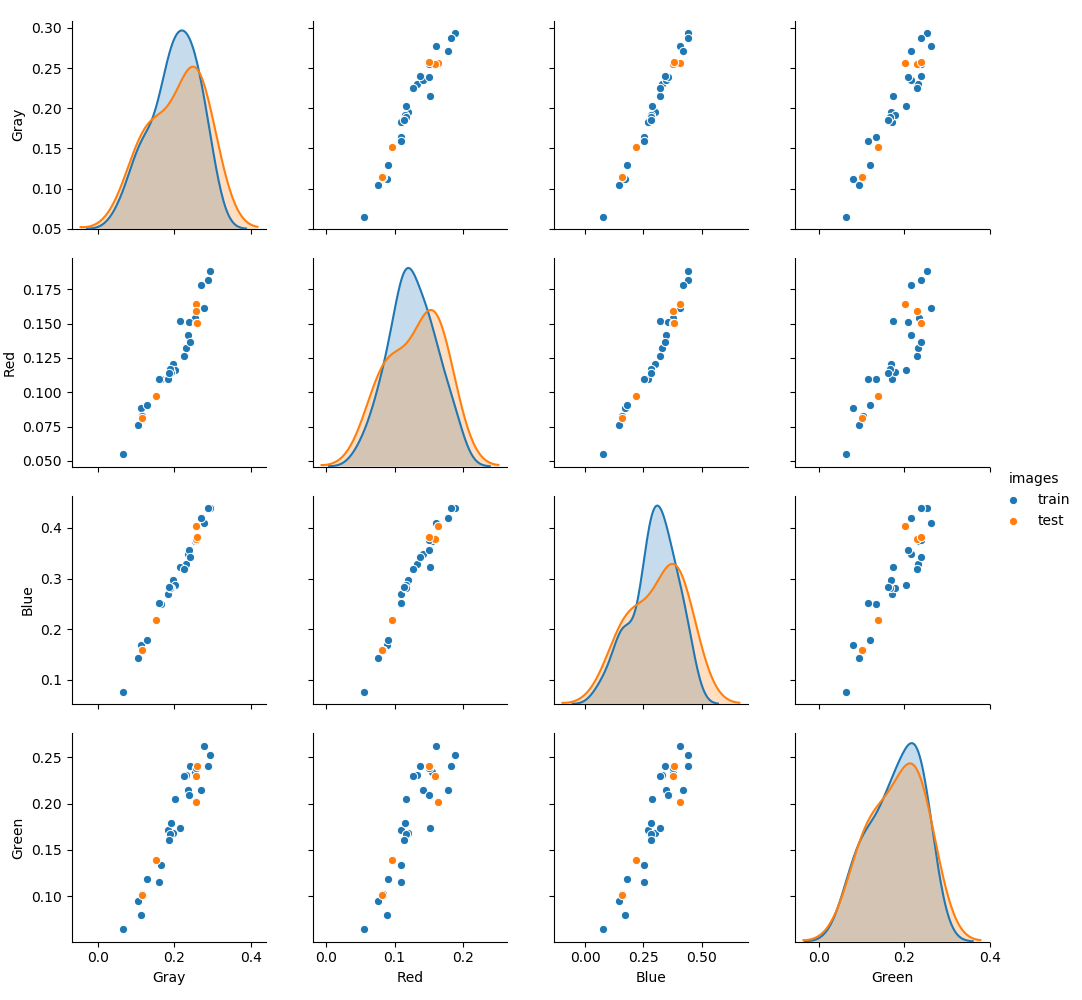
**E**

For step2, transfer learning was used to train a CNN classifier based on pretrained Inception V3 model. The dataset was split to 25 training images and 5 validation images. Only the foreground contains the centers of the nuclei will be used to the model here. Particularly, 27\*27 image patches around the center of the nucleus were extracted to train the model. In total, there are 9450 image patches in training set and 1760 image patches in validation dataset. The image patches were labeled as 1, 2, 3 and 4 for epithelial, fibroblast, inflammatory, and other nuclei, respectively. Label preserving data augmentation methods such as rotation, horizontal flip and shifting were used. All input image patches were normalized as step 1. 4 classes softmax function was used as the loss function. Batch size 32 and epochs 10 were used. As in the step 1, the Adam optimization method, learning rate decay and dropout regularization method were used.

Figure 2 A and B. The intensity values comparison between training and validation dataset for each color channel before (A) and after normalization (B).

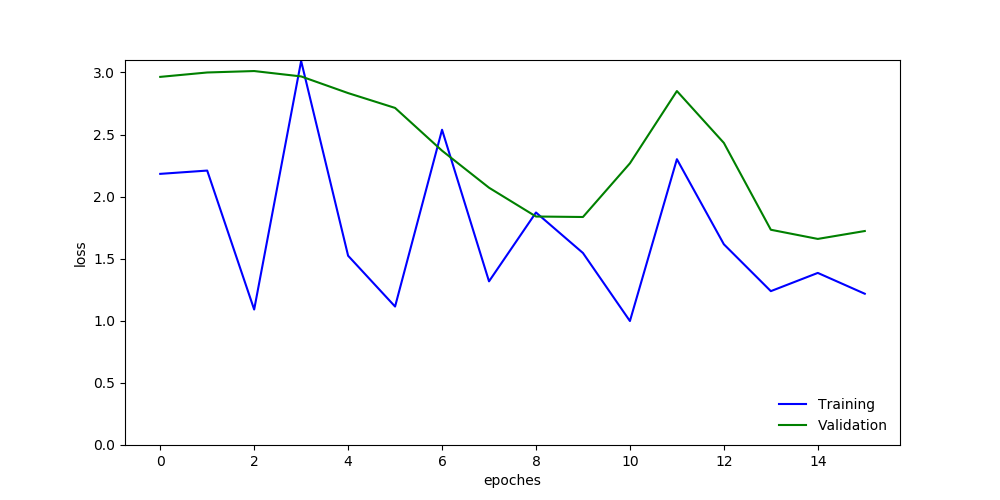


**A**



**B**

Figure 3. The tracking of loss of U-net training. Blue line is the training dataset and green line is the validation dataset.

****

## 2.4 Results

In our proposal, we plan to use Precision, Recall, and F1 score. These were planning to be used to quantitatively assess the detection performance. As the SC-CNN paper, the region within a radius of 6 pixels from the annotated center of each nucleus as its ground truth. Thus, median distance, 1st quartile, and 3rd quartile will also be calculated. We planned to calculate the F1 score for each class of nuclei and their average weighted by the number of nucleus samples to summarize the overall classification performance. Eventually, the weighted average F1 score for combined detection and classification for all class labels will be used. In the combined case, Precision is defined as the proportion between the number of correctly detected and classified nuclei in a class and all detected nuclei classified as that class. Similarly, Recall is the proportion of true positive and the total number of nuclei of that class in the ground truth. At the last, we will also present the average execution time per image.

However, we have faced some difficulties in our project and due to the time and other limitations, we were not able to fully achieve the ideas we had in the proposal.

In general, the task is to detect the center of nuclei from the background and classify detected nuclei into 4 classes: 1, 2, 3, and 4 represent epithelial, fibroblast, inflammatory, and other nuclei, respectively. For the first semantic segmentation stage. The U-net encountered a convergence issue. The loss did not decrease as more epochs were used. For the second step, overall, in our current approach, without a weighted softmax loss function for the CNN classifier, the model tends to classify an unknow nucleus to class 2 with the accuracy of 0.3906 at the end.

## 2.5 Discussion

There are several points could be improved. Better masks might improve the training of the U-net architecture. Transfer learning using pretrained U-net weights might also help resolve the converge issue. The architectures should be modified to allow the input image size as 500\*500 for step 1 and 27\*27 for step 2. More appropriate data augmentation methods should be used to increase the input sample significantly. A weighted softmax loss function should be applied for the CNN model at the second step to tackle the unbalanced learning problem as well. The overlapping sliding window approach can also increase the accuracy of the model by taking the average the probabilities of results from multiple windows. Last but not the least, according to the references, regression method not the classification method can further increase the accuracy of the model.

# 3 Conclusions

Two independent convolutional networks were trained in this work. Firstly, the FCN U-net architecture was trained from the beginning to conduct pixel-wise binary segmentation. However, the model had the convergence issue. For the second stage, an Inception V3 model was trained with transfer learning to conduct multi-classes classification and end up with an accuracy of 0.3906.

**Acknowledgment**

The authors wish to thank Fuyong, Xing. This work is a class project for BIOS 7718 at University of Colorado Anschutz medical Campus.

**References**

1. Hu, Zilong, et al. “Deep Learning for Image-Based Cancer Detection and Diagnosis − A Survey.” Pattern Recognition, vol. 83, 2018, pp. 134–149., doi:10.1016/ j.patcog.2018.05.014.
2. Sirinukunwattana, Korsuk, et al. “Locality Sensitive Deep Learning for Detection and Classification of Nuclei in Routine Colon Cancer Histology Images.” IEEE Transactions on Medical Imaging, vol. 35, no. 5, 2016, pp. 1196–1206., doi:10.1109/tmi.2016.2525803.
3. F. Xing, T. C. Cornish, T. Bennett, D. Ghosh and L. Yang, "Pixel-to-pixel Learning with Weak Supervision for Single-stage Nucleus Recognition in Ki67 Images," in IEEE Transactions on Biomedical Engineering. doi: 10.1109/TBME.2019.2900378.
4. Y. Xie, F. Xing, X. Kong, H. Su, and L. Yang, “Beyond classification: Structured regression for robust cell detection using convolutional neural network,” in Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. New York: Springer, 2015, pp.358–365.
5. Y. Xie, X. Kong, F. Xing, F. Liu, H. Su, and L. Yang, “Deep voting: A robust approach toward nucleus localization in microscopy images,” in Medical Image Computing and Computer-Assisted Intervention- MICCAI 2015. New York: Springer, 2015, pp. 374–382.
6. Shin, Hoo-Chang, et al. “Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning.” IEEE Transactions on Medical Imaging, vol. 35, no. 5, 2016, pp. 1285–1298., doi:10.1109/tmi.2016.2528162.
7. Jacobs, J.G., et al. "Intravascular Imaging and Computer Assisted Stenting, and Large-ScaleAnnotation of Biomedical Data and Expert Label Synthesis". SpringerInternational Publishing, Cham, pp. 77–86, http://dx.doi.org/10.1007/978-3-319-67534-39.
8. Janowczyk, A., Madabhushi, A., 2016. "Deep learning for digital pathology imageanalysis: a comprehensive tutorial with selected use cases. J. Pathol". Inform. 7,29, http://dx.doi.org/10.4103/2153-3539.186902.
9. Ronneberger, Olaf, et al. “U-Net: Convolutional Networks for Biomedical Image Segmentation.” Lecture Notes in Computer Science Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, 2015, pp. 234–241., doi:10.1007/978-3-319-24574-4\_28.
10. Szegedy, Christian, et al. “Rethinking the Inception Architecture for Computer Vision.” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, doi:10.1109/cvpr.2016.308.
11. Xing, F., Xie, Y., Yang, L., 2016. An automatic learning-based framework for robustnucleus segmentation. IEEE Trans. Med. Imaging 35, 550–566, http://dx.doi.org/10.1109/TMI.2015.2481436.
12. Xu, J., Xiang, L., Liu, Q., Gilmore, H., Wu, J., Tang, J., Madabhushi, A., 2016. Stackedsparse autoencoder (SSAE) for nuclei detection on breast Cancerhistopathology images. IEEE Trans. Med. Imaging 35, 119–130, http://dx.doi.org/10.1109/TMI.2015.2458702.
13. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning appliedto document recognition. Proceedings of the IEEE vol. 86, 2278–2324, http://dx.doi.org/10.1109/5.726791.
14. Meijering, E., 2012. Cell segmentation: 50 years down the road [Life sciences]. IEEESignal Process. Mag. 29, 140–145, http://dx.doi.org/10.1109/MSP.2012.2204190.
15. Vink, J., Van Leeuwen, M., Van Deurzen, C., De Haan, G., 2013. Efficient nucleusdetector in histopathology images. J. Microsc. 249, 124–135, http://dx.doi.org/10.1111/jmi.12001.
16. Kost, H., Homeyer, A., Molin, J., Lundström, C., Hahn, H., 2017. Training nucleidetection algorithms with simple annotations. J. Pathol. Inform. 8, 21, http://dx.doi.org/10.4103/jpi.jpi 3 17.