A Robust Mean Teacher Framework for Semi-Supervised Cell Detection in Histopathology Images

Ziqi Wen, Chuyang Ye

School of Integrated Circuits and Electronics, Beijing Institute of Technology, Beijing, China

Introduction

- Cell detection in histopathology images facilitates clinical diagnosis, and deep learning methods have been applied to the detection problem with substantially improved performance.
- Cell detection methods based on deep learning usually require a large number of annotated training samples, which are costly and time-consuming to obtain, and it is desirable to develop methods where detection networks can be adequately trained with only a few annotated training samples.
- Since unlabeled data is much less expensive to obtain, it is possible to address this problem with semi-supervised learning, where abundant unlabeled data is combined with the limited annotated training samples for network training.
- We propose a semi-supervised object detection method for cell detection in histopathology images, which is based on and improves the mean teacher framework.

Methods

- We create a teacher model and a student model with the same network structure but different network weights, denoted as θ and θ' respectively for convenience, in accordance with the *mean teacher* (MT) framework.
  - We denote the image patch corresponding to a cell of interest detected by the teacher model at the t-th iteration in the unlabeled images by I_t^{u}, and the image patch corresponding to a randomly selected cell of the same class in the labeled images is represented by I^{a}. Then, for each image patch I_t^{u}, we mix it with the randomly selected I^{a} as

$$\tilde{I}_t^{\mathrm{u}} = I_t^{\mathrm{u}} \cdot p_t^{\mathrm{u}} + P(I^{\mathrm{a}}) \cdot (1 - p_t^{\mathrm{u}}), \tag{1}$$

    where p_t^{u} is the confidence of the teacher prediction for the image patch I_t^{u} and $P(\cdot)$ represents the resizing operation with bilinear interpolation to match the patch size of I^{a} to that of I_t^{u}.
  - The pseudo-label for the synthetic patch $\tilde{I}_t^{\mathrm{u}}$ is still the hard label of the teacher prediction, and each mixed patch $\tilde{I}_t^{\mathrm{u}}$ then replaces the corresponding original image patch I_t^{u} in the unlabeled images.
- Formally, suppose the classification result of the n-th detected cell in the synthetic unlabeled training images given by the student model is $\tilde{c}_n^{\mathrm{u}}$, the corresponding pseudo-label at the t-th iteration is $\tilde{d}_{n,t}^{\mathrm{u}}$, and the total number of the cells detected by the student model in the synthetic unlabeled images is $\tilde{N}_{\mathrm{u}}$.
  - The following unsupervised loss term is used at the t-th iteration, which can be combined with $\mathcal{L}_{\mathrm{sup}}$ to train the student model.

$$\mathcal{L}_{\mathrm{unsup}} = \frac{1}{\tilde{N}_{\mathrm{u}}} \sum_{n=1}^{\tilde{N}_{\mathrm{u}}} H(\tilde{c}_n^{\mathrm{u}}, \tilde{d}_{n,t}^{\mathrm{u}}). \tag{2}$$

  - An additional term $\mathcal{L}_{\mathrm{reg}}$ is computed for the student predictions on the synthetic unlabeled images as

$$\mathcal{L}_{\mathrm{reg}} = \frac{1}{\tilde{N}_{\mathrm{u}}} \sum_{n=1}^{\tilde{N}_{\mathrm{u}}} \|\tilde{c}_n^{\mathrm{u}}\|_p^p. \tag{3}$$

  - The complete loss function $\mathcal{L}$ for training the student model at the t-th iteration becomes

$$\mathcal{L} = \mathcal{L}_{\mathrm{sup}} + \lambda_{\mathrm{u}} \mathcal{L}_{\mathrm{unsup}} + \lambda_{\mathrm{r}} \mathcal{L}_{\mathrm{reg}}, \tag{4}$$

    where λ_{u} and λ_{r} are weights for the loss terms $\mathcal{L}_{\mathrm{unsup}}$ and $\mathcal{L}_{\mathrm{reg}}$, respectively.
- After the student model is trained at the t-th iteration, the network weights of the teacher model are also updated based on the student model and the current teacher model with EMA:

$$\theta' \leftarrow \theta' \cdot \sigma + \theta \cdot (1 - \sigma), \tag{5}$$

  where σ is the EMA decay rate to be specified. The iterative update of the teacher and student models is performed until convergence, and the teacher model is used for the final detection.
- A graphical illustration of the proposed framework is shown in Fig. 1.

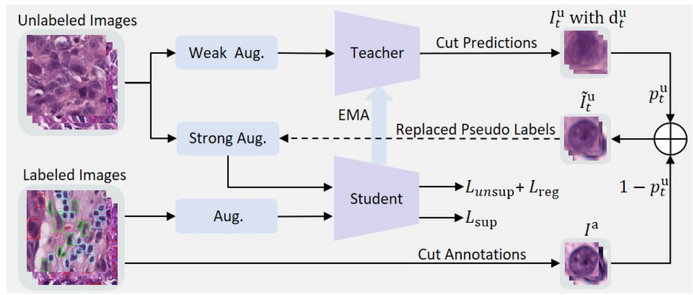

Figure 1: An overview of the proposed framework for semi-supervised cell detection.

Results

- To evaluate the proposed method, experiments were performed on the corrected single-rater subset of the publicly available NuCLS dataset that aims to detect multiple types of cells in breast cancer, which consisted of 1744 images with annotated cells.

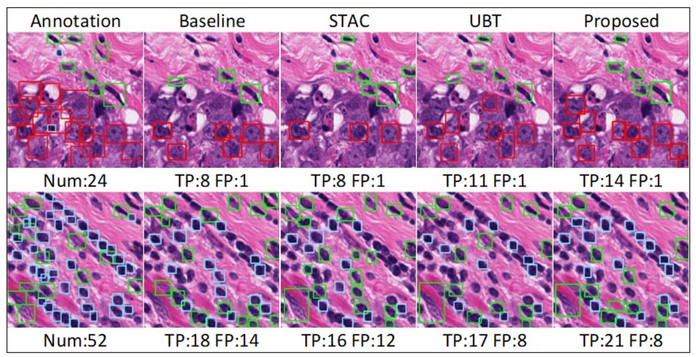

Figure 2: Examples of detection results on test images (achieved with 2% labeled training images) shown together with the annotation. The tumor, stromal, lymphocyte classes are represented by red, green, and blue boxes, respectively. The numbers of *true positive* (TP) and *false positive* (FP) detection results are indicated in the figure for each case. The numbers of annotated cells are also shown for reference.

- We randomly divided the images into a training, validation, and test set in a 7:1:2 ratio and split the training set into a labeled training set and an unlabeled training set, with annotations available only for the labeled training set in several cases (2%, 5%, 10%, and 20% of the training set) while the remaining training images were used as the unlabeled training set.
- We compared our proposed method with three other methods that used the same Faster R-CNN [2] detection network: the baseline method, STAC [3], and the *unbiased teacher* (UBT) method [1].
- As Fig. 2 shows, our method compares favorably with the competing methods by producing more true positive boxes than the competing methods without increasing the number of false positive boxes.
- As Tabel 1 shows, the proposed method outperforms the competing methods in terms of F1-score in all cases. As Fig. 3 shows, the proposed method without regularization has a higher mAP than competing methods, but lower than the complete proposed method, suggesting the benefits of both mixing and sparse regularization in the proposed method.

Table 1: The F1-score (%) achieved with different amounts of labeled training data (2%, 5%, 10%, and 20%) for each cell type. The tumor, stromal, lymphocyte classes are represented by Tum, Str, and Lym, respectively. The best results are highlighted in bold.

Method	2%			5%			10%			20%		
	Tum	Str	Lym	Tum	Str	Lym	Tum	Str	Lym	Tum	Str	Lym
Baseline	53.0	27.9	43.1	59.8	27.7	59.0	60.6	37.1	56.7	64.0	40.9	60.2
STAC	48.4	18.4	21.1	55.0	20.9	39.8	61.0	36.3	50.8	52.7	33.8	56.9
UBT	55.1	29.0	38.6	59.5	25.9	56.6	60.0	35.6	53.3	62.9	40.8	49.5
Proposed	**56.7**	**31.0**	**53.1**	**60.5**	**35.9**	**61.6**	**62.9**	**41.6**	**61.8**	**65.5**	**43.1**	**62.3**

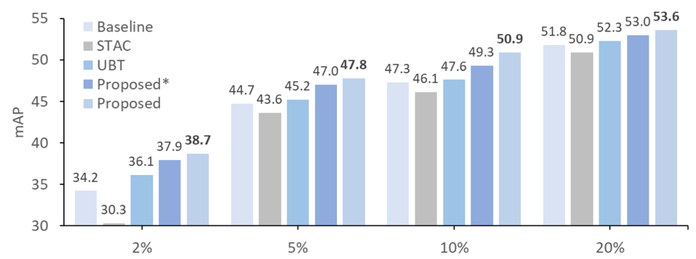

Figure 3: The mAP (%) computed for the detection results for each amount of the labeled training data (2%, 5%, 10%, and 20%). The best results are highlighted in bold. The 'Proposed*' denotes our proposed method without sparse regularization. Our proposed model can efficiently leverage the unlabeled data and perform favorably against the existing semi-supervised object detection works, including STAC and Unbiased Teacher.

Conclusion

- We have proposed a semi-supervised approach to cell detection in histopathology images.
- Based on the mean teacher framework, we have a developed a training procedure that is more robust to the noise in the teacher prediction.
- The experimental results on a publicly available dataset show that our method can improve the performance of semi-supervised cell detection.

References

[1] Yen-Cheng Liu, Chih-Yao Ma, Zijian He, Chia-Wen Kuo, Kan Chen, Peizhao Zhang, Bichen Wu, Zsolt Kira, and Peter Vajda. Unbiased teacher for semi-supervised object detection. In *International Conference on Learning Representations*, 2020.

[2] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1137–1149, 2017.

[3] Kihyuk Sohn, Zizhao Zhang, Chun-Liang Li, Han Zhang, Chen-Yu Lee, and Tomas Pfister. A simple semi-supervised learning framework for object detection. In *arXiv:2005.04757*, 2020.