

Segmentation of Nucleus from Histopathology Images

Group Member:

Prantik Howlader AND Arghya Bhattacharya

(Netid: phowlader, 112279275; Netid: argbhattacha, 112276117)

I. OBJECTIVE OF PROJECT

Semantic Segmentation of nuclei in pathology images : Automated nucleus segmentation is a prerequisite for various quantitative analyses including automatic morphological feature computation. However, it remains to be a challenging problem due to the complex nature of histopathology images. In this project, the goal is to suggest techniques to solve the situation where a pathology image is given in the input and the nuclei need to be segmented. For each pixel in the input image, there needs to be a prediction if it belongs to any nucleus or not.

We extended the project scope to apply instance segmentation on the tissue images. As the semantic segmentation problem may have overlap of nucleus, as different nucleus that are overlapping are segmented together in U-net.

II. MOTIVATION BEHIND THE PROJECT

The work in this field has been motivated by the fact that there are significant differences in the nuclear architecture of cancer cells compared to non-cancer cells [1]. Specific tumour types are associated with their characteristic alterations. The Structural alterations in tumour cells include changes in nucleoli and also the appearance of the perinucleolar compartment. The size of nucleus of different types of cells changes in order to maintain appropriate nuclear-to-cytoplasmic volume relationships [2]. Aberrations in nuclear size is associated with certain disease states, most notably cancer, as the size of the Nucleus and its morphology impacts cellular and nuclear functions.

The morphometric and appearance features such as density, nucleus-to-cytoplasm ratio, average size, and pleomorphism are not only useful for finding the cancer grades but also for predicting treatment effectiveness [3] - [6]

A. challenges

Image segmentation techniques: **Otsu thresholding** [7] and marker controlled **watershed segmentation** [8] - [10] assume uniform and distinguishable colors or textures within a nucleus. The problem with this assumption is that in certain pathological conditions (such as hyperplasia or certain cancer subtypes), the nuclei enlarges and there is margination of chromatin, such that the inner body is lighter while there is

a slightly darker outer ring when stained using hematoxylin. Moreover in such conditions , the nucleoli becomes darker than rest of the nucleus. These situations lead to under segmentation or over segmentation [11]. It has been suggested to start with a deep convolutional network to generate a predictive map and based on that the iterative region merging is done. Apart from using Chan-Vese nuclei segmentation, there has been applications of selection based sparse shaped model [12]. Spatially constrained CNN is also applied in cancerous images to classify nuclei where Neighboring Ensemble Predictor has been used to avoid segmentation of the nuclei [13].

III. APPROACH

To perform semantic segmentation using Convolution Neural Network with **U-NET architecture** and **MaskRCNN**. Apart from using the standard architecture of these networks, we modified the training strategy and the architecture if the networks to achieve better results, that we have discussed about in the later sections of this report.

A. U-Net

The architecture consists of a contracting path to capture context and a symmetric expanding path that enables precise localization. [15] Before this work, The typical use of convolutional networks was on classification tasks, where the output to an image was a single class label. However, in many visual tasks, especially in biomedical image processing, the desired output should include localization, i.e., a class label is supposed to be assigned to each pixel, like here, each pixel actually has to be classified whether it is part of nucleus or not. The main idea in [16] is to supplement a usual contracting network by successive layers, where pooling operators are replaced by upsampling operators. Hence, these layers increase the resolution of the output. In order to localize, high resolution features from the contracting path are combined with the upsampled output. A successive convolution layer can then learn to assemble a more precise output based on this information.

It consists of a contracting path (left side) and an expansive path (right side). The contracting path follows the typical architecture of a convolutional network. It consists of the repeated application of two 3x3 convolutions (unpadded convolutions), each followed by a rectified linear unit (ReLU) and a 2x2 max pooling operation with stride 2 for downsampling.

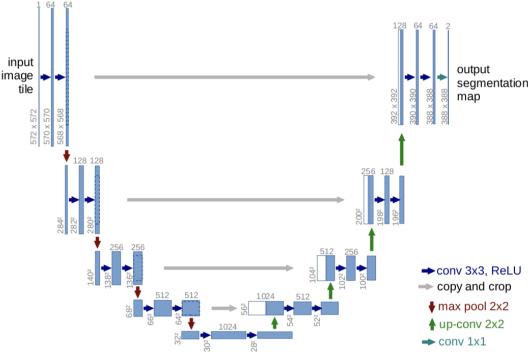


Fig. 1. This is the architecture of U-Net used for Biomedical Image Segmentation.

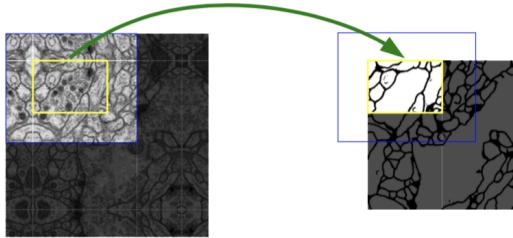


Fig. 2. This is the overlap tiles strategy shown in general. In order to get the segmentation done in the yellow data, we take the blue area as input. Hence, data augmentation is a large advantage in this architecture as mirroring helps to serve this purpose here.

At each downsampling step we double the number of feature channels. Every step in the expansive path consists of an upsampling of the feature map followed by a 2x2 convolution (up-convolution) that halves the number of feature channels, a concatenation with the correspondingly cropped feature map from the contracting path, and two 3x3 convolutions, each followed by a ReLU. The cropping is necessary due to the loss of border pixels in every convolution. At the final layer a 1x1 convolution is used to map each 64- component feature vector to the desired number of classes. In total the network has 23 convolutional layers (shown in Fig. 1). One important modification in the U-Net architecture is that in the upsampling part it has a large number of feature channels, which allows the network to propagate context information to higher resolution layers. As a consequence, the expansive path is more or less symmetric to the contracting path, and yields a u-shaped architecture. The network does not have any fully connected layers and only uses the valid part of each convolution, i.e., the segmentation map only contains the pixels, for which the full context is available in the input image. This strategy allows the seamless segmentation of arbitrarily large images by an overlap-tile strategy (shown in Fig. 2).

Training: : The energy function is computed by a pixel-wise soft-max over the final feature map combined with the cross entropy loss function. The cross entropy then penalizes at each position the deviation.

In deep networks with many convolutional layers and dif-

ferent paths through the network, a good initialization of the weights is extremely important. Otherwise, parts of the network might give excessive activation, while other parts never contribute. Ideally the initial weights should be adapted such that each feature map in the network has approximately unit variance. For a network with our architecture (alternating convolution and ReLU layers) this can be achieved by drawing the initial weights from a Gaussian distribution with a standard deviation of $2/N$, where N denotes the number of incoming nodes of one neuron [?].

a) Modifications We Introduced: As the base U-net has a robust structure that supports data augmentation, we have applied several data augmentation techniques and improving the architecture to improve our result.

- 1) Rotate the image by 90 degrees.
- 2) Divide the image in non-overlapping patches of 128*128.

- 3) **Introduction of Spatial Transformer Networks:** Spatial invariance is achieved within the network by this approach. This is how the model becomes invariant to translation, rotation, scaling or more generic warping or affine transforms. For the localization Net we are using: 2 convolution layers, with batch-normalisation and RELU, followed by two Linear Layers.

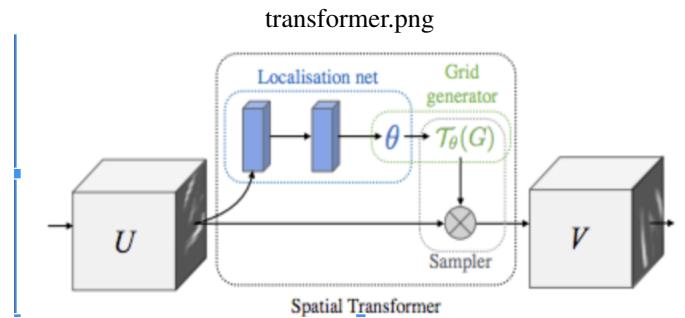


Fig. 3. The structure of Spatial Transformer is shown in this picture.

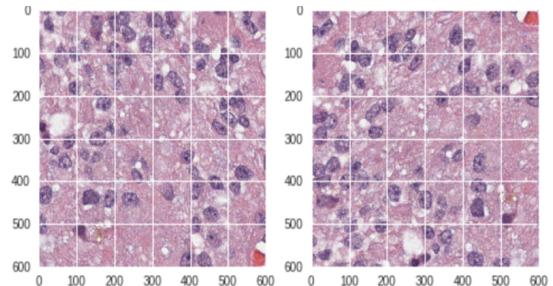


Fig. 4. The original image is at the left and the image rotated at 90 degree is at the right.

We also noticed that for the eosin stains present the image has different color ranges as they come from different manufacturer. So, we have transformed the image from RGB to Haematoxylin-Eosin-DAB (HED) color space. Then we have flexibility to change the ratio and produce same image with different colors.

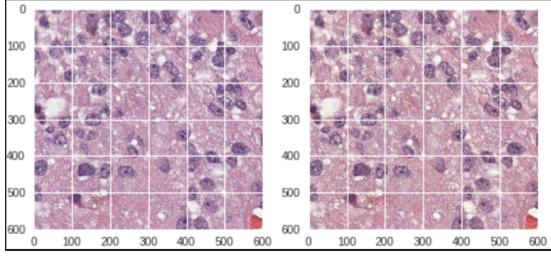


Fig. 5. The original image is at the left and the image after change in HED color ratio is at the right.

B. Mask-RCNN

Mask-RCNN is an instance based segmentation technique. As it requires multiple masks for multiple object instances present in a single image, the technique to enter the mask changes. The architecture of the network is shown below.

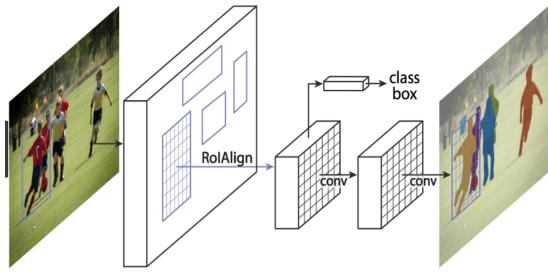


Fig. 6. This is the Mask-RCNN architecture.

A mask RCNN works in two stages.

- 1) First, it generates many Region of Interest(ROI) in the figure around each of the instances.
- 2) Secondly, it generates mask for the ROI.
- 3) The model can be divided in parts. One is RPN (Region Proposal Network) and another is Mask Classifier.
- 4) The ROI-align basically selects all the possible bounding boxes. The boxes are refined using a regression model.

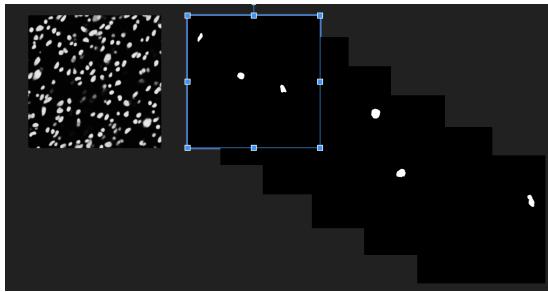


Fig. 7. The image at the left shows the mask for the U-net architecture. It has all the instances present in a single mask. The collection of image at the right shows the masks that are to be used for Mask-RCNN. The multiple object instances are to be trained separately.

Bottleneck of Mask-RCNN and the solution :: Mask RCNN creates mask for every region of interest (ROI) sequentially. So, as the number of ROI increases in a particular image, the running time increases significantly. So, we came up with a solution to these by patching the input images. We tried two

combination of the patching. While the original size of the images are 600*600, we calculated non overlapping patches of size 200*200 and 300*300. The improvement of running time and improvement in the accuracy is for two possible reasons.

- 1) As there is a non-linear relationship between the number of object instances and the running time, the total time per epoch is comparatively very high.
- 2) As we are patching the input images, the ratio of area of region of interest to area of overall image is becoming comparatively too high for the images. So, the network performance is better.

a) Size of Bounding Box :: As the nucleus is of very small size, the selection of bounding boxes have to be small. The base Mask-RCNN chooses the comparatively larger bounding boxes. So, we have modified the RPN anchor scales to be (8, 16, 64, 128, 256) in this range.

IV. DATASET

We used **MICCAI 2016** nuclei segmentation challenge dataset for this Project. It has 32 training images and 32 testing images. For every training image the mask is given. For every testing image, ground truth image is given in the dataset.

V. EXPERIMENTAL RESULTS

We used the dice coefficients, in terms of Dice-Not-Cool (DNC), Dice, Pixel-wise Accuracy for each image. I would also check how much percentage of the ground truth is covered in the predicted patch. If there is error related to over-segmentation (the segmented patches are fractured into several segments, there are increased chances to find boundaries and thus several superfluous boundaries are created) or under-segmentation (the predicted patch covers the nuclei but some significant amount of outer area as well) [14]. A marker-controlled watershed based on mathematical morphology is approached [8] where to further segment under-segmented cells and to merge over-segmented cells, context information among neighboring cells is employed.

The Dice Score is formulated as follows.

$$\text{Dice Score} = \frac{2 * (A \cap B)}{2 * (A \cap B) + (A) + (B)}$$

For the segmentation task it can be written as:

$$\text{Dice Score} = \frac{2 * \text{Number of TP}}{(2 * \text{Number of TP} + \text{Number of FP} + \text{Number of FN})}$$

Jaccard Index or intersection over Union shows as follows.

$$\text{IoU Score} = \frac{\text{mod } A \cap B}{\text{mod } A \cup B}$$

Below is the result obtained by Le-Hou by application of segmentation with thresholding.

Performance Measures	Dice-not-cool	Dice Score	Pixel wise Accuracy
Segmentation with Thresholding	0.6698	0.8317	0.9339

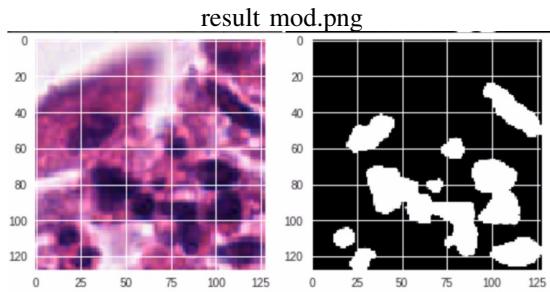
U-net and modified schemes:

The Result for the U-net and with the modified scheme using U-net is described below.

Performance Measure	old	new
Dice similarity score	40.10%	42.72%
Jaccard Index	88.15%	90.32%
Precision	34.80%	45.50%
Recall	56.71%	48.30%
F-score	43.13%	46.85%

result.png

A sample result from modified is shown for a 128*128 patch.



Mask-RCNN and modified schemes:

The segmented image for the Mask RCNN is shown below.

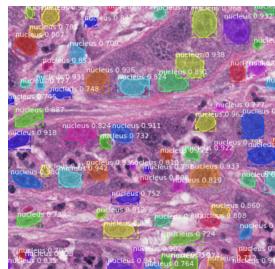


Fig. 8. This is the segmented output

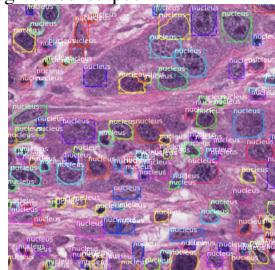


Fig. 9. This is the output of the RPN, the bounding boxes are located

Fig. 10. These are the two images for 600*600 patches

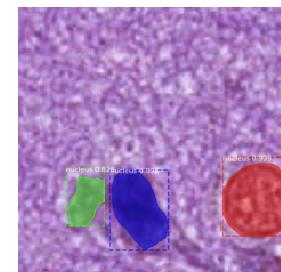


Fig. 11. This is the segmented output

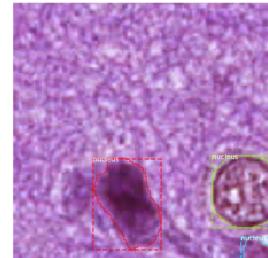


Fig. 12. This is the output of the RPN, the bounding boxes are located

Fig. 13. These are the two images for 200*200 patches

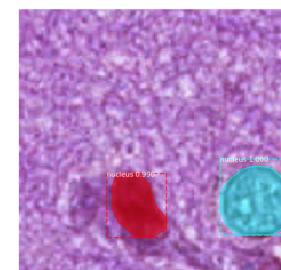


Fig. 14. This is the segmented output

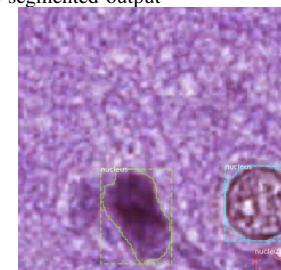


Fig. 15. This is the output of the RPN, the bounding boxes are located

Fig. 16. These are the two images for 300*300 patches

The results and comparisons are explained below.

rcnn result.png		
Performance Table	Accuracy	Avg time of training per epoch
Original Mask RCNN (600*600)	0.713	843.72
Mask RCNN with (300*300) patches	0.748	208.07
Mask RCNN with (200*200) patches	0.743	196.10

VI. HARDWARE USED

We used cuda-capable GPU from Computer Vision Lab for implementing the network, training and testing on the given dataset. We are indebted to Le Hou for giving us access.

VII. CODE SNIPPET USED

We used the code of base Mask-RCNN from the Github link cited below.

VIII. CONTRIBUTION

Prantik Contributed the codes of Mask-RCNN and the modification of U-Net. Arghya did the testing schemes and performance measures and base u-net architecture. Arghya contributed in the presentation and the report. Prantik delivered the major portion of the speech.

REFERENCES

- [1] Zink, Daniele, Andrew H. Fischer, and Jeffrey A. Nickerson. "Nuclear structure in cancer cells." *Nature reviews cancer* 4.9 (2004): 677.
- [2] Vukovi, Lidija D., et al. "New insights into mechanisms and functions of nuclear size regulation." *International review of cell and molecular biology*. Vol. 322. Academic Press, 2016. 1-59.
- [3] A. H. Beck et al., "Systematic analysis of breast cancer morphology uncovers stromal features associated with survival," *Sci. Transl. Med.*, vol. 3, no. 108, p. 108ra113, 2011.
- [4] H. Chang et al., "Invarient delineation of nuclear architecture in glioblastoma multiforme for clinical and molecular association," *IEEE Trans. Med. Imag.*, vol. 32, no. 4, pp. 670682, Apr. 2013.
- [5] P. Filipczuk, T. Fevens, A. Krzyzak, and R. Monczak, "Computer-aided breast cancer diagnosis based on the analysis of cytological images of fine needle biopsies," *IEEE Trans. Med. Imag.*, vol. 32, no. 12, pp. 21692178, Dec. 2013.
- [6] A. Sethi, L. Sha, R. J. Deaton, V. Macias, A. H. Beck, and P. H. Gann, "Computational pathology for predicting prostate cancer recurrence," in Proc. *AACR 106th Annu. Meeting*, Aug. 2015, p. LB-285.
- [7] J.-H. Xue and D. M. Titterington, "t-tests, F-tests and Otsus methods for image thresholding," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 23922396, Aug. 2011.
- [8] X. Yang, H. Li, and X. Zhou, "Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and Kalman filter in time-lapse microscopy," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 53, no. 11, pp. 24052414, Nov. 2006.
- [9] M. Veta, P. J. van Diest, R. Kornegoor, A. Huisman, M. A. Viergever, and J. P. W. Pluim, "Automatic nuclei segmentation in H&E stained breast cancer histopathology images," *PLoS ONE*, vol. 8, no. 7, p. e70221, 2013.
- [10] A. Vahadane and A. Sethi, "Towards generalized nuclear segmentation in histological images," in Proc. *IEEE 13th Int. Conf. Bioinform. Bioeng. (BIBE)*, Nov. 2013, pp. 14.
- [11] Kumar, Neeraj, et al. "A dataset and a technique for generalized nuclear segmentation for computational pathology." *IEEE transactions on medical imaging* 36.7 (2017): 1550-1560.
- [12] F. Xing, Y. Xie, L. Yang, "An automatic learning-based framework for robust nucleus segmentation", *IEEE Trans. Med. Imag.*, vol. 35, no. 2, pp. 550-566, Feb. 2016.
- [13] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. J. Snead, I. A. Cree, N. M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images", *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1196-1206, May 2016.
- [14] Sigut, Fumero, Nunez. Over- and Under-Segmentation Evaluation based on the Segmentation Covering Measure. *WSCG 2015 Conference on Computer Graphics, Visualization and Computer Vision*.
- [15] Olef Ronneberger, Philipp Fischer, Thomas Brox, "U-Net: Convolutional Network for Biomedical Image Segmentation." *arXiv:1505.04597v1* 18th may, 2015.
- [16] Long, J., Shelhamer, E., Darrell, T, "Fully convolutional networks for semantic segmentation" *arXiv:1411.4038 [cs.CV]* 2014.
- [17] He, K., Zhang, X., Ren, S., Sun, J., "Delving deep into rectifiers: Surpassing human-level performance on Imagenet classification" *arXiv:1502.01852 [cs.CV]* 2015.
- [18] Jaderberg, Max et al., "Spatial Transformer Networks" <https://arxiv.org/abs/1506.02025> 2016.
- [19] Matterport, "Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow" <https://github.com/matterport/MaskRCNN>