You Only Look on Lymphocytes Once

Mart van Rijthoven
Radboud University
Medical Centre
mart.vanrijthoven@gmail.com

Zaneta Swiderska-Chadaj
Radboud University
Medical Centre

Katja Seeliger
Radboud University
Donders Institute

Jeroen van der Laak
Radboud University
Medical Centre

Francesco Ciompi
Radboud University
Medical Centre

Abstract

Understanding the role of immune cells is at the core of cancer research. In this paper, we boost the potential of the You Only Look Once (YOLO) architecture applied to automatic detection of lymphocytes in gigapixel histopathology whole-slide images (WSI) stained with immunohistochemistry by (1) *tailoring* the YOLO architecture to lymphocyte detection in WSI; (2) guiding training data sampling by exploiting *prior knowledge* on hard negative samples; (3) pairing the proposed sampling strategy with the *focal loss* technique. The combination of the proposed improvements increases the F_1-score of YOLO by 3% with a speed-up of 4.3X.

1 Introduction

Lymphocytes are immune cells that accumulate at sites of disease in the event of an immune response. In the presence of tumors, it has been shown that the amount of lymphocytic infiltration correlates with clinical outcome [1]. Therefore, quantifying tumor-infiltrating lymphocytes is of paramount interest in cancer research. Immunohistochemistry (IHC) is a staining technique used to highlight cells of interest, such as lymphocytes, in histopathology tissue samples. As a result of IHC, a blue nucleus and a brown rim will become visible in the presence of lymphocytes, but also artifacts with dark regions and brown dots which may look like lymphocytes appear on tissue. Currently, tissue examination and cell density assessment is performed by pathologists, often inspecting glass slides under the microscope. Recently, advances in digital pathology have made high-resolution digitalized whole-slide images (WSI's) largely available, de facto allowing the rise of computational pathology, with the aim of providing algorithms for accurate, objective and reproducible analysis of WSI's.

Following the trend of recent fast development in object detection deep learning technology in computer vision, You Only Look Once (YOLO) [2] comes as an improvement upon Fast and Faster R-CNN [5], resulting in a model that is simpler and faster. Furthermore, additional improvements were presented in [3] (YOLOv2) and only very recently in [4] (YOLOv3). In this work, we leverage YOLOv2 for detection of lymphocytes in histopathology WSI's of breast, colon and prostate cancer, stained with IHC. Although research has been done in applying object detection approaches to medical imaging, to the best of our knowledge, the YOLO architecture has never been used to tackle detection of lymphocytes in histopathology images. Since the problem of lymphocyte detection differs from object detection in natural images both in the appearance of the target objects and in the size of input images, we implement and investigate the effectiveness of novel features in the context of YOLO for lymphocyte detection. First, we tailor the classification model to the problem of lymphocyte detection by simplifying the network architecture. Second, we take advantage of prior knowledge on hard negative samples, typically consisting of staining artifacts and brown background areas, to guide the sampling procedure during training, by applying a simple yet effective analysis in the image color space. Finally, we investigate the effectiveness of combining the proposed sampling strategy with the recently presented focal loss [6]. Inspired by YOLO, we named the proposed architecture "You Only Look on Lymphocytes Once" (YOLLO).

1st Conference on Medical Imaging with Deep Learning (MIDL 2018), Amsterdam, The Netherlands.

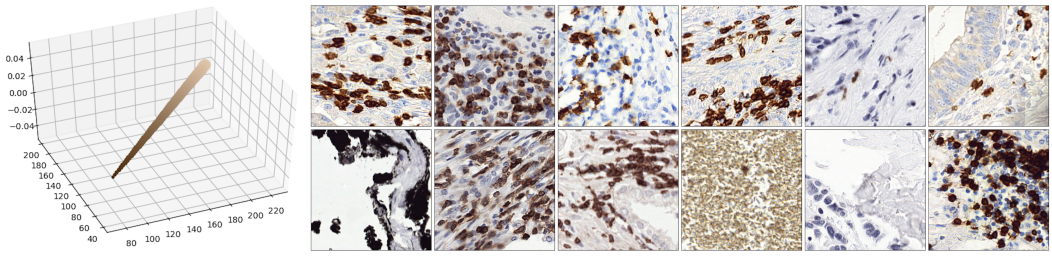

Figure 1: Left: Vector V of RGB values defining a range of brown. Right (top row): a mini-batch of 6 patches sampled without using the sampling strategy. Right (bottom row): a mini-batch of 6 patches including hard negative samples which were enforced through the sampling strategy.

2 Method

YOLLO. The input to the proposed method is an RGB patch of 256×256 pixels extracted at a resolution of ≈0.5 μm/px. As in YOLO, we divide the input image into a $S \times S$ grid. The output of the network is a tensor of size $S \times S \times (B \times 5 + C)$, where B is the number of bounding boxes predicted per grid cell, C is the number of classes, and 5 is the number of parameters predicted for each bounding box (i.e., center coordinates (x, y), (width, height) and confidence score). The average size of lymphocytes is 6-8μm, and although they often tend to form clusters, they never overlap or occlude each other. Based on this prior knowledge, we made several assumptions that allowed to simplify the network. First, we postulated that lymphocytes are covered by bounding boxes of one single size of approximately 12x12 px, de facto getting rid of anchor boxes with specific ratios for different objects. Consequently, we enforced each grid cell to only predict one lymphocyte (B=1) by using a grid size of 32x32, i.e., a grid cell size of 8x8 px. As done in YOLO, non maximum suppression at inference time is used to address overlapping predictions. Finally, because we only want to detect lymphocytes, we set C=1. We further simplify the network by trimming the original 23 layers of YOLO, pre-trained on the Pascal VOC dataset, down to a 8-layer convolutional network, named YOLLO$_{8L}$, which we trained from scratch. Furthermore, we investigate the effect of further reducing the network size to a 4-layer model, which we name YOLLO$_{4L}$. Common to all networks are 3 max-pooling layers, ReLU activations and batch normalization.

Sampling strategy. Sources of difficult negative samples can be identified in WSI areas containing brown areas of artifacts and dots without lymphocytes (see Figure 1). Based on this prior knowledge, a *brown score* $\mathcal{B}_i$ was computed for each training image patch I_i as $\mathcal{B}_i = \sum_{c=1}^{3} \sum_{v=1}^{V} |T(c, v) - \delta < I_i(c) < T(c, v) + \delta|$, where c indicates the three channels in RGB images, T is a look-up table containing V RGB combinations of brown colors, ranging from light brown to black (see Figure 1, left), and $|\cdot|$ indicates cardinality of non-zero pixel values. As a result, $\mathcal{B}_i$ is proportional to the amount of brown in a patch. During training, we sample patches based on the distribution of $\mathcal{B}_i$, allowing YOLLO to focus on difficult negative samples, de facto implementing hard negative mining "on-the-fly", without the need for a two-stage detector.

Focal loss. While the proposed sampling strategy offers YOLLO the possibility of focusing on difficult negative examples, their contribution to learning can be supervised by the *focal loss* [6], which balances sample weights by detecting easy and difficult training examples. For this purpose, we implemented the focal loss as described in [6].

3 Experimental results

Materials. The data consisted of 58 slides and contained breast, colon and prostate tissue. Slides were stained for CD3 and CD8 and were collected from 6 different medical centers in the Netherlands. The 3D-Histech Pannoramic Flash II scanner was used to generate WSI's. 109,841 annotations were made within ROI's containing both sparse lymphocytes and densely distributed lymphocytes as

Table 1: Performance on the validation data. The symbol + indicates training done with guided sampling, ++ indicates training with both guided sampling and focal loss.

Network	Precision	Recall	F_1-score	Speed-up
YOLLO$_{4L}$	0.819	0.653	0.727	**6.8X**
YOLLO$_{8L}$	0.799	0.672	0.730	4.3X
YOLLO$_{8L+}$	0.811	0.693	**0.747**	4.3X
YOLLO$_{8L++}$	0.589	0.751	0.660	4.3X
YOLO	0.717	0.730	0.723	1.0X

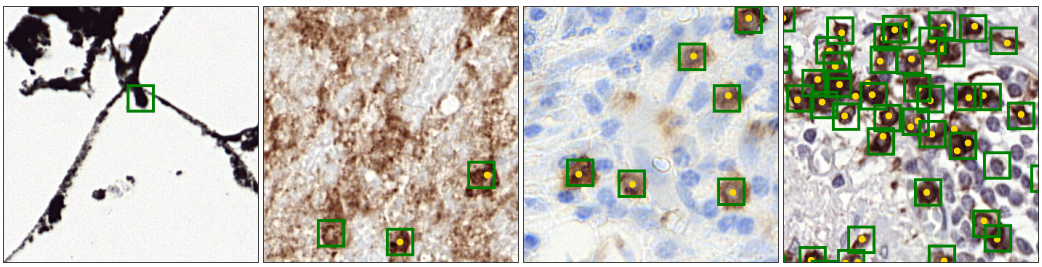

Figure 2: Results on samples from the validation set. From left to right: black artifacts, brown artifacts, sparse lymphocytes and a cluster of lymphocytes. Detections from YOLLO$_{8L+}$ are shown as green rectangles. True annotations are visualized as yellow dots.

well as artifacts. Fig 1 shows examples of these sites. The 58 slides were divided into a training set of 37 WSI's, a validation set of 6 WSI's and a test set of 15 WSI's. The training and validation sets contained WSIs from 2 medical centers. The WSI's in the test set came from 6 medical centers.

Experiments. We compared the performance of the original YOLO, YOLLO$_{8L}$ and YOLLO$_{4L}$ using the data sets described in section Materials in case of (1) YOLO pre-trained on Pascal VOC and fine tuned using data at hand, (2) YOLLO$_{8L}$ and YOLLO$_{4L}$ trained from scratch, (3) presence of the proposed sampling strategy, (4) combination of (3) with focal loss. For training purposes, we extracted patches at a resolution of 0.49 μm/px (20X magnification). In all networks, grid cells that contained objects were weighted 5 times higher than grid cells only containing background. Network parameters were updated with the Adam optimizer and its default parameters, and a learning rate of 5×10^{-5}. During training, we monitored model performance using the F_1 score on the validation set. We also made a comparison in terms of computation time performance at inference time. For this purpose, we ran the network 100 times, and averaging the time it takes to process a mini-batch of 32 samples. In Table 1 we report the performance on the validation set for all the considered approaches. It can be noted that the original YOLO and YOLLO$_{4L}$ achieve comparable results, but with a speed-up factor of 6.8X in favor of YOLLO$_{4L}$. The slightly deeper architecture YOLLO$_{8L}$ achieves an F_1 score of 0.73, which is further improved when the proposed sampling strategy is used for training, achieving an F_1 score of 0.747. Visual results for this approach are shown in Figure 2. Training with focal loss did not result in better performance. Although this was not expected initially, the same effect was reported in the very recent YOLOv3 paper [4]. Consequently, the best performing model on the validation set (YOLLO$_{8L+}$), was used for processing the independent test set, where it scored an precision of 0.83, a recall of 0.6 and an F_1 score of 0.7.

4 Discussion and conclusion

We presented YOLLO, a YOLOv2-based model tailored to detection of lymphocytes in histopathology WSI stained with IHC. The proposed modifications, namely simplified architecture and guided sampling strategy, allowed to gain a speed up of 4.3 with an increase of 3% in detection performance. However, application of focal loss did not increase performance. Improvements due to the guided sampling were mostly observed in ROIs with artifacts and clusters of lymphocytes. Future work could focus on a more in depth analysis of misclassified lymphocytes. With YOLLO$_{8L}$, a gigapixel whole-slide image of 100,000$\times$100,000 pixels can be fully processed in 16 minutes using a GeForce GTX 1080, which can become < 5 minutes when image background is removed and only patches containing tissue are processed.

References

[1] A. J. Gentles et al. *The prognostic landscape of genes and infiltrating immune cells across human cancers* Nature Medicine, 21:938-945, Aug. 2015.

[2] J. Redmon et al. *You only look once: Unified, real-time object detection.* CoRR, abs/1506.02640, 2015.

[3] J. Redmon and A. Farhadi. *YOLO9000: better, faster, stronger.* CoRR, abs/1612.08242, 2016.

[4] J. Redmon and A. Farhadi. *YOLOv3: An Incremental Improvement*, 2018.

[5] Ren et al. *Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks* Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1 (NIPS 2015).

[6] Tsung-Yi Lin et al. *Focal Loss for Dense Object Detection* CoRR, abs/1708.02002, 2017.