Weakly-Supervised Cell Segmentation for Multiplex Immunohistochemistry Images

Anonymous ECCV submission

Anonymous Organization

Abstract. Multiplex immunohistochemistry (mIHC) is a novel scalable method of staining multiple cell types in a single tissue slice. In this paper, we propose a new method to automatically segment multiple cell types from a mIHC whole slide image. Our method only requires domain experts to provide a limited number of weak annotations (i.e., labeled dots placed at the centers of cells), while still achieving high quality segmentation. In particular, we (1) expand dot labels to mask annotations via superpixels; (2) introduce a multi-resolution supervision loss; and (3) leverage color deconvolution networks to further refine segmentation boundaries. Empirical evaluation on pancreatic cancer tissue slides demonstrates the efficacy of our method in providing an unprecedented amount of data from a single tissue section. Combining mIHC and the cell segmentation methods described herein would enable large scale studies of the immune contexture of cancer with minimal annotation effort from domain experts.

Keywords: Cell Segmentation · Multiplex Immunohistochemistry.

1 Introduction

Multiplex techniques allow the study of multiple cell types and the spatial relationships between them while maximizing the amount of information acquired from a single sample [13,4,8]. This is particularly important for the study of the tumor immune microenvironment, which has become an intense area of translational research focus. Multiplex immunohistochemistry (mIHC) and immunofluorescence (mIF) allow simultaneous labeling of 5 or more distinct cell types in the same tissue sample using colored chromogens or fluorophores, respectively. mIF tends to be more costly and requires a specialized microscope for image capture. On the other hand IHC is already routinely used in clinical medicine, and mIHC images can be captured by traditional bright field microscopy with a single, low-cost imaging step (10% of the cost of fluorescent staining), making mIHC the most rational choice for future large scale studies.

While scalable staining and image capture platforms are readily available for mIHC, automated image analysis platforms are lacking. Cell segmentation in mIHC images requires distinguishing between cell classes primarily based on color. Compared with other microscopic image modalities, there is high cell appearance heterogeneity in mIHC images. Cell shapes are highly variable; most

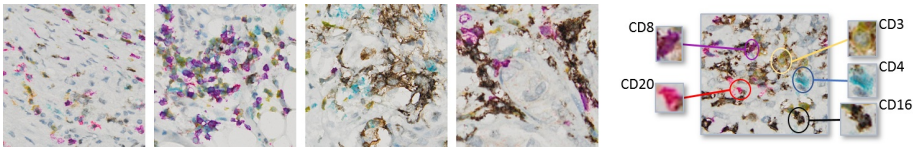

Fig. 1. Left: Four sample images from mIHC whole slide images taken at the same scale and size illustrating the fuzzy cell boundaries and large variation in immune cell size and shape. Right: Stained immune cell types are magnified and labeled.

lymphocytes are nearly circular with an average diameter of 8 micrometers (um), while macrophages range from discrete rounded cells of similar size, to elongated cells (∼20um) with projections extending in all directions. The edges of the cells may appear fuzzy due to chromogen properties, and boundaries between cells may be difficult to detect when cells are in close proximity. Furthermore, the appearance of the same cell types can be variable. Compared with H&E and fluorescence microscopy images, chromogenic staining has less clear delineation of the cell boundaries. Furthermore, nearby cells tend to be fused together and are hard to separate (see Fig. 1).

mIHC works by tagging an enzyme to a specific protein that is uniquely produced by a given cell type; the enzyme acts on the chromogen, producing a colored dye localized only at the cell type of interest. Individual cells of a given class may produce varying amounts of the specific proteins targeted by the mIHC stains, leading to differences in staining color intensities across a class. Furthermore, color spectra of different chromogen stains overlap significantly, making it difficult to distinguish between some classes. While all cell nuclei in the tissue are stained with hematoxylin (nuclear counterstain, blue), target proteins are often localized at the cytoplasm of the cell. In the case of lymphocytes, much of the cell volume is taken up by the nucleus, which introduces a significant co-staining issue. Therefore, a direct color deconvolution (unmixing different colors as cell masks) [24,17,21,6,7,23] is both challenging and insufficient.

Another challenge in mIHC image segmentation is the lack of high quality manual annotations for training. Manual generation of highly detailed, high quality training data is challenging even for expert pathologists due to interactions between nearby biomarkers, co-existence of multiple stains in a cell, biomarker protein expression variation, and the large variation of macrophage shapes. There is a need for segmentation methods that can use weak annotations, for example, dots placed at (approximately) the cell centers.

In this paper, we propose a novel approach for weakly-supervised semantic cell segmentation of mIHC images. Our goal is to achieve high segmentation quality despite limited supervision and cell appearance heterogeneity in mIHC. In our approach, we first extend dot annotations to mask annotations using superpixels. Second, we introduce multi-resolution supervision, i.e., comparing the prediction result at different layers/resolutions of the neural network. Finally, we observe that focusing on stains/colors can capture the fine-details of the

cell near the boundary. Thus, we use a color deconvolution method (detecting masks corresponding to different colors/stains) to complement our segmentation network. Empirically, our method attains high quality segmentation results on sample patches from 4 whole slide images of pancreatic cancer tissue and we believe opens the way to future large-scale studies.

Related work. Cell and nucleus segmentation have been studied mainly in the context of H&E and fluorescence microscopy [5,15,3,25,22]. Most of the earlier work relies on the availability of high quality, curated pixel-wise annotations that delineate the boundaries of cells and nuclei. Recent work has looked at methods for overcoming the need for high-quality annotation data in fluorescent microscopy datasets [18,11]. [19] is the first weakly supervised method that incorporates dot annotations for nuclei segmentation in H&E stained histopathology images. They assumed that most nuclei shapes are convex regions, so segmentation masks are generated with Voronoi diagrams. After the segmentation network is trained, they used a fine-tuning step with a dense conditional random field loss. In the context of mIHC segmentation, [4] is the first work to enable cell segmentation of more than 5 immunostaining cell classes in prostate cancer tissue. Also, [14] performed segmentation in the bone marrow tissue from patients with acute lymphoblastic leukemia. Both works rely on traditional segmentation methods such as the watershed segmentation method and Otsu's thresholding method, respectively. Recently, [12] proposed a multi-stage weakly supervised cell detector with only dot annotations using mIHC images of breast cancer tissue. Segmentation masks are generated by labeling pixels within a fixed radius from dot annotations as centers. The detection process is performed in two stages. First, a segmentation network learns to predict the cell locations and includes an auxiliary loss to regress the number of cells. Then, a cell classifier model based on a feature extractor is trained with small patches that cover the cell areas.

2 Methodology

We use weak annotation labels in the form of dots placed in the centers of the cells. These dot annotations are used to train and to validate our automatic segmentation method. We use superpixels to expand the labels of each dot to its adjacent area, so that we have a per-pixel annotation to train our network. We compute superpixels in each input image using SLIC [2] (one of the commonly used superpixel methods). Each superpixel is assigned the same semantic label (i.e., the stain/color) as the dot within it. This means all pixels within a superpixel are given the same semantic label. An empty superpixel is assigned the background label.

Fig. 2 illustrates the overview of our method. We first convert dot annotations into a superpixel mask. Using this per-pixel annotation, we train a semantic segmentation network. We adopt the UNet [20] architecture. To address the large cell shape and size variations, we introduce multi-scale supervision,

namely, provide loss-based supervision at different resolutions/layers of the decoder module. This way the network learns accurate representations at different resolutions. Details are provided in Section 2.1. Finally, since the superpixel mask itself is inaccurate in delineating cell boundaries, the segmentation model trained with the mask cannot be expected to produce correct details. To address this problem, we introduce a color decomposition network that is able to capture fine grain stain presence and composition. We then employ an ensemble method that combines the segmentation results of both networks. The final results enjoy the advantages of both networks and are of higher quality. See Section 2.2 for more details.

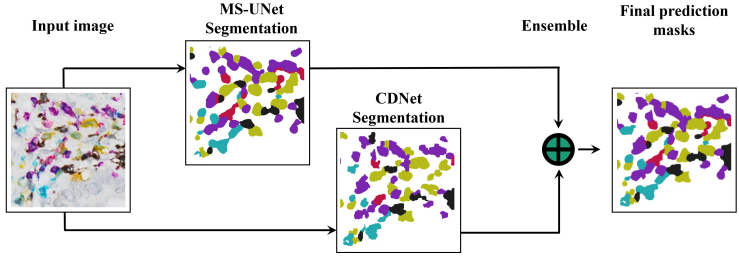

Fig. 2. Pipeline of the ensemble method, which combines the predictions of the multi-resolution UNet (MS-UNet) and the color decomposition model (CDNet).

2.1 Segmentation Network

In this section, we propose a semantic segmentation network based on the UNet [20] architecture. The segmentation network has to deal with the fact that cells exhibit a large variation in shape and size (see Fig. 1 for an illustration). To address this challenge, we draw inspiration from how pathologists study tissue images. Pathologists often use multiple magnifications jointly so that they can take into consideration cell/tissue architectural features at different scales. By viewing images (especially mIHC images) at different scales, pathologists capture contextual information about shape and size variations.

Based on this insight, we introduce a multi-resolution supervision component to the segmentation network. In particular, we introduce additional supervision to different intermediate layers of the decoder module of UNet. The additional supervision enables the network to learn more discriminative features in those intermediate layers so that their feature representation better captures cells of different sizes/shapes. In the literature, this technique, called a deeply supervised network [16], has been applied in various domains [9,26,27,10]. To the best of our knowledge, *we are the first to apply deep supervision to microscopic image analysis tasks.*

The architecture of our multi-resolution UNet (MS-UNet) is illustrated in Fig. 3. More specifically, we input the intermediate feature representation of each of the layers into a convolutional layer for semantic label prediction. The prediction is of the same resolution as the layer. We use a downsampled superpixel mask to supervise this prediction (via cross-entropy loss). In this manner, we can enforce the segmentation network to learn better representations at different resolutions, from fine to coarse. The total loss L can be formulated as follows

$$L = L_{CE}(Y, \widehat{Y}) + \sum_{i=1}^{n} \lambda_i L_{CE}(Y_i, \widehat{Y}_i) \tag{1}$$

where L_{CE} is the categorical cross-entropy loss. The first cross-entropy loss compares $\widehat{Y}$, the prediction of the network, and Y, the superpixel mask. The remaining terms compare an intermediate layer prediction $\widehat{Y}_i$ and a downsampled superpixel mask, Y_i (downsampled by a factor of 2^i). The weights of the different intermediate layer losses are controlled by λ_i's.

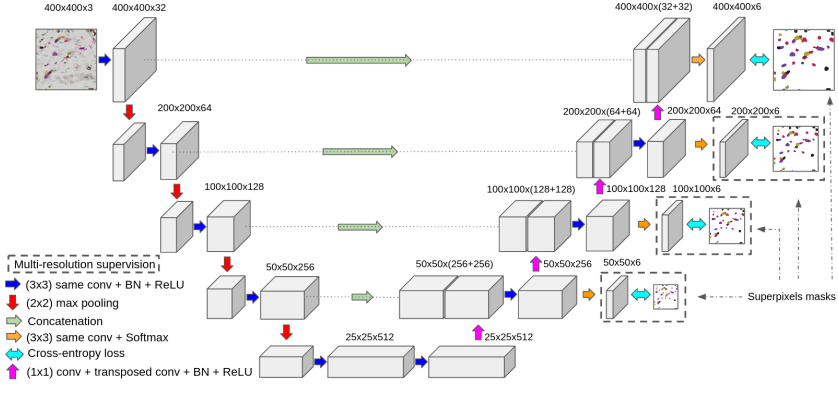

Fig. 3. Multi-resolution UNet (MS-UNet).

2.2 Ensemble with Color Decomposition Network

The segmentation (or segmentor) network is unable to capture fine cell details, especially near their boundaries. This is inherently unavoidable because the training labels (superpixel masks) are extended automatically from dot annotations and are not guaranteed to be accurate. Failing to detect fine scale cell boundaries can be detrimental to downstream analyses that require accurate measurements of pairwise distances between cells, cell sizes and cell distributions.

We observe that the cell boundary is well defined by color chromogenetic stains, which bind to proteins and express themselves in the cell cytoplasm. If

we can identify the color stains correctly, we can refine the prediction cell mask. Indeed, color deconvolution (or color decomposition), which is the process of finding masks corresponding to different stain colors, is a classic image analysis task. Various existing color deconvolution methods have been developed [24,17,21,6,7,23]. Most of these methods can only solve problems with up to 4 stains. In this paper, we adopt a recent deep autoencoder method that can unmix a large number of stains [1][1].

Fig. 4 shows an input mIHC image and the color deconvolution result. We show the concentration maps of different stains/colors. We can compute a segmentation mask for each stain by thresholding these concentration maps (Fig. 4 right). Our method uses this pretrained color deconvolution network, denoted as CDNet, to complement the segmentor MS-UNet. We observe that MS-UNet is better in detecting cells, while CDNet better delineates cell boundaries.

We propose an ensemble of the outputs of MS-UNet and CDNet and we denoted it as the *anchor* operation, where we exploit the benefits of both methods. In particular, *CDNet anchor MS-UNet* uses the MS-UNet predictions to determine cell locations. Then, we refine each located cell by taking the union of the MS-UNet prediction and the CDNet prediction. In other words, we drop any connected component of the CDNet prediction which does not intersect with the MS-UNet prediction. We take the union of the remaining CDNet prediction and the MS-UNet prediction as the final output mask. Note that this ensemble is carried out for each stain/color separately.

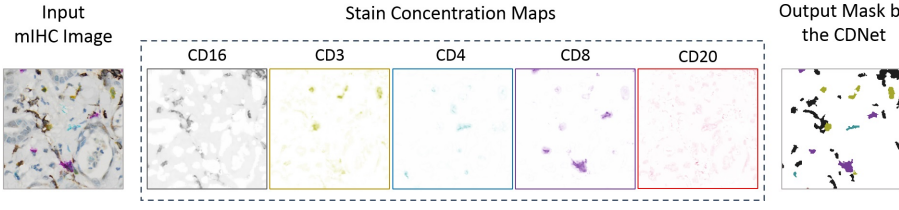

Fig. 4. Sample output from the color decomposition network (CDNet).

3 Experimental Results

Dataset and Settings. We evaluated the proposed method using multiplex IHC whole slide images (WSIs) of pancreatic cancer tissue that are stained with chromogenic biomarkers. The biomarkers are CD3 (yellow), CD4 (teal), CD8 (purple), CD16 (black), and CD20 (red), representing different types of immune cells. Fig. 1 depicts instances of these stained immune cell types. The training set consists of 300 patches of size 400×400 pixels randomly sampled from the tumor

[1] The code was obtained by communication with the authors

area of 5 WSIs and the validation set is 60 patches of the same dimensions from another WSI. The test set is 19 patches of size 1200×1920 pixels from 4 different WSIs. The physical resolution of all the patches is 0.174 microns per pixel. The MS-UNet and CDNet are each trained independently on the same dataset. The MS-UNet is trained with the Adam optimizer for 1200 epochs, with an initial learning rate of 0.001 and decreased by 10 every 400 epochs. Using the validation set, the model with the lowest cross entropy loss in only the last output layer is selected. In the final segmentation small connected components that are far less than the typical size of immune cells are removed. A size threshold of 100 pixels, which is around 3 square microns, is empirically chosen.

Quantitative Results. In our evaluation, F-scores are computed using the dot annotations as follows: (i) a true positive is a dot that intersects with the prediction segmentation mask of the same label, (ii) a false positive is a connected component in the prediction that does not intersect with a ground truth dot of the same label, and (iii) a false negative is a dot that does not intersect with any component of the same label in the prediction segmentation mask. We compare the results from the segmentation using CDNet alone, a typical UNet without deep supervision, and MS-UNet with different weights for the deep supervision layers. Performance is improved by varying the weight for the first deep supervision layer (λ_1). Varying the weights for the rest of the remaining deep supervision layers does not improve performance. In Table 1 we compare with different configurations for λ_1, further comparisons with varying λ_2 and λ_3 are in the supplementary material. Table 2 shows that the ensemble *CDNet anchor MS-UNet* method achieves higher performance than the segmentation network UNet and the color decomposition network CDNet. The mean F-score improves by 11.8% and 13% compared to CDNet and UNet respectively. Class-wise, there is a significant improvement in most classes compared to the individual model results in Table 1.

Qualitative results. Fig. 5 shows example qualitative results. For each patch, a magnified area is displayed in the following row. In the first magnified sample the red staining in the CD20 cell appears like purple staining. CDNet, which is sensitive to the staining, mistakes part of that cell for CD8, whereas MS-UNet correctly segments the whole cell. In the same magnified region, we see that the yellow staining on the right has a very light color that CDNet picks up more accurately and gives a better segmentation of the whole region than MS-UNet. In the last column we see the advantage of the proposed ensemble *CDNet anchor MS-UNet* which is able to capture the best of both models. In the second magnified sample, MS-UNet mistakenly connects the top cell to the bottom ones. This is due to being trained on coarse superpixels, on the other hand CDNet can better separate the cells. Because we take the union of the 2 results, we leave this issue to future work. Additionally, Fig. 6 shows segmentation predictions on two patches of the test set in full size. Further segmentation results of the ensemble method on the test set can be found in the supplementary material.

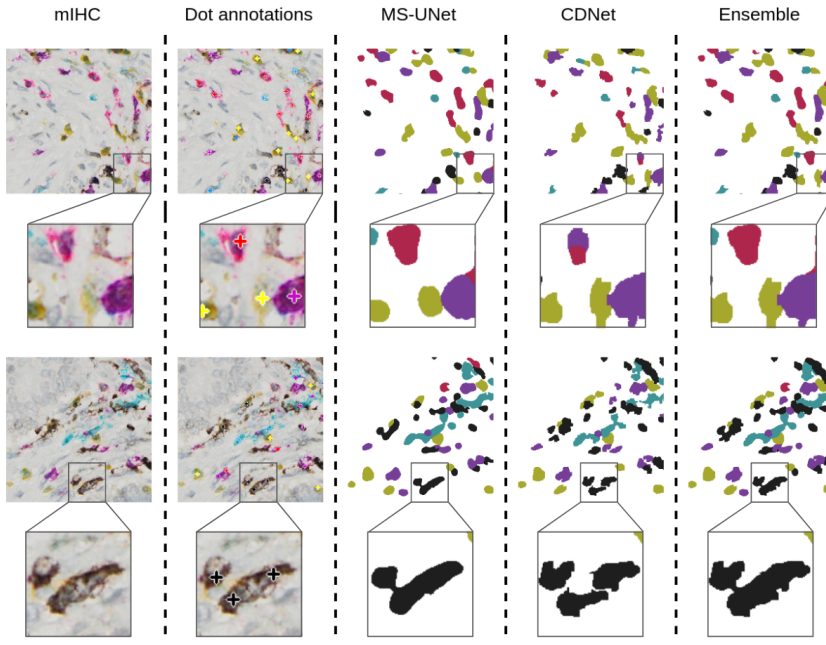

Fig. 5. Qualitative results on patches from the test set. Rows 1 and 3 are patches of 400x400 pixels. Rows 2 and 4 are magnifications of subregions of 100x100 pixels.

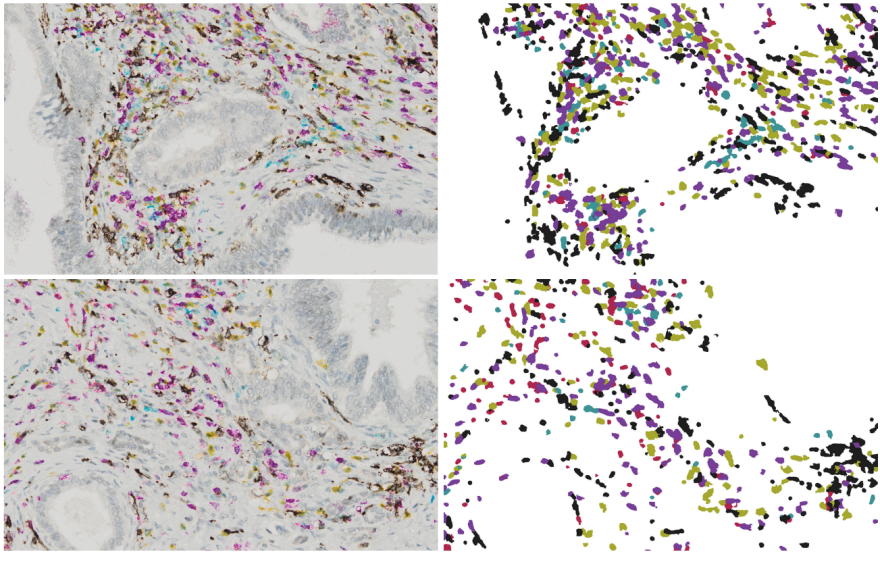

Fig. 6. Left: Two full size patches of 1200 × 1920 pixels from the test set. Right: Segmentation predictions using the ensemble method *CDNet anchor MS-UNet*.

Table 1. Evaluation of color decomposition network and individual segmentation networks

	F1-score					
Method	CD16	CD3	CD4	CD8	CD20	Mean
CDNet	0.6716	0.6168	0.6141	0.6336	0.2166	0.5505
UNet (λ_1=0)	0.7413	0.5876	0.5048	0.6166	0.2738	0.5448
MS-UNet(λ_1=0.50)	0.7383	0.6344	0.5907	0.6248	0.3342	0.5845
MS-UNet(λ_1=0.75)	**0.7473**	**0.6380**	0.5859	**0.6478**	0.3455	**0.5929**
MS-UNet (λ_1=1.00)	0.6881	0.6042	**0.6288**	0.6439	**0.3562**	0.5842

Table 2. Evaluation of ensemble models

	F1-score					
Method	CD16	CD3	CD4	CD8	CD20	Mean
MS-Unet anchor CDNet	0.6775	0.6246	**0.6211**	0.6429	0.2166	0.5565
CDNet anchor MS-UNet	**0.7877**	**0.6651**	0.6079	**0.6618**	**0.3556**	**0.6156**

4 Conclusions

We proposed a weakly supervised cell segmentation method for mIHC images. Our method leverages the fine details offered by the color decomposition network and the multi-scale representation learnt through deep supervision. It achieves high quality immune cell segmentation. We expect it to be very useful in downstream quantitative large-scale studies of tumor microenvironments.

References

1. Abousamra, S., Fassler, D., Hou, L., Zhang, Y., Gupta, R., Kurc, T., Escobar-Hoyos, L.F., Samaras, D., Knudson, B., Shroyer, K., et al.: Weakly-supervised deep stain decomposition for multiplex ihc images. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). pp. 481–485. IEEE (2020)
2. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: Slic superpixels compared to state-of-the-art superpixel methods. IEEE transactions on pattern analysis and machine intelligence **34**(11), 2274–2282 (2012)
3. Al-Kofahi, Y., Zaltsman, A., Graves, R., Marshall, W., Rusu, M.: A deep learning-based algorithm for 2-d cell segmentation in microscopy images. BMC bioinformatics **19**(1), 1–11 (2018)
4. Blom, S., Paavolainen, L., Bychkov, D., Turkki, R., Mäki-Teeri, P., Hemmes, A., Välimäki, K., Lundin, J., Kallioniemi, O., Pellinen, T.: Systems pathology by multiplexed immunohistochemistry and whole-slide digital image analysis. Scientific Reports **7**(1), 1–13 (2017)
5. Caicedo, J.C., Goodman, A., Karhohs, K.W., Cimini, B.A., Ackerman, J., Haghighi, M., Heng, C., Becker, T., Doan, M., McQuin, C., et al.: Nucleus segmentation across imaging experiments: the 2018 data science bowl. Nature methods **16**(12), 1247–1253 (2019)
6. Chen, T., Chefd'Hotel, C.: Deep learning based automatic immune cell detection for immunohistochemistry images. In: International workshop on machine learning in medical imaging. pp. 17–24. Springer (2014)
7. Chen, T., Srinivas, C.: Group sparsity model for stain unmixing in brightfield multiplex immunohistochemistry images. Computerized Medical Imaging and Graphics **46**, 30–39 (2015)
8. Dixon, A.R., Bathany, C., Tsuei, M., White, J., Barald, K.F., Takayama, S.: Recent developments in multiplexing techniques for immunohistochemistry. Expert review of molecular diagnostics **15**(9), 1171–1186 (2015)
9. Dou, Q., Yu, L., Chen, H., Jin, Y., Yang, X., Qin, J., Heng, P.A.: 3d deeply supervised network for automated segmentation of volumetric medical images. Medical image analysis **41**, 40–54 (2017)
10. Gao, Y., Liu, C., Zhao, L.: Multi-resolution path cnn with deep supervision for intervertebral disc localization and segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 309–317. Springer (2019)
11. Guerrero-Peña, F.A., Fernandez, P.D.M., Ren, T.I., Cunha, A.: A weakly supervised method for instance segmentation of biological cells. In: Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data, pp. 216–224. Springer (2019)
12. Hagos, Y.B., Narayanan, P.L., Akarca, A.U., Marafioti, T., Yuan, Y.: Concordenet: Cell count regularized convolutional neural network for cell detection in multiplex immunohistochemistry images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 667–675. Springer (2019)
13. Hofman, P., Badoual, C., Henderson, F., Berland, L., Hamila, M., Long-Mira, E., Lassalle, S., Roussel, H., Hofman, V., Tartour, E., et al.: Multiplexed immunohistochemistry for molecular and immune profiling in lung cancer—just about ready for prime-time? Cancers **11**(3), 283 (2019)
14. Hohtari, H., Brück, O., Blom, S., Turkki, R., Sinisalo, M., Kovanen, P.E., Kallioniemi, O., Pellinen, T., Porkka, K., Mustjoki, S.: Immune cell constitution in bone

marrow microenvironment predicts outcome in adult all. Leukemia **33**(7), 1570–1582 (2019)
15. Kumar, N., Verma, R., Anand, D., Zhou, Y., Onder, O.F., Tsougenis, E., Chen, H., Heng, P.A., Li, J., Hu, Z., et al.: A multi-organ nucleus segmentation challenge. IEEE transactions on medical imaging **39**(5), 1380–1391 (2019)
16. Lee, C.Y., Xie, S., Gallagher, P., Zhang, Z., Tu, Z.: Deeply-supervised nets. In: Artificial intelligence and statistics. pp. 562–570 (2015)
17. Macenko, M., Niethammer, M., Marron, J.S., Borland, D., Woosley, J.T., Guan, X., Schmitt, C., Thomas, N.E.: A method for normalizing histology slides for quantitative analysis. In: 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. pp. 1107–1110. IEEE (2009)
18. Nishimura, K., Bise, R., et al.: Weakly supervised cell instance segmentation by propagating from detection response. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 649–657. Springer (2019)
19. Qu, H., Wu, P., Huang, Q., Yi, J., Riedlinger, G.M., De, S., Metaxas, D.N.: Weakly supervised deep nuclei segmentation using points annotation in histopathology images. In: International Conference on Medical Imaging with Deep Learning. pp. 390–400 (2019)
20. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
21. Ruifrok, A.C., Johnston, D.A., et al.: Quantification of histochemical staining by color deconvolution. Analytical and quantitative cytology and histology **23**(4), 291–299 (2001)
22. Su, H., Xing, F., Kong, X., Xie, Y., Zhang, S., Yang, L.: Robust cell detection and segmentation in histopathological images using sparse reconstruction and stacked denoising autoencoders. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 383–390. Springer (2015)
23. Thommen, D.S., Koelzer, V.H., Herzig, P., Roller, A., Trefny, M., Dimeloe, S., Kiialainen, A., Hanhart, J., Schill, C., Hess, C., et al.: A transcriptionally and functionally distinct pd-1+ cd8+ t cell pool with predictive potential in non-small-cell lung cancer treated with pd-1 blockade. Nature medicine **24**(7), 994–1004 (2018)
24. Vahadane, A., Peng, T., Sethi, A., Albarqouni, S., Wang, L., Baust, M., Steiger, K., Schlitter, A.M., Esposito, I., Navab, N.: Structure-preserving color normalization and sparse stain separation for histological images. IEEE transactions on medical imaging **35**(8), 1962–1971 (2016)
25. Yi, J., Wu, P., Huang, Q., Qu, H., Liu, B., Hoeppner, D.J., Metaxas, D.N.: Multi-scale cell instance segmentation with keypoint graph based bounding boxes. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 369–377. Springer (2019)
26. Zeng, G., Zheng, G.: Multi-stream 3d fcn with multi-scale deep supervision for multi-modality isointense infant brain mr image segmentation. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). pp. 136–140. IEEE (2018)
27. Zhu, Q., Du, B., Turkbey, B., Choyke, P.L., Yan, P.: Deeply-supervised cnn for prostate segmentation. In: 2017 International Joint Conference on Neural Networks (Ijcnn). pp. 178–184. IEEE (2017)