**Title:** Pixel-to-Pixel Learning with Weak Supervision for Single-Stage Nucleus Recognition in Ki-57 Images

**Authors**: Fuyong Xing, Toby C. Cornish, Tell Bennett, Debashis Ghosh, Lin Yang

**Introduction**

Each year, approximately 12,000 individuals are diagnosed with neuroendocrine tumors (NETs) in the United States (1). Ki-67 labeling index has been used a biomarker for the proliferation rate of tumor cells, which determines the tumor grades as well as prognosis of patients diagnosed with NETs (2). Measurement of Ki-67 from pathology images requires correctly counting immunopositive and immunonegative tumors, while ignoring non-tumor cells (3). Hence, it is crucial to recognize and differentiate nuclei and cells of tumor (e.g., immunopositive or immunonegative) and non-tumor cells in the analysis of histopathology images. Common approaches such as “eyeballing” estimation method for Ki-67 counting is labor-intensive, expensive, and often results in large inter- and intra-observer variations (4,5).

To address these shortcomings, machine learning methods (6), especially deep learning (7), have been developed to facilitate quantification of different cell types. Deep neural networks are one of the state-of-the-art approaches that have been employed in medical image analysis and showed improved performance (7,8). However, these previous methods rely on a multi-stage, sequential process that consists of nucleus/cell detection followed by feature extraction and classification. As a consequence of this multi-stage processing pipeline, these methods are usually low-throughput and difficult to apply to large datasets.

Xing et al. proposed KiNet, a novel, single-stage processing approach for nucleus recognition in histopathology images (9). KiNet is comprised of two joint learning tasks. The main model trains the network to predict high values for pixels in nucleus centers using a structured regression model (nucleus identification). An auxiliary task, region-of-interest (ROI) extraction, is incorporated into the main task to further enhance nucleus recognition. The authors also comprehensively compared the performance of the proposed model against other deep learning models.

**Methods**

1. **Model construction**

The network is a modified variant of the recently developed neural network, U-Net (10). The author replaced the conventional convolution connections with residual learning (11) in order to alleviate the risk of gradient vanishing. Furthermore, they applied a multi-context aggregation to fusion of multi-level features for object localization to handle scale variation of nuclei. As a result of these modifications, the model can take any arbitrary-sized image as input and produce an identical-sized output without using a sliding window, which significantly reduces computational burden.

1. **Model evaluation**

The constructed model was evaluated on 38 pancreatic NETs cases, each of whom had three images of size 500 ×500. The image was cropped to represent different tissue appearance from whole slide scanned tissue microarray image data (Ki-67 immunohistochemistry stained). On each image, the position and category of nucleus were labeled, and weak ROI was annotated. The 38 cases were randomly split into two equally sized training and testing sets (each with 19 cases). 20% of (4 cases) the training set were further randomly selected as a validation set. There was no overlap between training, validation, and testing sets. The aim was to localize and correctly classify nuclei into the following three categories: immunopositive tumor, immunonegative tumor, and non-tumor.

The model was trained using stochastic gradient descent with Nesterov momentum (12) with parameters set to the following values: learning rate = 0.01, momentum = 0.9, weight decay = 10-6, batch size = 4 and number of iterations = 105. The training would be stopped if the performance on the validation set did not improve for 2×104 iterations. The model was implemented with PyTorch and trained and tested on a machine with 2,2 GHZ Intel(R) Xeon(R) COU and an Nvidia GeForce GTX 1080Xi GPU.

Model performance was evaluated on two applications: nucleus detection and classification. For each annotated nucleus center, the gold-standard region was defined as the area centered at this annotation with a radius *r* = 16. The automatic detection is considered matched if its detected center was within the gold-standard region. Matched detections are viewed as true positive (TP). Detections that are not matched with any gold-standard regions are considered false positive (FP). Gold-standard regions that did not have matched detections are false negative (FN). Accordingly, accuracy of nucleus detection was evaluated using precision (P), recall (R) and F1score, defined as below: P = TP/(TP+FP), R = TP/(TP+FN), F1 = 2P×R/(P+R). For classification, the three aforementioned matrices were calculated for each class of nuclei and then summarized as a weighted average over all three categories (13).

**Results**

1. **Comparison with other models**

In addition to KiNet, several other deep learning models were evaluated on accuracy of detection and classification, including FRCN (14), SFCNOPI (15), FCRNA (16), FCRNB (16), U-Net (10), and FCN-8s (17). The results were presented in **Table 1**. Compared with other models, KiNet showed the best performance for nucleus detection in the metrices of recall and F1 score. It also had the smallest mean distance between the true positives and gold-standard annotations. Similarly for nuclei classification, KiNet performed the best in the metrices of recall, F1 score, and the area under the precision-recall curve (AUC).

Table

Description automatically generated

1. **Learning with limited and weak supervision**

Chart, line chart

Description automatically generated

**Figure 6** showed the relationship between percentage of fine-grained annotation data and F1 score for both learning with (KiNet) and without (oKiNet) ROI annotation. For this experiment, the training set was generated using all the weak ROI annotation in the training data and a randomly selected subset of fine-grained annotation with nucleus labeling. As shown in the Figure, both KiNet and ookinetes performed poorly when limited fine-grained annotation was used, possibly due to model overfitting. As the percentage of fine-grained annotation data increased, the accuracy of nuclei classification for both methods improved. Collectively, it suggested that strong supervision was needed for nucleus detection and classification.

Additionally, **Figure 6** showed that KiNet outperformed OKiNet at any given percentage of fine-grained annotation. This may be explained by the fact that joint learning improved the ability of the neural network to learn more general representations.

1. **Joint learning vs. stage-wise training**

The author also compared the model performance of KiNet (one-stage) against its sequential or multi-stage variant, sKiNet, and summarized results in **Figure 7**.

**Figure 7** compared the performance of KiNet, sKiNet, and FRCN (baseline model) in detecting and Chart, bar chart

Description automatically generatedclassify nuclei. KiNet outperformed the other two models in terms of recall and F1 score for both detection and classification, suggesting that joint learning is superior to sequential learning. Besides, both KiNet and sKiNet outperformed the baseline model in terms of recall and F1 score, which showed that the proposed network was effective.

**Discussion**

The proposed KiNet model has several strengths. First, it outperforms multiple recently developed deep learning models in nucleus detection and classification. Secondly, the auxiliary task (i.e., ROI extraction), which only requires weak data annotation, further boosts the performance of KiNet. Last but not least, KiNet is a single-stage processing pipeline that allows for high-throughput analysis of large datasets. However, there are also several limitations of the proposed model. First, the model was evaluated only using images from pancreatic NETs patients. Given that Ki-67 has been utilized as a prognostic biomarker in cancers of other organs such as the breast (18), the performance of KiNet in recognizing nuclei from images of breast cancer patients remains unknown. Besides, the authors did not provided details on how the gold-standard annotations were determined. If they were determined by a single pathologist, then there might have been some errors. Lastly, Ki-67 is not only expressed in tumor cells but also in other normal proliferating cells. Thus, its clinical utility as a prognostic biomarker for cancer patients may be questionable if challenges still remain in differentiating tumor versus non-tumor cells.

In conclusion, the authors developed a single-stage processing approach for nucleus recognition in histopathology images that showed competitive performance over other deep learning models. Future work is warranted to address the aforementioned limitations and further optimize the network.

**References**

1. Cancer.Net. 2023 Neuroendocrine Tumors: Statistics. <<https://www.cancer.net/cancer-types/neuroendocrine-tumors/statistics>>.

2. Davey MG, Hynes SO, Kerin MJ, Miller N, Lowery AJ. Ki-67 as a Prognostic Biomarker in Invasive Breast Cancer. Cancers (Basel) **2021**;13(17) doi 10.3390/cancers13174455.

3. Xing F, Su H, Neltner J, Yang L. Automatic Ki-67 counting using robust cell detection and online dictionary learning. IEEE Trans Biomed Eng **2014**;61(3):859-70 doi 10.1109/TBME.2013.2291703.

4. Gurcan MN, Boucheron LE, Can A, Madabhushi A, Rajpoot NM, Yener B. Histopathological image analysis: a review. IEEE Rev Biomed Eng **2009**;2:147-71 doi 10.1109/RBME.2009.2034865.

5. Polley MY, Leung SC, Gao D, Mastropasqua MG, Zabaglo LA, Bartlett JM*, et al.* An international study to increase concordance in Ki67 scoring. Mod Pathol **2015**;28(6):778-86 doi 10.1038/modpathol.2015.38.

6. Sommer C, Gerlich DW. Machine learning in cell biology - teaching computers to recognize phenotypes. J Cell Sci **2013**;126(Pt 24):5529-39 doi 10.1242/jcs.123604.

7. Fuyong X, Yuanpu X, Hai S, Fujun L, Lin Y. Deep Learning in Microscopy Image Analysis: A Survey. IEEE Trans Neural Netw Learn Syst **2018**;29(10):4550-68 doi 10.1109/TNNLS.2017.2766168.

8. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M*, et al.* A survey on deep learning in medical image analysis. Med Image Anal **2017**;42:60-88 doi 10.1016/j.media.2017.07.005.

9. Xing F, Cornish TC, Bennett T, Ghosh D, Yang L. Pixel-to-Pixel Learning With Weak Supervision for Single-Stage Nucleus Recognition in Ki67 Images. IEEE Trans Biomed Eng **2019**;66(11):3088-97 doi 10.1109/TBME.2019.2900378.

10. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors2015 2015//; Cham. Springer International Publishing. p 234-41.

11. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016 27-30 June 2016. p 770-8.

12. Sutskever I, Martens J, Dahl G, Hinton G. On the importance of initialization and momentum in deep learning. In: Sanjoy D, David M, editors. Proceedings of the 30th International Conference on Machine Learning. Volume 28. Proceedings of Machine Learning Research: PMLR; 2013. p 1139--47.

13. Sirinukunwattana K, Raza SEA, Tsang YW, Snead DRJ, Cree IA, Rajpoot NM. Locality Sensitive Deep Learning for Detection and Classification of Nuclei in Routine Colon Cancer Histology Images. IEEE Transactions on Medical Imaging **2016**;35(5):1196-206 doi 10.1109/TMI.2016.2525803.

14. Xie Y, Xing F, Shi X, Kong X, Su H, Yang L. Efficient and robust cell detection: A structured regression approach. Med Image Anal **2018**;44:245-54 doi 10.1016/j.media.2017.07.003.

15. Zhou Y, Dou Q, Chen H, Qin J, Heng P-A. Sfcn-opi: Detection and fine-grained classification of nuclei using sibling fcn with objectness prior interaction. 2018.

16. Xie W, Noble JA, Zisserman A. Microscopy cell counting and detection with fully convolutional regression networks. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization **2018**;6(3):283-92 doi 10.1080/21681163.2016.1149104.

17. Shelhamer E, Long J, Darrell T. Fully Convolutional Networks for Semantic Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **2017**;39(4):640-51 doi 10.1109/TPAMI.2016.2572683.

18. Nielsen TO, Leung SCY, Rimm DL, Dodson A, Acs B, Badve S*, et al.* Assessment of Ki67 in Breast Cancer: Updated Recommendations From the International Ki67 in Breast Cancer Working Group. J Natl Cancer Inst **2021**;113(7):808-19 doi 10.1093/jnci/djaa201.