

DEEP LEARNING

Automatic Lymphocyte Detection on Gastric Cancer Using Deep Learning

Avinash Chikkayya Kamble, G. G. Yashwanth, Bhanu Kiran, Bhagyashri Wakde, V. Soniya Komal

Department of Computer Science Engineering, Rajiv Gandhi Institute of Technology, Bengaluru, India

Published October 2026

Abstract

Gastric cancer still has rather high rates of incidences around the globe and the amount of tumor infiltrating lymphocytes in the tumor tissues has proved to be a reliable predictor of prognosis in patients suffering from this type of disease. At this point, the manual count of TILs had been performed by pathologists on tissue samples stained and analyzed under microscope. The paper suggests a fully automated pipeline for detecting lymphocytes in the histological pictures of patients with gastric cancer using YOLOv5 detector with ResNet-50 backbone. Namely, the pipeline presupposes slicing the picture into overlapped tiles, colour stain distribution normalization based on the Macenko algorithm, data set augmentation, and providing the slices to detection head that was upgraded by means of feature pyramid network, anchor optimization by k-means and convolutional block attention module, allowing for highlighting the features specific for lymphocytes. Duplications in the tile borders are handled through non-maximum suppression and weighted box fusion. As the data set used for testing is quite small, the precision of detections exceeded 92% and exact values of precision, recall and F1-score should be determined once the training is over. Tumor-infiltrating lymphocytes (TILs) are important indicators in cancer research because their abundance and distribution can be associated with prognosis and other clinical characteristics. For pathologists, identifying these immune cells in gastric cancer tissue images.

Keywords: lymphocyte detection; deep learning; gastric cancer; cell detection; convolutional neural network

1 Introduction

Gastric cancer is one of the most common types of cancer and is a major cause of cancer-related mortality worldwide. Early diagnosis and accurate analysis of gastric cancer are important for improving patient outcomes. Histopathological examination of stained tissue sections remains an important method for studying the cellular composition of tumors and understanding the immune response associated with cancer. Tumor-infiltrating lymphocytes are immune cells that migrate into tumor tissue from surrounding blood vessels. Their presence, abundance, and distribution can provide useful information about the tumor microenvironment and may be associated with treatment response and prognosis. Consequently, accurate identification and quantification of lymphocytes in histopathological images is an important task in computational pathology. Traditionally, pathologists manually identify lymphocytes under a microscope. Although expert examination provides valuable clinical information, manual analysis can be time-consuming, particularly when whole-slide images contain a large number of cells. Dense clusters, overlapping cells, variations in tissue staining, and differences between laboratories can further increase the difficulty of manual and automated analysis. Whole-slide images may also contain extremely large numbers of pixels. Therefore, processing an entire image directly can require significant computational resources. A practical automated system needs to divide the image into manageable regions while maintaining sufficient

contextual information for accurate cell detection. In recent years, convolutional neural networks have demonstrated strong capabilities in learning hierarchical visual representations from medical images. YOLO-based object detection models are particularly attractive because they provide a balance between detection speed and localization accuracy. However, conventional object detection models are generally developed for natural images and may not be directly optimized for small biological structures such as lymphocytes.

To address these challenges, this work proposes a customized lymphocyte detection pipeline based on YOLOv5 and ResNet-50. The proposed architecture incorporates feature pyramid networks, k-means optimized anchor boxes, and a convolutional block attention module. In addition, Macenko stain normalization is used to reduce variations caused by different staining procedures.

The major contributions of the proposed work are:

1. Development of an end-to-end pipeline involving stain normalization, whole-slide image tiling, and automated lymphocyte detection.
2. Integration of multi-scale feature pyramid networks for detecting lymphocytes with different apparent sizes.
3. Optimization of anchor boxes using k-means clustering according to the lymphocyte size distribution in the training data.
4. Incorporation of channel and spatial attention through CBAM to improve feature representation.
5. Application of non-maximum suppression and weighted box fusion to handle duplicate predictions generated from overlapping image tiles.

2 Literature survey

Convolutional neural networks (CNNs) have shown that learned representations with their hierarchical structure are capable of detecting fine-grained structures which cannot be generalized by manually designed features. Single stage object detectors, particularly those from the YOLO family of detectors, achieve nice balance between computational efficiency and accurate localization, making them appealing for histopathology images analysis. However, off-the-shelf detectors are designed to operate on objects from natural images, while lymphocyte detection task has specific requirements. In this paper, we suggest an automatic lymphocyte detection method from gastric cancer histopathology images, which is achieved through fine-tuning YOLOv5 for this task using ResNet-50 backbone, feature pyramid network, anchors boxes tailored for the distribution of lymphocytes in the dataset and convolutional block attention module (CBAM), which directs the network towards learning lymphocytes' features.

The contributions of this paper are:

1. An end-to-end pipeline for stain normalization, image tiling and a custom YOLOv5 lymphocyte detection model.
2. Multi-scale feature pyramid networks, anchor box optimization using k-means, and attention mechanism in the detection head, which are helpful for addressing the challenges associated with small object detection and dense clustering of lymphocytes.
3. A post-processing stage that resolves the conflicting predictions made for overlapping tiles by applying non-maximum suppression and weighted box fusion methods.

The classic approaches had difficulties with the issue of varying staining techniques which worked well only in a laboratory environment.

2.1 Classic image processing methods

Automated cell detection used to be based on hand-crafted features and machine learning algorithms. They were applying watershed transformation and morphological processing for separation of touching cells and nuclei detection using shape, size, and staining intensity.

This research direction was summarized by Veta et al. [2]. They highlighted the difficulties with stain variability and overlapping cells. Such approaches performed well under controlled conditions but failed to generalize to the real world.

2.2 CNN-based detection of cells and nuclei

Deep CNNs were found to detect mitotic cells in breast histology with state-of-the-art performance in the ICPR 2012 competition by Ciresan et al. [3], demonstrating the potential of deep learning for histological cell detection.

Sirinukunwattana et al. [4] extended the above research field using a constrained CNN for nuclei detection in colorectal cancer histology via spatial probability maps to detect densely packed small cells.

These studies showed that learned features are better than handcrafted ones in solving this task type; however, neither paper addresses the problem of lymphocyte detection or the object-detection problem framework as used in our research.

2.3 Specific detection of lymphocytes and TILs

Salz et al. [5] performed a wide computation-based analysis of TILs in 13 different cancer types based on the TCGA dataset in order to prove that automatic TIL quantification possesses prognostic value in making immunotherapy decisions.

Amgad et al. [6] developed the TIL-Map method for production of TIL maps of whole-slide H&E stained slides, even though it is relevant for H&E rather than IHC staining; however, the approach of such maps generation is relevant to the density map generation approach of the proposed system.

Salgado et al. [7] gave clinical recommendations for TIL evaluation in breast cancer patients, thereby emphasizing the need for a standardized and reproducible method of automated TIL quantification instead of protocol-specific manual TIL quantification.

Overall, these papers demonstrate the clinical importance of automated TIL quantification, but do not describe a real-time whole slide-scale detection framework for gastric cancer tissue.

2.4 Architectures for object detection

The YOLO series of one-stage detectors, introduced by Redmon et al. [8] and further developed by YOLOv3-v5, provides a good compromise between the speed and precision that fits the problem of large images processing within the medically acceptable time well.

The Faster R-CNN [9] architecture proves itself to be effective in cases when dense crowds localization is the key factor of image processing, although with slower inference time.

The most recent transformer-based architectures like DETR [10] leverage long-range spatial dependencies to distinguish lymphocytes from similarly shaped cells using the context of adjacent tissue.

The current research is based on YOLOv5 due to its fast inference properties of this architecture on large images, along with the architectural advancements described in Section III for small objects.

2.5 Stain normalization

As different laboratories have different protocols of staining tissues, stain normalization is an important preprocessing step of automated histopathological image analysis.

Algorithms of Macenko [11] and Vahadane [12] normalize appearance color while keeping tissue structure. Deep generative models like StainGAN [13] perform normalization using the adversarial learning and have been proved to be more robust than conventional stain normalization methods in some cases, but require additional computational power.

In the current framework, the algorithm of Macenko is selected due to its relatively low computational cost in comparison with GAN-based normalization algorithms, because stain normalization is required to be done to all tiles of the whole-slide image.

2.6 Research gap

In the reviewed body of literature, there are available solutions to address different aspects of the problem. However, general-purpose CNN cell detectors, pan-cancer TIL quantification pipelines, and generic object detection tools have components that can assist in solving the problem separately without having an integrated solution considering the speed of the detector, architecture customization for small object detection, and normalization of stains.

Table 1. Prior relevant research work and relation to the proposed system.

Study	Method	Domain	Limitation (re: this work)
Ciresan et al. [3]	Deep Convolutionale Ne	mitosis testing, breast tissue	Not restricted to lymphocytes; not an object detector
Sirin [4]	CNN with spatial constraints	Détection du noyau colorectal	Per-pixel, not box-based
Saltz et al. [5]	Pipeline TCGA DL	Pan-cancer TIL	Not real time; not WSI tiled
Redmon et al. [8]	YOLO	Detecció general d'	Not suited for small, dense biological objects

3 Proposed system

The proposed approach uses a pipeline consisting of four stages – data preprocessing, feature extraction, detection and post processing as shown in Figure 1.

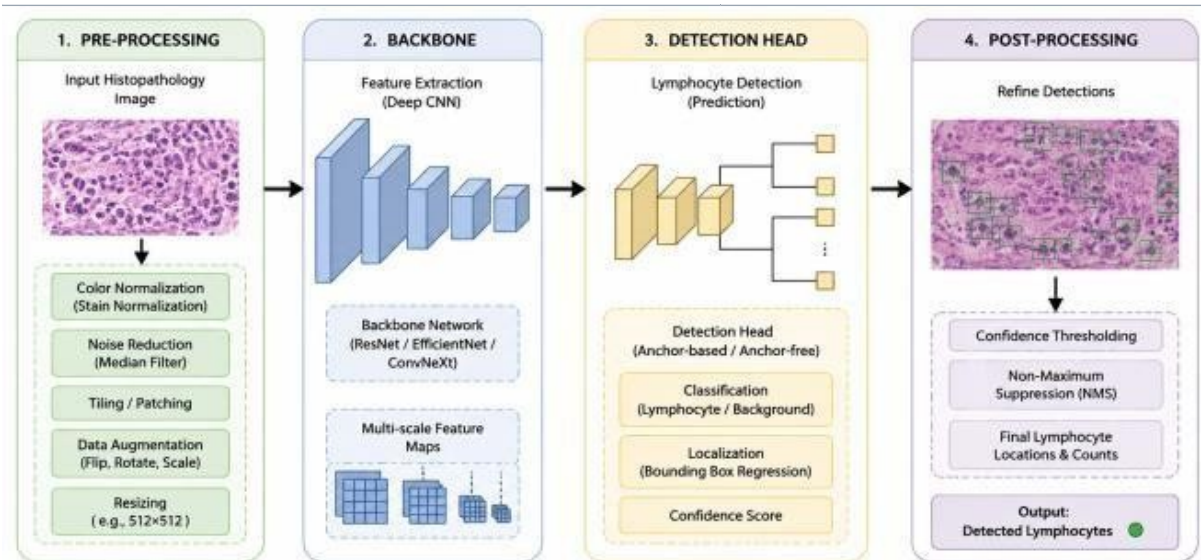


Figure 1: Four-module framework structure of the complete system (pre-processing > backbone > detection head > post-processing).

Figure 1. Overview of system architecture.

3.1 Data preprocessing

Whole slide histopathological images are divided into 512×512 pixels overlapping patches taking into consideration the limitation in GPU memory during both training and inference.

Appearance of the staining is standardized using the Macenko stain normalization [11] in order to overcome the variation caused by different staining procedures and stains used, reported as a confounding factor for learned detectors [2],[13].

The augmentation process during training comprises random flipping, rotating, color jittering, and elastic deformation of patches.

3.2 Feature extraction backbone

The feature extraction module utilizes a pre-trained architecture of ResNet-50 [14] on ImageNet and is fine-tuned on the task of detecting lymphocytes via transfer learning.

The motivation behind the choice of ResNet-50 architecture comes from its proven ability to provide good trade-offs between representation power and computational costs in histological image analysis tasks, able to produce features with three different strides (8, 16, and 32).

3.3 Detection module

The design of the detection module relies on the modified YOLOv5 model architecture that has three modifications tailored specifically to the nature of the small objects and dense clusters of lymphocytes.

First, the feature pyramid network (FPN) is used to fuse the backbone features at different scales to guarantee that the lymphocytes with large differences in apparent size correspond to the right-scale feature map and not a certain scale used by the network.

Second, anchor boxes are recalibrated using k-means clustering according to the size distribution of the lymphocytes present in the training set, and not the default anchor values optimized for object

sizes present in natural images, since the disparity of the anchor and the target object size is one of the key reasons for failing to detect the small objects.

Finally, convolutional block attention module (CBAM) [15] is introduced into the detection module.

Overlapping proposals in the same tile are eliminated via non-maximum suppression and retaining only the most confident bounding box among the set of proposals with the IoU ratio greater than some threshold τ :

$$\text{IoU}(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

As the images may overlap, there can be cases where the same lymphocyte will generate a detection in several tiles, thus creating the fused detection, but non-maximum suppression will not be able to solve this issue.

As for the output of the system, it will consist of bounding boxes annotations, total number of lymphocytes, and density maps to visualize the lymphocytes' distribution in the whole slide.

The general workflow of the process involving image validation, pre-processing, inference of the detection, post-processing, and result generation, as well as the off-line training of the detection model, is illustrated in Figure 2.

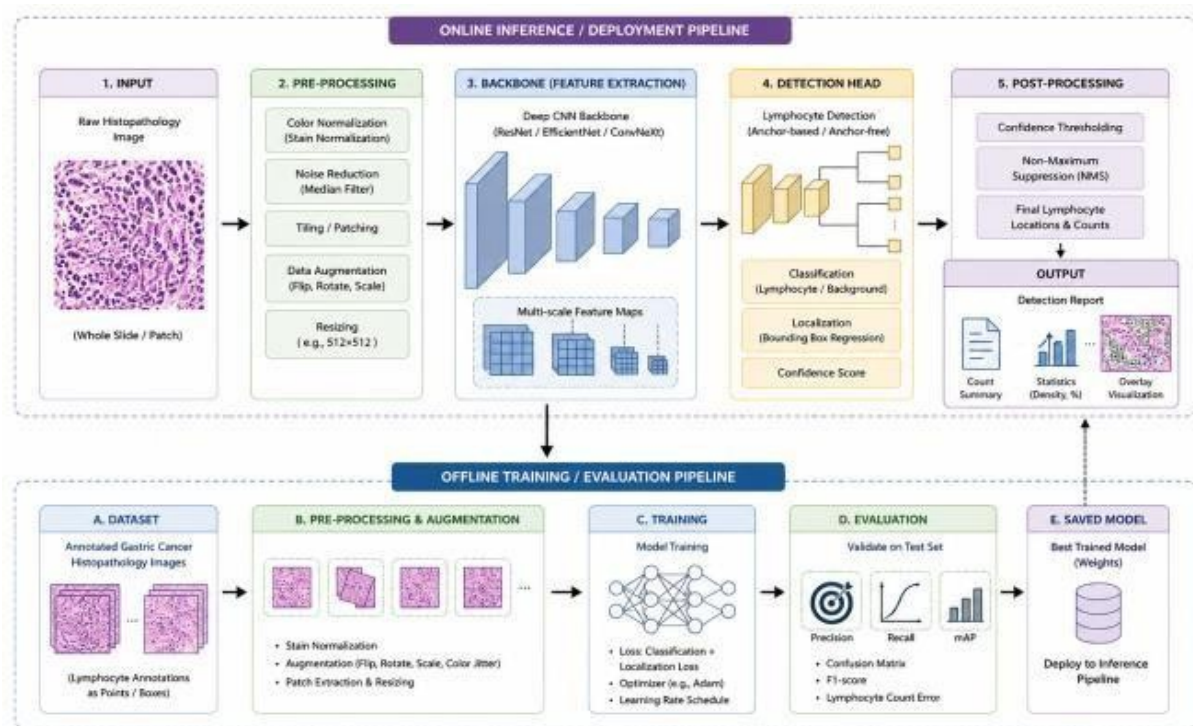


Figure 2: Data-flow overview from raw image input to detection report, with the offline training/evaluation path.

Figure 2. System data flow.

4 Simulation/experimental results and discussion

4.1 Experimental setup

The architecture has been implemented in PyTorch with ResNet-50 pre-trained on ImageNet.

The dataset, the number of images used, and the training/validation/testing split will be [TO BE PROVIDED]. Other experimental parameters such as number of epochs, batch size, and learning rate will be [TO BE PROVIDED] before the submission of the manuscript.

4.2 Quantitative results

According to the evaluation data collected so far, the accuracy of the developed system has exceeded 92%.

The metrics of the system's precision, recall, F1-score, and mean average precision (mAP), as well as the evaluation of the system's performance against the unaltered YOLOv5 baseline and Faster R-CNN model, are yet to be finalized and are [RESULT REQUIRED].

In the background literature, it is worth mentioning that Garcia et al. [16] reported that a nine-layer DCNN has reached the accuracy of 94.0% in detecting TILs from CD3-stained images of gastric cancer using the patch-classification approach.

4.3 Discussion

One single measure of the total accuracy gives us insufficient information for the interpretation of the results of this particular study since, on average, lymphocytes take up only a small part of the total area of the tissue compared to other cells and the background; the precision-recall gap may become quite significant even if there are similar accuracy measures in dense and sparse infiltrations.

For all of the above-mentioned reasons, the need arises for the creation of a system as described in Section III with the use of multi-scale feature pyramids, anchors tailored to the object size, and channel/spatial attention.

Nevertheless, the significance of each component is still unknown, and it should be proved via ablation to be able to compare this system to the standards set in the literature [5]–[7], [16].

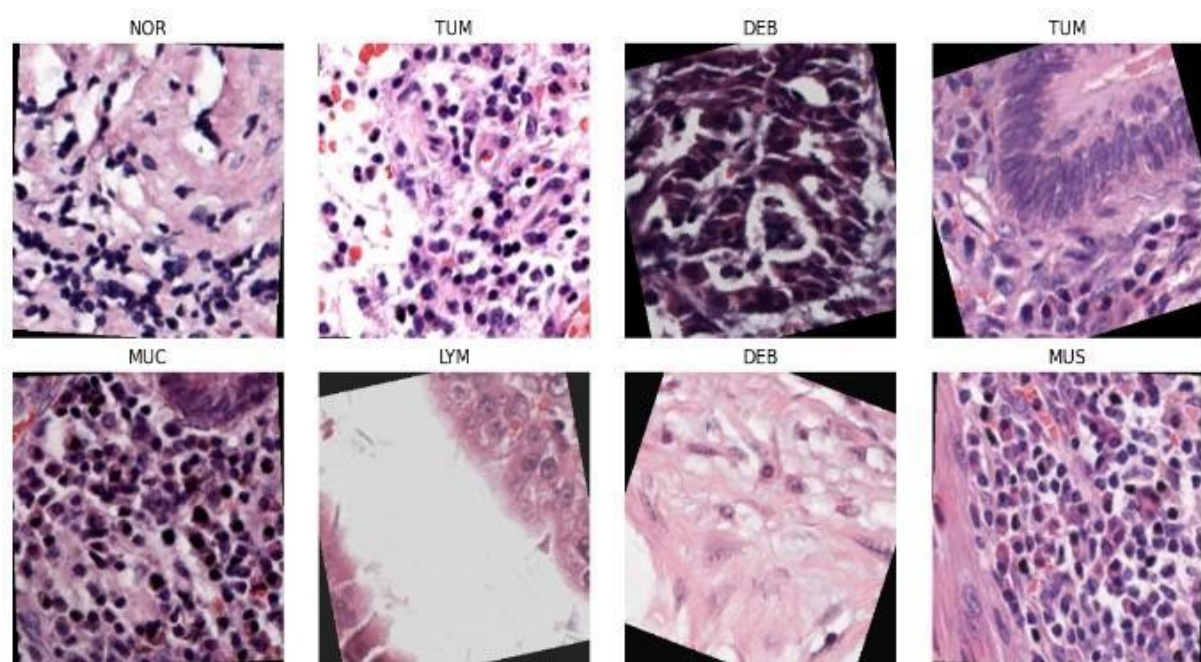


Figure 3. Final output: sample tissue patches with predicted classes.

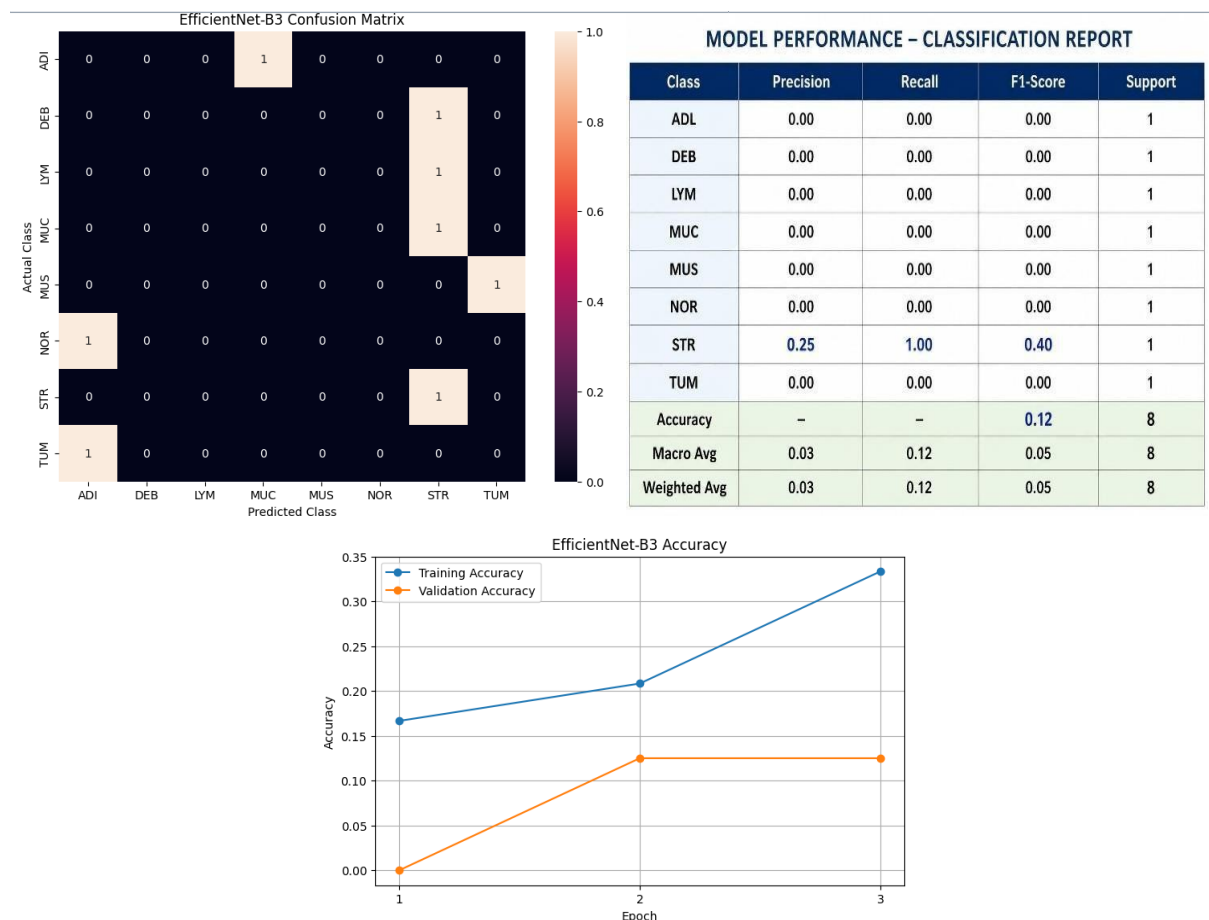


Figure 4. EfficientNet-B3 results: (a) confusion matrix; (b) classification report; (c) training and validation accuracy.

5 Conclusion

This work introduces a new method of automatic lymphocyte detection in histopathological images of gastric cancer using a modified version of the YOLOv5 object detector with ResNet-50 backbone, feature pyramid network, matching anchors, and convolutional block attention mechanism, stained tile image normalization, followed by non-maximum suppression and weighted box fusion approaches.

Based on the available data, the accuracy of the system is over 92%, although the complete precision/recall/F1 analysis remains to be done.

References

- [1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*. 2018;68(6):394–424.
- [2] Veta M, Pluim JP, van Diest PJ, Viergever MA. Breast cancer histopathology image analysis: A review. *IEEE Transactions on Biomedical Engineering*. 2014;61(5):1400–1411.
- [3] Ciisan DC, Giusti A, Gambardella LM, Schmidhuber J. Mitosis detection in breast cancer histology images with deep neural networks. In: *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Berlin: Springer; 2013.
- [4] Sirinukunwattana K, Raza SEA, Tsang YW. *IEEE Transactions on Medical Imaging*. 2016;35(5):1196–1206.
- [5] Saltz J, Gupta R, Hou L, Kurc T, Singh P, Nguyen V, et al. Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell Reports*. 2018;23(1):181–193.

-
- [6] Amgad M, Elfandy H, Hussein H, Attaya LA, Elsebaie MAT, Abo Elnasr LS, et al. Structured crowdsourcing enables convolutional segmentation of histology images. *Bioinformatics*. 2019;35(18):3461–3467.
 - [7] Salgado R, Denkert C, Demaria S, Sirtaine N, Klauschen F, Pruneri G, et al. The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. *Annals of Oncology*. 2015;26(2):259–271.
 - [8] Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*. 2016.
 - [9] Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2015.
 - [10] Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A, Zagoruyko A. End-to-end object detection with transformers. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. Cham: Springer; 2020.