Panoptic segmentation with highly imbalanced semantic labels

Josef Lorenz Rumberger*[1,2,6], Elias Baumann*[5], Peter Hirsch*[1,2],
Andrew Janowczyk[3,4], Inti Zlobec[5], Dagmar Kainmueller[1,2]
[1] *Max-Delbrueck-Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany*
[2] *Humboldt-Universität zu Berlin, Faculty of Mathematics and Natural Sciences, Berlin, Germany*
[3] *Case Western Reserve University, Department of Biomedical Engineering, USA*
[4] *Lausanne University Hospital, Precision Oncology Center, Switzerland*
[5] *University of Bern, Institute of Pathology, Bern, Switzerland*
[6] *Charité University Medicine, Berlin, Germany*

***Abstract*—This manuscript describes the panoptic segmentation method we devised for our submission to the CONIC challenge at ISBI 2022. Key features of our method are a weighted loss that we specifically engineered for semantic segmentation of highly imbalanced cell types, and an existing state-of-the art nuclei instance segmentation model, which we combine in a Hovernet-like architecture.**

***Index Terms*—nuclei segmentation, cell classification, digital pathology, challenge submission**

I. Introduction

The Colon Nuclei Identification and Counting (CoNIC) Challenge 2022 dataset of H&E stained whole-slide images comes with drastic semantic class imbalance: Among the six cell types described in [1], neutrophils and eosinophils each make up less than 1% of the cells in the dataset. Furthermore, nucleus size varies considerably across cell types, while at the same time some cell types, though frequent, also exhibit high intra-class variation in shape. To simultaneously handle the imbalances that stem from (1) cell types that rarely appear at all in an image, as well as (2) cell types that are small and thus occupy relatively few pixels in an image, we adopt ideas for importance sampling and loss weighting from the recent work of [3]. We complement these ideas with a state-of-the-art nuclei instance segmentation model adapted from [5].

II. Methodology

The CoNIC: Colon Nuclei Identification and Counting Challenge consists of two different tasks: *Task 1: Nuclear segmentation and classification* and *Task 2: Prediction of cellular composition*. We make some task specific adjustments, but keep the overall architecture the same.

A. Importance sampling

To treat image-level class imbalance, i.e., to cope with the fact that some classes rarely appear in an image at all, we adapt the training image sampling method presented in [3]: We sample training images such that the class distribution in the sampled training data is uniform. To this end, we determine the probability for a pixel in training image n to belong to class c as

$$X_{c,n} = \frac{1}{|h||w|} \sum_{h,w} m_{c,n,h,w} \tag{1}$$

where $m_{c,n,i,j} \in \{0, 1\}$ denotes the semantic mask of class c for training image n at pixel index (h,w). We then calculate the probability for a training image to be drawn as

$$p_n = \frac{1}{|c|} \sum_{c} \frac{X_{c,n}}{\sum_m X_{c,m}} \tag{2}$$

This has the effect that training images that do contain rare classes are preferably sampled. In each epoch, we sample with replacement a new subset from the training data using the sampling probabilities p_n.

B. Weighted focal loss for semantic segmentation

To cope with pixel-level class imbalance, i.e., the fact that some classes occupy fewer pixel per image than others, we adapt the focal loss from [3]. Based on the training labels we calculate an exponential moving average class prior

$$X_c^{t+1} = \gamma X_c^t + (1 - \gamma) X_{c,n}. \tag{3}$$

The loss is then calculated as weighted cross-entropy, with class weights $w_c = (1 - X_c^{t+1})^\rho$, with hyperparameter $\rho = 3$ and label smoothing [4] parameter set to 0.05.

C. Three-label model and auxiliary center-point vector regression for instance segmentation

We employ a three-label model (nucleus interior, nucleus boundary and background) for instance segmentation [10]. During training, we additionally regress nucleus center point vectors as auxiliary task, as described in [5]. These vectors are not used at test time. As training loss we employ the sum of the cross-entropy (classification) loss and the L2 (vector regression) loss, as in [5].

* equal contribution

D. Hovernet-like model architecture

The first 5 levels of an imagenet pre-trained EfficientNet B7 encoder [6] (or alternatively an EfficientNet-L2 encoder [13]) with two U-Net [7] decoders with 20% dropout per layer are used. One of the decoders is optimized for the semantic loss whereas the other is optimized for both instance segmentation losses.

E. Training recipe

We shuffle the dataset and split it into 4481 training tiles and 500 validation tiles. For color augmentation during training, we apply color deconvolution to the RGB image to obtain an HED instensity image [9]. The image is then augmented by randomly adjusting each channels contrast and transformed back to RGB. In addition, we apply RGB Color Jitter. Furthermore, gaussian blur and slight gaussian noise is applied. Spatial augmentations include random mirroring, translations, scaling, zooming, rotations, shearing and elastic transformations. We use AdamW optimizer [12] with $1e-4$ weight decay and cosine annealing learning rate scheduler. Models are trained for 600k training steps with batch size 4 and the model with the highest $mPQ+$ on the validation set is picked as the best checkpoint.

F. Test-time augmentations and inference

During inference we apply random HED color augmentation and change staining intensities by 10% and apply random flips and $90°$ rotations. Furthermore we use dropout during test-time [8]. We thus obtain $predictions$ from 16 forward passes, 8 per model in the ensemble for *Task 1*, and obtain our final prediction by pixel-wise averaging.

G. Post-processing

For each task, we draw per-class seeds from the predicted nucleus interior class and use watershed to delineate instances as described in [5]. Foreground and seed thresholds for watershed are optimized on our validation set. To split distant false merges in the predictions, we apply the connected-components algorithm on each individual instance mask. Then, we remove small as well as large instances again with class specific thresholds and filter out instance masks where the instance solidity is below a cell type specific threshold. We fill any holes in each instance mask. Parameters are optimized by hyperparamter search on the validation set. For each instance mask we assign the cell type that maximizes the sum of the cell type's softmax scores over all pixels.

H. Results

We evaluated two different models on the preliminary test set. Models differ only in their encoder architecture. The larger EfficientNet B7 considerably improved performance compared to the EfficientNet B5 model that we used for method development.

We submit two different versions per tasks for the final test set. For *Task 1*, we use an ensemble model consisting of one U-Net with EfficientNet-L2 backbone, and one U-Net with EfficientNet-B7 backbone and treat the results the same as TTAs from one model. For *Task 2*, we use only the U-Net with EfficientNet-B7 backbone and transform the output by center cropping the masks to 224×224 and count the number of instances per class.

Model	mPQ+	PQ	R2
EfficientNet B5	0.45684	0.62954	0.68697
EfficientNet B7	0.48471	0.64458	0.75432

TABLE I

PRELIMINARY TEST SET RESULTS FOR TWO MODEL CONFIGURATIONS

REFERENCES

[1] S. Graham, M. Jahanifar, A. Azam, M. Nimir, Y. W. Tsang, K. Dodd, E. Hero, H. Sahota, A. Tank, K. Benes, N. Wahab, F. Minhas, S.E.A. Raza, H. El Daly, K. Gopalakrishnan and N. Rajpoot, "Lizard: A Large-Scale Dataset for Colonic Nuclear Instance Segmentation and Classification". In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 684-693).

[2] S. Graham, M. Jahanifar, Q.D. Vu, G. Hadjigeorghiou, T. Leech, D. Snead, S.E.A. Raza, F. Minhas and N. Rajpoot, "CoNIC: Colon Nuclei Identification and Counting Challenge 2022". arXiv preprint arXiv:2111.14485 (2021).

[3] N. Araslanov and S. Roth, "Self-supervised Augmentation Consistency for Adapting Semantic Segmentation". In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 15384-15394)

[4] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the inception architecture for computer vision". In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 2818-2826).

[5] P. Hirsch and D. Kainmueller, "An auxiliary task for learning nuclei segmentation in 3d microscopy images". In Medical Imaging with Deep Learning (pp. 304-321).

[6] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks". In International conference on machine learning (pp. 6105-6114).

[7] O. Ronneberger, P. Fischer and T. Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention (pp. 234-241).

[8] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning". In international conference on machine learning (pp. 1050-1059).

[9] A.C. Ruifrok and D.A. Johnston "Quantification of histochemical staining by color deconvolution". Analytical and quantitative cytology and histology, 23 4, (pp. 291-299).

[10] J.C. Caicedo, J. Roth, A. Goodman, T. Becker, K.W. Karhohs, M. Broisin, M. Csaba, C. McQuin, S. Singh, F. Theis and A.E. Carpenter, "Evaluation of deep learning strategies for nucleus segmentation in fluorescence images". Cytometry Part A, 95(9), 952-965.

[11] S. Graham, Q.D. Vu, S.E.A Raza, A. Azam, Y.W. Tsang, J.T. Kwak and N. Rajpoot. "Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images". Medical Image Analysis, 58, 101563.

[12] I. Loshchilov and F. Hutter, "Decoupled Weight Decay Regularization". International Conference on Learning Representations

[13] X. Xie, M.T Luong, E. Hovy and Q.V. Le, "Self-training with noisy student improves imagenet classification". In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10687-10698).