Weakly Supervised Cell Instance Segmentation for Multi-Modality Microscopy

Ming XUE
`mingxue.mailbox@gmail.com`

Abstract

Instance segmentation of multi-modality high-resolution microscopy images is an important task in computational pathology. We extended HoVer-Net[1], originally developed for segmentation and classification of nuclei in multi-Tissue histology images, to apply it under weakly supervised situation. According to the final tests, this modification also works for multi-modality microscopy.

1 Introduction

Microscopy is a central technology of biomedical research, mentioned in nearly one million PubMed-indexed scientific papers to date. Various microscopy techniques allow capturing structural and functional properties of biological model systems, including cultured cells, tissues and organoids. As microscopy makes progress to capture such systems in greater detail and throughput and as the development of novel assays reveals more complex properties of living organisms, the need for robust and easy to use microscopy image analysis methods becomes critical to answer a wider variety of biological questions. Many image analysis workflows involve the identification (segmentation) of cell nuclei as a first step to extract meaningful biological signals. Research studies may involve counting cells, tracking moving populations, localizing proteins and classifying phenotypes or profiling treatments; in all of these and more, the nucleus is a reliable compartment of reference for identifying single cells in microscopy images.

Cell segmentation is usually the first step for downstream single-cell analysis in microscopy image-based biology and biomedical research. Deep learning has been widely used for image segmentation, but it is hard to collect a large number of labeled cell images to train models because manually annotating cells is extremely time-consuming and costly. Furthermore, datasets used are often limited to one modality and lacking in diversity, leading to poor generalization of trained models. This competition aims to benchmark cell segmentation methods that could be applied to various microscopy images across multiple imaging platforms and tissue types. The cell segmentation problem can be simplified as a weakly supervised learning task to encourage models that use limited labeled and many unlabeled images for cell segmentation as unlabeled images are relatively easy to obtain in practice[2, 3, 4, 5, 6, 7].

We modified the deep learning approach HoVer-Net, which was originally developed for simultaneous segmentation and classification of nuclear instances in sliding window mode for weakly supervised cell instance segmentation. The network is based on the prediction of horizontal and vertical distances of cell pixels to their centres of mass, which are subsequently leveraged to separate clustered cells.

Since the challenge is basiclly a instance segmentation task, we discard the classification branch of HoVer-Net and applied the model under weakly supervised situation.

The following part will basically describe the main idea and modifications made to the model and some detailed analysis. First, an overall introduction of the complete method will be introduced. Then, details about prepossessing, model structure, Post-processing and weakly supervised strategy will be presented.

2 Method

Our overall framework for automatic cell instance segmentation can be summarized as below: cell pixels are first detected and then, a tailored post-processing pipeline is used to segment cell instances. The framework is based upon the horizontal and vertical distance maps. Each cell pixel denotes either the horizontal or vertical distance of pixels to their centres of mass.

To get a better understanding of the original structure of HoVer-Net, here we make a brief introduction to the model: in order to extract a strong and representative set of features, a pre-activated residual network with 50 layers (Preact-ResNet50) is employed. Then nearest neighbour up-sampling is performed via two distinct branches to simultaneously obtain accurate nuclear instance segmentation: (i) nuclear pixel (NP) branch; (ii) HoVer branch, note here we discard the classification branch (nuclear classification (NC)) originally introduced in HoVer-Net, since classfication task is not necessary in this task. The NP branch predicts whether or not a pixel belongs to the nuclei or background, whereas the HoVer branch predicts the horizontal and vertical distances of nuclear pixels to their centres of mass. For more info , you can refer to the original model[1].

For the semi-supervised semantic segmentation[8, 9, 10, 11, 12], it aims to generalize from a combination set of pixel-wise labeled images and unlabeled images. This technology attracts more and more attention in recent years. We performs selective re-training via prioritizing reliable unlabeled images based on holistic prediction-level stability[13]. Concretely, several model checkpoints are saved in the first stage supervised training, and the discrepancy of their predictions on the unlabeled image serves as a measurement for reliability. The image-level selection offers holistic contextual information for learning.

2.1 Preprocessing

For the preprocessing part, we apply common geometric augmentations (RandAxisFlipd, RandRotate90d and RandZoomd) to each pair of input and label image on-the-fly during training. Additionally, we explore different types of pixel-wise augmentations using RandGaussianNoised, RandHistogramShiftd and so on.

The following code snippet shows the main prepossessing pipeline.

```
// Preprocessing
ScaleIntensityd(
    keys=["img"],
    allow_missing_keys=True
),
SpatialPadd(
    keys=["img", "label"],
    spatial_size=args.input_size
),
RandSpatialCropd(
    keys=["img", "label"],
    roi_size=args.input_size,
    random_size=False
),
RandAxisFlipd(
    keys=["img", "label"],
    prob=0.5
),
RandRotate90d(
    keys=["img", "label"],
    prob=0.5,
    spatial_axes=[0, 1]
),
RandGaussianNoised(
    keys=["img"],
    prob=0.25,
    mean=0,
    std=0.1
),
```

```python
RandAdjustContrastd(
    keys=["img"],
    prob=0.25,
    gamma=(1, 2)
),
RandGaussianSmoothd(
    keys=["img"],
    prob=0.25,
    sigma_x=(1, 2)
),
RandHistogramShiftd(
    keys=["img"],
    prob=0.25,
    num_control_points=3
),
RandZoomd(
    keys=["img", "label"],
    prob=0.15,
    min_zoom=0.8,
    max_zoom=1.5,
    mode=["area", "nearest"],
),
```
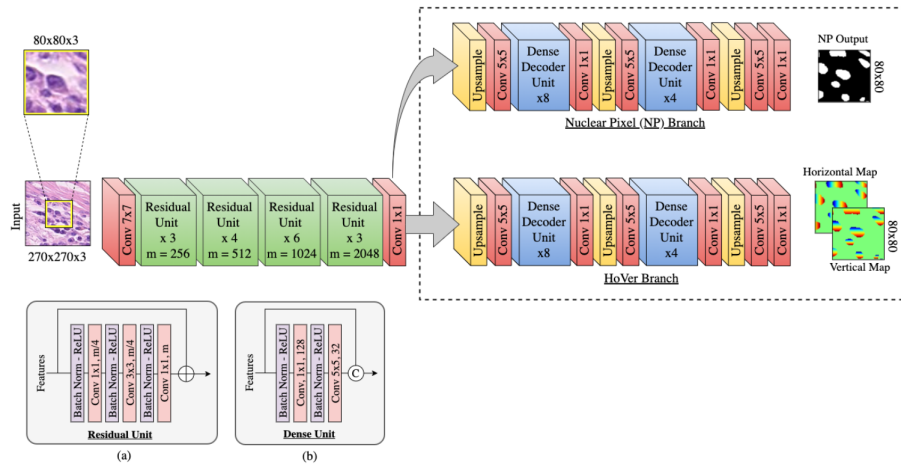
2.2 Model

Figure 1: modified model, adapted from HoVer-Net[1].

We have modified HoVer-Net for instance segmentation that not only detects nuclei with accuracy, but also effectively separates clustered nuclei. There are two up-sampling branches: 1) the nuclear pixel branch that separates nuclear pixels from the background; 2) the HoVer branch that regresses the horizontal and vertical distances of nuclear pixels to their centres of mass.

2.3 Post-processing

For the Post-processing, we mainly followed the original Hover-net. Within each horizontal and vertical map, pixels between separate instances have a significant difference. Therefore, calculating the gradient can inform where the nuclei should be separated because the output will give high values between neighbouring nuclei, where there is a significant difference in the pixel values. Specifically, compute the horizontal and vertical derivative approximations by the gradient maps.

The following code snippet shows the main post-processing pipeline.

```
// Post-processing
GenerateWatershedMaskd(
    keys=HoVerNetBranch.NP.value,
    softmax=True
),
GenerateInstanceBorderd(
    keys='mask',
    hover_map_key=HoVerNetBranch.HV,
    kernel_size=3
),
GenerateDistanceMapd(
    keys='mask',
    border_key='border',
    smooth_fn="gaussian"
),
GenerateWatershedMarkersd(
    keys='mask',
    border_key='border',
    threshold=0.7,
    radius=2,
    postprocess_fn=FillHoles()
),
Watershedd(
    keys='dist',
    mask_key='mask',
    markers_key='markers'
)
```

2.4 Weakly supervised training

For the weakly supervised training algorithm, we mainly referred to ST++[13], the main idea of this algorithm with selective re-training basically consists of the following steps:

1) Step 1: train on labeled images (Supervised training);
2) Step 2: select reliable images for the 1st stage re-training;
3) Step 3: pseudo label reliable images using best checkpoints ever;
4) Step 4: re-train on labeled and reliable unlabeled images;
5) Step 5: pseudo labeling unreliable images for the 2nd round;
6) Step 6: re-train on labeled and all unlabeled images for the 2nd round;

The whole workflow looks similar to normal semi-supervised learning, the main difference lies in step 2, in which phase reliable images should be selected. Here we calculate F1-score between the prediction from different checkpoints and the final checkpoint, which is usually the best or almost best ever.

3 Experiments

3.1 Dataset

In the NeurIPS 2022 Cell Segmentation challenge, the following datasets have been provided:

3.1.1 Training set

1000 labeled image patches from various microscopy types, tissue types, and staining types, and 1500+ unlabeled images.

There are four microscopy modalities in the training set, including:

1) Brightfield (300 patches);
2) Fluorescent (300 patches);
3) Phase-contrast (200 patches;

Figure 2: demo 1, from left to right: image, label, prediction

4) Differential interference contrast (200 patches);

3.1.2 Validation/Tuning set

101 images from various microscopy types with different size.

3.1.3 Testing set

more than 200 images (Hidden).

3.2 Implementation details

3.2.1 Environment settings

The development environments and requirements are presented in Table 1.

Table 1: Development environments and requirements.

System	Ubuntu 18.04.5 LTS
CPU	Intel(R) Xeon(R) Silver 4216 CPU @ 2.10GHz
RAM	64GB
GPU (number and type)	Four NVIDIA V100 32G
CUDA version	11.2
Programming language	Python 3.8
Deep learning framework	Pytorch

4 Results and discussion

4.1 Quantitative results on tuning set

We divided the training set (including labeled and unlabeled) into several parts, here are some demos showing our results from Figures 2 to Figures 6.

From the above demos, we can see that the model seems to perform better for Phase-contrast and Differential interference contrast than Brightfield microscopy images.

We conjecture that the model has still not learned the ability of distinguishing between cytoplasm and nucleus. We are still trying to find the root cause.

4.2 Qualitative results on validation set

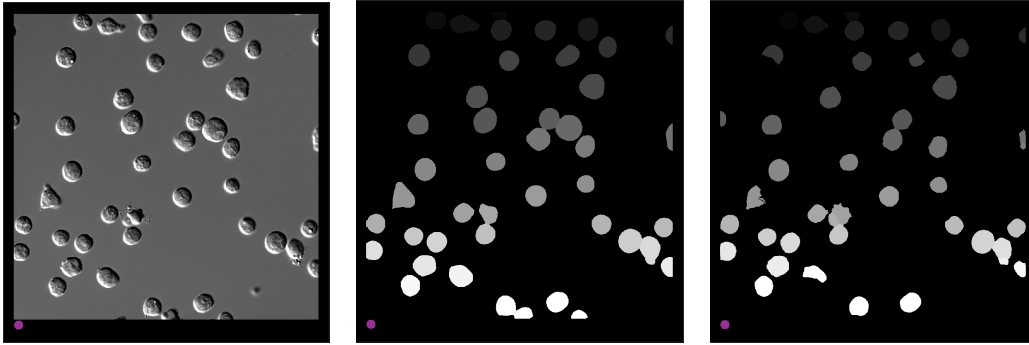For the leader board, we got F1-Score of 0.57, ranking 55/113.

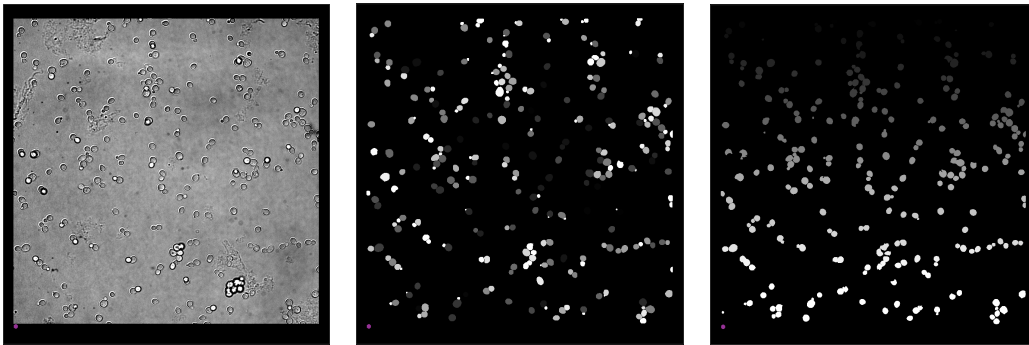Figure 3: demo 2, from left to right: image, label, prediction

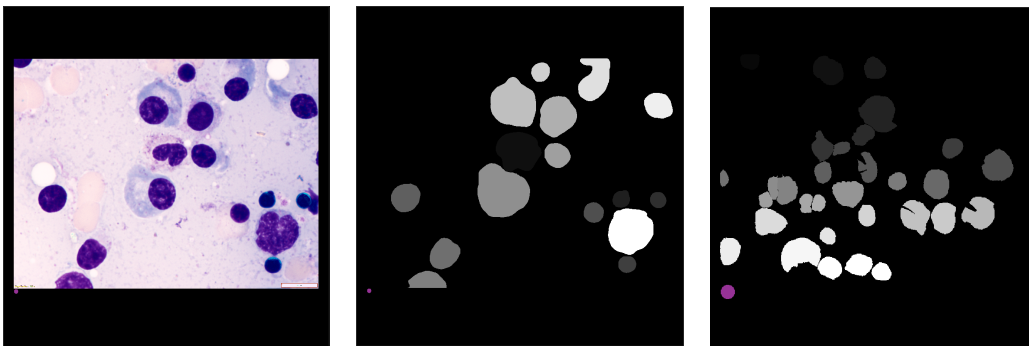Figure 4: demo 3, from left to right: image, label, prediction

Figure 5: demo 4, from left to right: image, label, prediction

Figure 6: demo 5, from left to right: image, label, prediction

4.3 Limitation and future work

The model is still not able to adapt to multi-modality microscopy with robustness, and more augmentations could be added to the traning process.

Besides, other weekly supervised learning skills could be tried and more ablation studies need to be conducted.

5 Conclusion

In this work, we construct a weakly supervised instance segmentation model via introducing strong data augmentations to unlabeled images, in hope of alleviating overfitting noisy labels as well as decoupling similar predictions. The whole pipeline is proposed to progressively leverage the unlabeled images.

Due to time limitation, we still have a lot to improve, e.g. more ablation study, why the model performs worse for certain microscopy modality.

Acknowledgement

References

[1] Simon Graham, Quoc Dang Vu, Shan E Ahmed Raza, Ayesha Azam, Yee Wah Tsang, Jin Tae Kwak, and Nasir Rajpoot. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical Image Analysis*, page 101563, 2019.

[2] Yves Grandvalet and Yoshua Bengio. Semi-supervised learning by entropy minimization. In L. Saul, Y. Weiss, and L. Bottou, editors, *Advances in Neural Information Processing Systems*, volume 17. MIT Press, 2004.

[3] Philip Bachman, Ouais Alsharif, and Doina Precup. Learning with pseudo-ensembles. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014.

[4] Mehdi Sajjadi, Mehran Javanmardi, and Tolga Tasdizen. Regularization with stochastic transformations and perturbations for deep semi-supervised learning. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.

[5] Samuli Laine and Timo Aila. Temporal ensembling for semi-supervised learning. In *International Conference on Learning Representations*, 2017.

[6] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *NeurIPS*, 2017.

[7] Takeru Miyato, Shin-ichi Maeda, Masanori Koyama, and Shin Ishii. Virtual adversarial training: a regularization method for supervised and semi-supervised learning. *TPAMI*, 2018.

[8] Qizhe Xie, Zihang Dai, Eduard Hovy, Minh-Thang Luong, and Quoc V Le. Unsupervised data augmentation for consistency training. In *NeurIPS*, 2020.

[9] Ilija Radosavovic, Piotr Dollár, Ross Girshick, Georgia Gkioxari, and Kaiming He. Data distillation: Towards omni-supervised learning. In *CVPR*, 2018.

[10] I Zeki Yalniz, Hervé Jégou, Kan Chen, Manohar Paluri, and Dhruv Mahajan. Billion-scale semi-supervised learning for image classification. *arXiv:1905.00546*, 2019.

[11] Qizhe Xie, Minh-Thang Luong, Eduard Hovy, and Quoc V Le. Self-training with noisy student improves imagenet classification. In *CVPR*, 2020.

[12] Barret Zoph, Golnaz Ghiasi, Tsung-Yi Lin, Yin Cui, Hanxiao Liu, Ekin D Cubuk, and Quoc V Le. Rethinking pre-training and self-training. In *NeurIPS*, 2020.

[13] Lihe Yang, Wei Zhuo, Lei Qi, Yinghuan Shi, and Yang Gao. St++: Make self-training work better for semi-supervised semantic segmentation. In *CVPR*, 2022.