

Deep Learning Nuclei Instance Segmentation: A U-Net Convolutional Layer Approach

Alexander Swanepoel¹

¹University of Warwick, Computer Science MSc

This paper seeks to undergo implementation of a convolutional neural network to model instance segmentation of nuclei in histological images. We employ machine learning functionalities yielding descriptions of the malignancy of the glandular objects considered in microscopic whole-slide images visual fields. We also demonstrate implementation of the HoVerNet[2] deep learning network.

Keywords: Machine Learning, Convolutional Neural Networks, Segmentation, Colon Cancer, Intestinal Gland, Computational Pathology, Deep Learning

INTRODUCTION

The grading of cancers has served as a fundamental preliminary identifier to the severity of the disease, the variance of the malignancy dictates the course of treatment. Whilst the availability of digitised histological images are pervasive, the exploratory analysis with respect to the tumour microenvironment; ranging from glandular objects to fibroblasts, organelles and the extracellular matrix are computationally unrepeatable due to the variance of morphological features. The repeated accurate analysis of the objects in question would provide pathologists and oncologists alike a precision tool for prognosis and planned treatment and practice. We seek to delve into application of convolutional neural networks and deep learning implementations to provide accurate identification of the organelles and nuclei available in a given whole-slide image, in particular we undergo analysis of the GlaS contest dataset provided by the University of Warwick TIA Centre, initially developed for the MICCAI 2015 challenge. With respect to the glandular object, we consider a cross-sections of intestinal glands, moreover the colonic crypt residing in the epithelium, to highlight architectural deficiencies within these structures, under aim of identification of colorectal cancer.

DATASETS

I. GlaS Dataset

The GlaS dataset consisted of a set of 330 images such that there were 165 ground truth test images, and 165 masks corresponding to each ground truth image. The ground truth images themselves were .bmp extension files with the dimensions {775, 52, 3}, where 3 defines the RBG value; with a colour depth of 24, where the masks had a bit depth of 8 respectively. The content of the image was a digitised Colon histological cross sections, where the cell tissue had undergone hematoxylin and eosin staining, pre-transformation into a whole-slide image as via. usage of the Zeiss MIRAX MIDI Slide Scanner[5], resolved to a x20 pixel magnification (0.620 μ m).

The contents visual fields were analysed by an expert pathologist into the categories of benign or malignant with respect to the glandular structure, with respect to the contest, the histological grades were hidden from contestants to avoid any bias. Moreover, due to the dimensionality, the projection of the 3 dimensional objects into 2 dimensional space raises the issue of apparent separate objects are in reality a single object due to the cross-section capture. The degree of malignancy of the cells were noted to be Tumour-3 and Tumour-4 colorectal adenocarcinoma grades, such that Tumour-3 is in reference to the tumour growth extending into the lining of bowel wall, and tumour-4 is through the wall, respectively.

METHODOLOGY

Model initialisation

We begin initialisation by constructing the convolutional neural network model we seek to implement. The basis of architecture will be U-Net[4], arranged such that the convolutional responses, result in an instance segmentation model. The architecture itself contains two paths, contraction(encoder), and expansion(decoder), on the left and right, respectively. The concatenation of the U-net feature map provides the basis of localisation information. The mutation of the U-net architecture to produce an optimal implementation for the glandular objects insists that the padding of the input layer(top left of 1) is the same, pragmatically, this ensures the input image on the third and final convolution is equivalent to the final output convolution. We employ a 3×3 ReLU(Rectifier) such that the pixels on the edge of the input under consideration is treated as such. Strategic maxpooling was undergone using a 2×2 matrix such that selection of maximum values replace that 2×2 matrix.

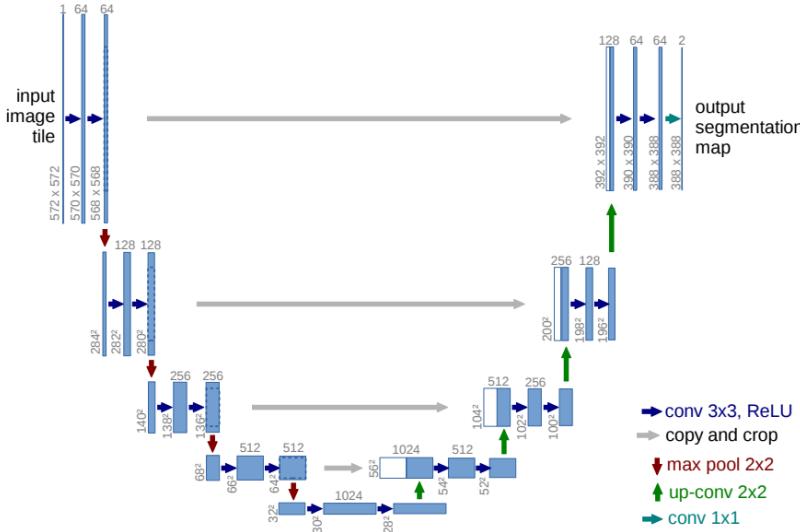


Figure 1. U-Net Convolutional Layer Architecture [4]

Data Preprocessing

Upon progression through the convolutional layers we iterate through and upsample until a final convolution has occurred. Post model construction, data preprocessing is required to satisfy the limits of the architecture. The masks insisted on resizing due to the RGB nature of the image, requiring a binary transform under the `np.putmask` function, allocating a 1 to all non-zero RGB entries of the mask image matrix.

Segmentation

The U-net model is designed for semantic segmentation; as we previously eluded, the dimensionality of the structures being considered indicates that two structures, may be indicated as overlapping, however they are a cross-sectional view of the same single substructure. This insists we implement instance segmentation; the requirements dictate we adjust our initial model by applying a threshold method via the `opencv` module. We then seek to implement the watershed algorithm to overcome the dimensionality issue stated previously, yielding a full instance segmentation model trained on 150 epochs. For intuition we show in 2 that the segmentation yielded non-overlapping accurate qualitative descriptions of the ground truth images, showing clear distinct edges and no intersections on masking.

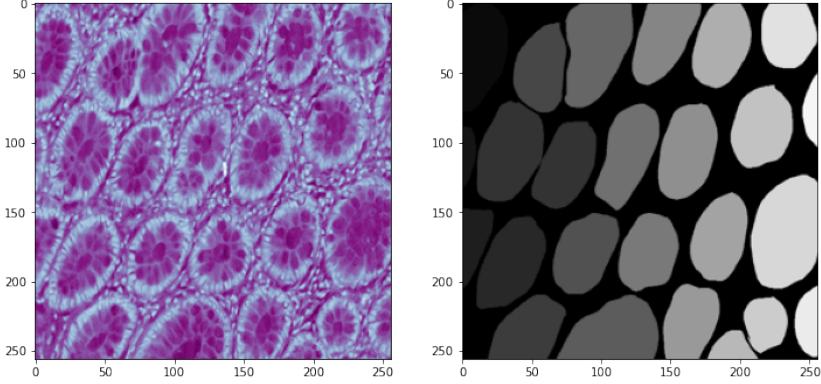


Figure 2. Instance Segmentation Visualisation

Accuracy Metrics

Given the nature of the image, default accuracy metrics do not satisfy the constraints. Intuitively speaking this is due to the observation that the background of an image may for example consist of 99% of the image, whereas 1% of the image is relevant to the analysis, then the accuracy of the image may be calculated as an accuracy of 0.99 however, the nuclei may have all been incorrectly identified. This insists on implementation of Jaccard and Dice metrics. The accuracy index chosen was the Dice; given that the Jaccard score punishes single instance incorrect classification more heavily than the Dice, it can be deduced that the Jaccard is a measure of the worst-case accuracy, with the implication that the Dice is a prediction of the average performance. The Dice index is given by,

$$DICE(A, B) = \frac{2|A \cap B|}{|A| + |B|}$$

such that $\{A\}$ is $\{\text{ground truth}\}$ and $\{B\}$ is $\{\text{segment}\}$ of a glandular object for a given class. Then, let \mathcal{G}_a denote a set of ground truth objects in image a, and \mathcal{S}_a denote a set of segmented objects in image a. Let, $S_i \in \mathcal{S}_a$ denote the i'th segmented object in image a and $G_i \in \mathcal{G}_a$ denote a ground truth object that maximally overlaps S_i in image a. Furthermore, let $\tilde{G}_i \in \mathcal{G}_a$ denote the i'th ground truth object in image a, and $\tilde{S}_i \in \mathcal{S}_a$ denote a segmented object that maximally overlaps \tilde{G}_i in image a. From that, $\mathcal{G} = \cup_a \mathcal{G}_a$ is a set of all ground truth objects. Moreover, $\mathcal{S} = \cup_a \mathcal{S}_a$ gives the set of all segmented objects, n_S denoting the total number of segmented objects in \mathcal{S} . Lastly, n_G denote the total number of ground truth objects in \mathcal{G} . It then follows, the Dice index is given by, We define the object-level Dice index[1] as

$$\text{Dice}_{\text{object}}(G, S) = \frac{1}{2} \left[\sum_{i=1}^{n_S} \omega_i \text{Dice}(G_i, S_i) + \sum_{i=1}^{n_G} \tilde{\omega}_i \text{Dice}(\tilde{G}_i, \tilde{S}_i) \right]$$

where

$$\omega_i = |S_i| / \sum_{j=1}^{n_S} |S_j|, \tilde{\omega}_i = |\tilde{G}_i| / \sum_{j=1}^{n_G} |\tilde{G}_j|.$$

Post analysis, the dice score for the given modified instance segmentation U-Net model was 0.744.

HoVerNet Deep Learning Network

We continued exploratory analysis in the implementation of the HoVerNet[2] Deep learning module, yielding excellent results of the identification of nuclear instance segmentation and classification. The network itself is designed on the premise of a devoted up-sampling branch, with built-in distance consideration functionalities that calculate norms with respect to the centre of mass of the glandular object. Moreover, simultaneously undergoing nuclei identification(classification) whilst also segmentation of the image itself, and is a truly novel technology. Let us delve into the results of the deep learning network, shown in figures, 3,4,5. We may note the excellent quality of the non-overlapping instance segmentation produced by the HoVerNet, with a dice accuracy metric outperforming our initial model by a substantial margin. The values of the dice scores and f1 accuracies can be fully explored in their

documentation <https://github.com/vqdang/hovernet/blob/master/README.md>.[3]. Let us observe the colour delineation denoting blue as epithelial, red as inflammatory, yellow as a spindle-shaped cell, and cyan being miscellaneous, showing the qualitatively the accuracy of classification of the deep learning network.

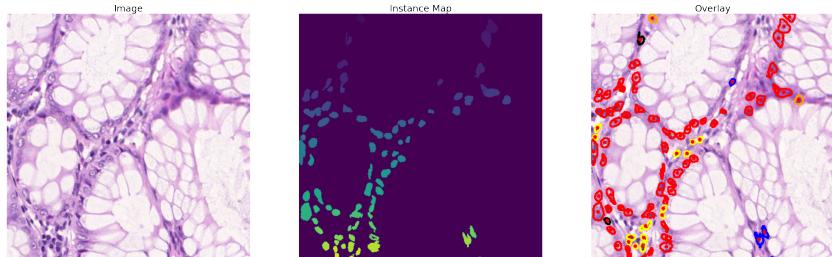


Figure 3. HoVerNet Instance Segmentation and Classification Image 1; Ground Truth Image 1

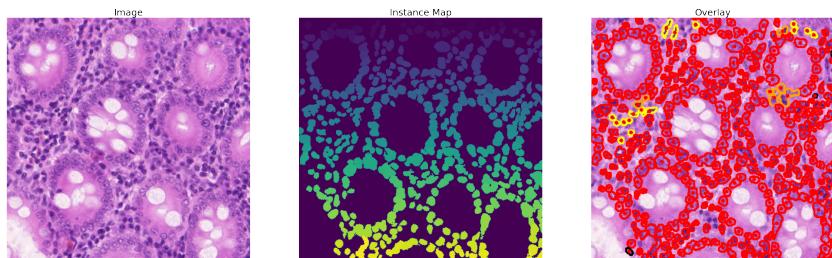


Figure 4. HoVerNet Instance Segmentation and Classification Image 2; Ground Truth image 10

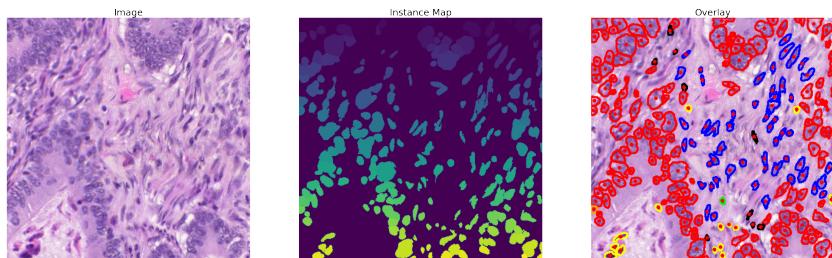


Figure 5. HoVerNet Instance Segmentation and Classification Image 3; Ground Truth image 34

CONCLUSION

In this paper we have formally developed a convolutional neural network under the basis of the U-Net architecture; upon inquisition mutated the architecture via binary transform and thresholding, whilst applying a watershed algorithm to yield an instance segmentation model with respect to our glandular objects. We underwent an implementation of the dice accuracy score, to yield a score of 0.744. We then implemented the HoVerNet deep learning network, providing a basis for whole-slide image classification whilst simultaneously leveraging segmentation.

From the following we are able to conclude that with respect to the GlaS contest entries, the naive model produced with respect to our initial convolutional neural network is significantly outperformed. This can be seen in the table below:

Implementation	Dice Index
Mutated U-Net	0.724
CUMedVision1	0.85
CUMedVision2	0.868
CVML	0.647
ExB1	0.858
ExB2	0.851
ExB3	0.855
Freiburg2	0.852
Freiburg1	0.853
LIB	0.744
vision4GlaS	0.705

Whilst, the U-Net was naive and inefficiently implemented, the model shows significant promise with respect to the contestants, however whilst significantly underperforming with respect to the contestants. From a biological perspective, the model itself, whilst not being efficient or the most accurate, may still serve as a tool in aid of pathologists and oncologists alike; such that the tool may be used to guide decision-making in prognosis and protocol of practice, as opposed to replacing the diagnostic interface completely. The model itself, and all models prior or posterior, should serve to supplicate decision making of the experts in the field, and should provide a more quantitative intuition with respect to the degree of malignance of a glandular object. Possible extensions may include refining the parameters of the U-Net module, and implementing a more rigorous thresholding and watershedding algorithm, providing a stronger basis and would most likely yield a higher dice index.

REFERENCES

- [1] University of Warwick tissue image analytics (tia) centre. https://warwick.ac.uk/fac/cross_fac/tia/data/glascontest/evaluation/.
- [2] S. Graham, Q. D. Vu, S. e Ahmed Raza, J. T. Kwak, and N. M. Rajpoot. XY network for nuclear segmentation in multi-tissue histology images. *CoRR*, abs/1812.06499, 2018.
- [3] S. Graham, Q. D. Vu, S. E. A. Raza, A. Azam, Y. W. Tsang, J. T. Kwak, and N. Rajpoot. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical Image Analysis*, page 101563, 2019.
- [4] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.
- [5] K. Sirinukunwattana, J. P. W. Pluim, H. Chen, X. Qi, P. Heng, Y. B. Guo, L. Y. Wang, B. J. Matuszewski, E. Bruni, U. Sanchez, A. Böhm, O. Ronneberger, B. B. Cheikh, D. Racoceanu, P. Kainz, M. Pfeiffer, M. Urschler, D. R. J. Snead, and N. M. Rajpoot. Gland segmentation in colon histology images: The glas challenge contest. *CoRR*, abs/1603.00275, 2016.