Symmetric Dense Inception Network for Simultaneous Cell Detection and Classification in Multiplex Immunohistochemistry Images

Hanyun Zhang HANYUN.ZHANG@ICR.AC.UK
The Institute of Cancer Research, London, UK

Tami Grunewald* TAMI.GRUNEWALD.18@UCL.AC.UK
Ayse U. Akarca* A.AKARCA@UCL.AC.UK
University College London Hospital, London, UK.

Jonathan A. Ledermann J.LEDERMANN@UCL.AC.UK
UCL Cancer Institute, University College London, London, UK

Teresa Marafioti† T.MARAFIOTI@UCL.AC.UK
University College London Hospital, London, UK

Yinyin Yuan† YINYIN.YUAN@ICR.AC.UK
The Institute of Cancer Research, London, UK

Editors: M. Atzori, N. Burlutskiy, F. Ciompi, Z. Li, F. Minhas, H. Müller, T. Peng, N. Rajpoot, B. Torben-Nielsen, J. van der Laak, M. Veta, Y. Yuan, and I. Zlobec.

Abstract

Deep-learning based automatic analysis of the multiplex immunohistochemistry (mIHC) enables distinct cell populations to be localized on a large scale, providing insights into disease biology and therapeutic targets. However, standard deep-learning pipelines performed cell detection and classification as two-stage tasks, which is computationally inefficient and faces challenges to incorporate neighbouring tissue context for determining the cell identity. To overcome these limitations and to obtain a more accurate mapping of cell phenotypes, we presented a symmetric dense inception neural network for detecting and classifying cells in mIHC slides simultaneously. The model was applied with a novel stop-gradient strategy and a loss function accounted for class imbalance. When evaluated on an ovarian cancer dataset containing 6 cell types, the model achieved an F1 score of 0.835 in cell detection, and a weighted F1-score of 0.867 in cell classification, which outperformed separate models trained on individual tasks by 1.9% and 3.8% respectively. Taken together, the proposed method boosts the learning efficiency and prediction accuracy of cell detection and classification by simultaneously learning from both tasks.

Keywords: Deep learning, Digital pathology, Multiplex immunohistochemistry

* These authors contributed equally.
† These authors contributed equally.

1. Introduction

Multiplex immunohistochemistry (mIHC) is an important technique to resolve the spatial arrangement of multiple cell phenotypes within the tissue, which has been used to elucidate biological processes and predict therapeutic outcomes (Zahir et al., 2020), (Lu et al., 2019). A promising utility of mIHC is to locate the programmed cell death-1 (PD-1) and assess its correlation with CD8+, CD4+ and FOXP3+ tumour-infiltrating lymphocytes, which are of great interest to pathologists given the role of PD1 as an immune checkpoint protein and an important target of immunotherapy (Diana et al., 2016), (Halse et al., 2018).

To automatically identify the expression of antigens, several deep learning models have been developed to detect and classify cell types in mIHC images (Hagos et al., 2019, 2021; Narayanan et al., 2021). However, these methods were designed to process the two tasks separately, which has several limitations. Firstly, the use of separate models is costly and inefficient in both computing time and resources, which constrains the usage of models in large scale datasets (Song et al., 2019). Moreover, classifying cells with limited local context features might lead to a reduction in performance. For example, assembling predictions from multiple neighbouring regions has been shown to produce more accurate results than classification performed on the central region alone (Sirinukunwattana et al., 2016). On the other hand, the cell detection model without considering class-wise abundance was shown to miss out on rare cell types in an unbalanced dataset (Hagos et al., 2021). As a result, models combining features from cell detection and classification are desired.

Recently, several efforts have been made to simultaneously detect and classify cells on H&E images. Song et al. (2019) proposed a synchronized asymmetric deep-learning framework to parallelly detect and classify cells in bone marrow specimens. Graham et al. (2019) constructed a three-branch network to concurrently segment and classify nuclei on histology images. However, both strategies required exhaustive manual annotations of nuclei boundary, which might not be essential for cell classification in mIHC samples since distinct types of cells were readily marked by chromogenic staining. In fact, Fassler et al. (2020) has demonstrated that a cell segmentation model trained on arbitrary cell masks generated using only annotations of cell centres could achieve desired performance in mIHC images.

Despite that cells are distinguishable by the expression of markers, it is still nontrivial to identify cell types on mIHC slides. Specifically, some cell types may express multiple markers, and some exhibit high variability in staining intensities. One solution to overcome the intermix of staining is to deconvolve colours into separate channels, then identify cells based on the combined positive or negative signal in different channels (Fassler et al., 2020; Blom et al., 2017; Chen and Srinivas, 2015; Duggal et al., 2017; Abousamra et al., 2020; Lahiani et al., 2018). However, such a strategy requires prior knowledge about the color range associated with each marker, which limits its application in cases where staining colour spans a broad spectrum. Moreover, methods relying on colour deconvolution and segmentation of homogeneous colours are insufficient to identify individual cells in close proximity, which compromise the accuracy of spatial analysis involving distributions of single cells (Fassler et al., 2020; Lahiani et al., 2018).

To overcome these shortcomings of previous methods, we proposed a neural network to detect and classify cells simultaneously on mIHC slides. We compared the performance of different network structures and evaluated the dependency between detection and clas-

sification tasks. The proposed method outperforms baseline models trained separately on cell classification and detection. To the best of our knowledge, this is the first end-to-end solution for detecting and classifying cells in mIHC images.

2. Material and Methods

2.1. Dataset and annotations

Our dataset contained 9 whole slide images of high-grade serious ovarian cancer stained for CD8, CD4, FOXP3 and PD1. To train and validate the proposed method, a total of 3674 single cell annotations were collected from 9 slides with experts putting colourful dots at the cell centre to label 6 dominant cell types distinguished by expressions of markers. For sake of brevity, we denoted the 6 cell types as CD8+, CD4+FOXP3-, CD4+FOXP3+, PD1+CD4-CD8-, PD1+CD4-CD8+, and PD1+CD4+CD8-. Antigens expressed by each cell type were detailed in Table 1. Examples of the annotated cells were shown in Figure 1b. A total of 257 image patches containing annotations were extracted from 4 slides and randomly split with a 7:3 ratio into training (180) and validation (77) datasets. Cells from the other 5 slides were used for testing. The composition of each dataset was described in Table 1.

Table 1: Number of cells in training, validation, and test datasets.

Antigen expression				Cell types	Training	Validation	Test
PD1	CD8	CD4	FOXP3				
-	+	-	-	CD8+	110	33	344
-	-	+	-	CD4+FOXP3-	380	137	515
-	-	+	+	CD4+FOXP3+	192	110	368
+	-	-	-	PD1+CD4-CD8-	33	26	21
+	+	-	-	PD1+CD4-CD8+	453	183	612
+	-	+	-	PD1+CD4+CD8-	50	33	74

2.2. Training data preparation

Slides were scanned at 40x magnification and rescaled to 20x with a resolution of 0.44 µm/pixel. Regions with cell annotations were cropped into 224×224 patches $I \in \mathbb{R}^{224\times224\times1}$ with a stride of 120 pixels. For classification, we generated a binary mask per cell type $M_k \in \mathbb{R}^{224\times224\times1}$ by marking a circle area with a radius $r = 2$ pixels centred at each dot annotation. The distance threshold r was set empirically to cover a considerable cell area while avoiding the overlap of adjacent annotations. M_k for a total of 6 cell classes were stacked into a classification mask $M_c \in \mathbb{R}^{224\times224\times6}$. We generated the detection mask $M_d \in \mathbb{R}^{224\times224\times1}$ by taking the maximum values of M_c channel-wise, and a background mask $M_b \in \mathbb{R}^{224\times224\times1}$ as the inversion of M_d. The detection mask labelled locations of all cells regardless of cell types, and the background mask represented tissue regions without identified cells. We added the background mask to M_c as the 7^{th} channel. The i^{th} input in the training dataset was represented as (I^i, M_c^i, M_d^i). Data augmentation was performed by randomly flipping the input horizontally and vertically.

2.3. Network architecture and training

We proposed a fully convolutional network Symmetric Distance Regularized Dense Inception neural Network (S-DRDIN) to simultaneously detect and classify 6 immune cell types in the mIHC images. The network structure was built on the original DRDIN (Narayanan et al., 2021), with modifications to the decoder to enable parallel predictions for cell classes and locations.

As shown in Figure 1, the network was constructed following a U-net structure (Ronneberger et al., 2015) with inception blocks (Szegedy et al., 2015) as the basic convolution module. The encoder comprised 4 inception blocks linked by 2×2 averaged pooling layers to down sample features by a factor of 2 after each convolutional step. In comparison with the original DRDIN which comprised a single-branch decoder, S-DRDIN has two decoders to predict for class map and detection map simultaneously.

As suggested by Song et al. (2019), cell classification relied on less information than detection. To reduce the potential deleterious impact on classification induced by redundant information learnt from the detection task, we applied stop-gradient to all the skip connections between the encoder path and the detection branch (Figure 1). This operation prohibited the direct gradient transfer from the detection branch to the encoder, thereby disentangled the detection and classification information for low-level features and encouraged the network to optimize the feature map for the cell classification task. To compensate for the potential loss of useful information caused by the stop-gradient operation, outputs of the last inception blocks of the classification branch was incorporated into the detection branch, followed by two convolution layers with 3×3 kernels and a final convolution layer with 1x1 kernels to generate the detection output. For the classification branch, the last convolution layer was immediately added after the last inception block to produce an output of size $224 \times 224 \times 7$. Relu activation was applied to all the convolution layers except for the last convolution layers of classification and detection branches, which were activated with Softmax and Sigmoid respectively. The model was trained for 100 epochs with a batch size of 4. Weights were initialized using uniform glorot (Glorot and Bengio, 2010) and optimized using Adam (Kingma and Ba, 2015) with a learning rate of 10^{-3}.

2.4. Loss function for cell classification and detection

We used cross-entropy to calculate loss between the classification branch output and class map. To tackle the imbalance of class occurrences, the loss of each pixel was assigned with a weight w_c calculated as the inverse proportion of the number of pixels of the corresponding class in the batch, as given by,

$$w_c = \frac{N \cdot N_c \cdot B}{\sum_{i=1}^{N} x_i^c} \tag{1}$$

where N_c and N denoted the number of channels and number of pixels in each channel of the class map input. B is the batch size and x_i^c is the pixel value in channel c. The pixel-weighted cross-entropy loss function is defined as

$$L_{class} = -\frac{1}{N_c \cdot N} \sum_{c=1}^{N_c} \sum_{i=1}^{N} w_c \cdot \log(p_c(x_i)) \tag{2}$$

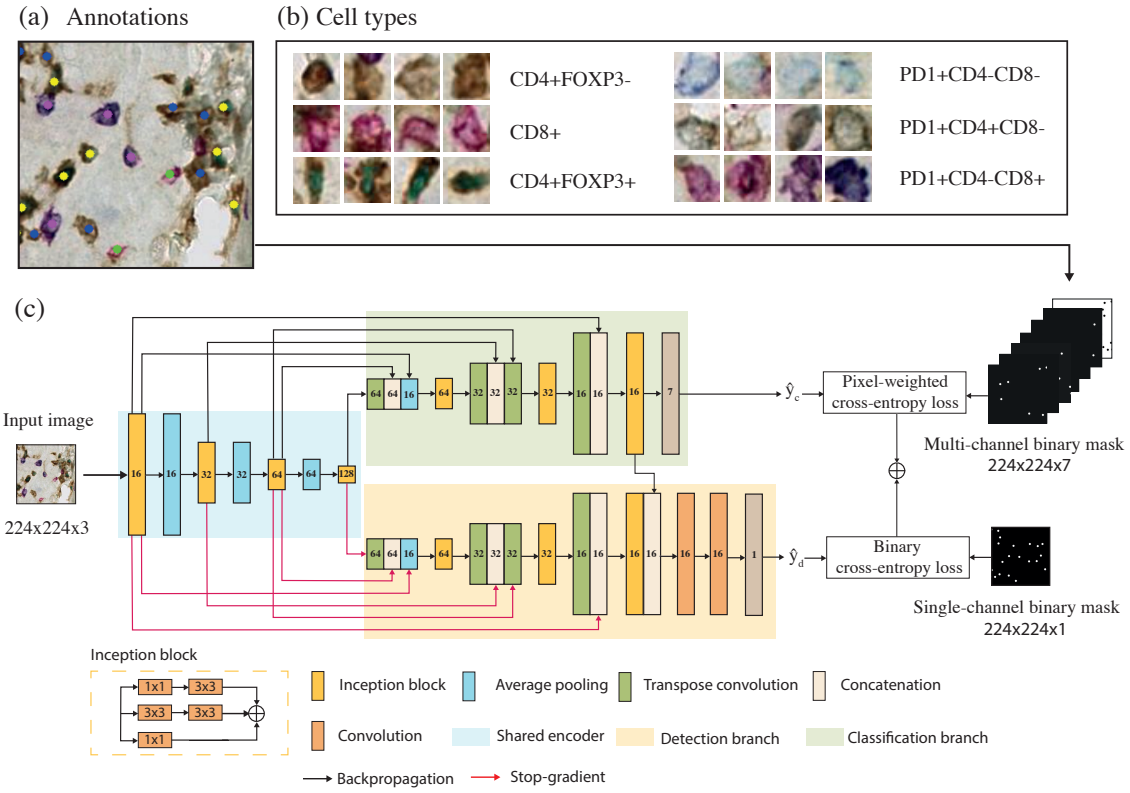

Figure 1: Framework of the proposed pipeline.

where $p_c(x_i)$ denotes the predicted probability from the classification branch after the Softmax activation. We separately calculated the binary cross-entropy for the detection branch. The model trained with a combined loss function from both cell detection and classification is referred to as the proposed S-DRDIN. To evaluate the intra-influence between the two tasks, we also tested models trained on detection and classification separately and trained without the stop-gradient operation. The three model variants are defined as follows,

1. DRDIN-detection: does not contain the classification branch
2. DRDIN-classification: does not contain the detection branch
3. S-DRDIN-gradient: allows back-propagation for all connections between encoder layers and the detection branch

2.5. Post-processing and evaluation

To obtain the locations of detected cells on the slide, we applied a post-processing pipeline to the probability map generated from the detection branch. Firstly, a threshold optimized

for the F1-score of each model was applied to binarize the predicted detection mask. Next, the connected components with an area smaller than 50 were discarded to remove noise in the background. Then we calculated the distance transform and identify the local maxima with a sliding window of size 15×15. The local maxima were dilated by a disk with a radius of 2 pixels. Lastly, centres of the instances were recorded as detected cell locations.

For cell classification, predicted values from each of the 6 cell-type channels were averaged for the 49×49 square region surrounded each identified cell. The predicted cell class k was determined by

$$\hat{k} = \underset{k}{\operatorname{argmax}} \frac{1}{N} \sum_{i=1}^{N} p_c^k(x_i) \tag{3}$$

where $p_c^k(x_i)$ represents the network output for pixel x_i at the channel corresponding to class k. N is the total number of pixels within the square region surrounding the detected cell. For robust evaluation of model performance, we trained each model separately on 3 datasets with training and validation split at different randomized states. The average model performance was reported for the 5 hold-out testing slides containing expert annotations on every identifiable cell within a given region. A predicted cell inside the region was considered as true positive if it fell within 10 pixels to an expert annotation, otherwise false positive. The false negative was counted as the number of annotated cells missed out by the model. The precision, recall and F1-score were reported for cell detection evaluation. Performance of cell classification was assessed for all annotated cells detected by a model, with precision, recall and F1-score weighted by the proportion of cell types computed for model comparisons.

3. Results

3.1. Cell detection performance

We evaluate the detection performance for different variants of the proposed S-DRDIN, and compared it against other states of the art U-Net (Ronneberger et al., 2015) and CONCORDe-Net (Hagos et al., 2019). The proposed S-DRDIN obtained an F1 score of 0.835, which was 2.7% and 6.7% higher than the U-Net and the CONCORDe-Net respectively. It also outperformed DRDIN-detection and S-DRDIN-gradient by 1.9% and 4.2% (Table 2). The improvement was mainly attributed to the increase in precision and the reduction in false positives, while S-DRDIN was less sensitive to true positive cells as compared to U-Net, DRDIN-detection and S-DRDIN-gradient, as reflected by the lower recall (Table 2). Specifically, the model tended to overlook CD4+FOXP3- and PD1+CD4+CD8- cells, with both classes showing a detection recall lower than 0.6 (Table A.5). These results were not fully explained by the data imbalance, as CD4+FOXP3- was the second most abundant class, and the averaged recall obtained for PD1+CD4-CD8- cells was 0.833 despite it being the rarest cell type Table 1. It is likely that the staining colour for CD4+FOXP3- and PD1+CD4+CD8- makes them less distinguishable from the tissue background as compared to other cell types (Figure A.3).

Interestingly, the S-DRDIN-gradient model, which jointly learned from classification and detection without stopping backpropagation for the detection branch, achieved the best recall but performed worse than the DRDIN-detection concerning the F1-score. This

observation was consistent with previous findings in Song et al. (2019) where the intra-influence between detection and classification reduces performance for both tasks. Here we demonstrated the stop-gradient operation as an effective approach to reduce conflicts between the two tasks and restore the detection accuracy. A potential explanation was that blocking the gradient flow from detection branch derived feature maps more representative for cell class identities, which limited the detection for cells with ambiguous class identity, therefore generated a more accurate cell map.

Table 2: Cell detection performance evaluated for different models.

Methods	Precision	Recall	F1-score	True positives	False positives
U-Net	0.848 ± 0.011	0.772 ± 0.016	0.808 ± 0.003	1493 ± 30	268 ± 28
CONCORDe-Net	0.921 ± 0.037	0.660 ± 0.017	0.768 ± 0.001	1276 ± 33	111 ± 58
DRDIN-detection	0.851 ± 0.017	0.783 ± 0.021	0.816 ± 0.004	1515 ± 41	266 ± 43
S-DRDIN-gradient	0.783 ± 0.098	$\mathbf{0.809 \pm 0.039}$	0.793 ± 0.032	1565 ± 76	456 ± 274
S-DRDIN	$\mathbf{0.941 \pm 0.016}$	0.750 ± 0.031	$\mathbf{0.835 \pm 0.013}$	1450 ± 61	92 ± 30

To better understand the contribution of stop-gradient operation for cell detection, we qualitatively compared the feature maps produced by the last layer of the encoder of different models. As shown in Figure A.2, activation produced by S-DRDIN was more localized to cell regions as compared to S-DRDIN-gradient and DRDIN-detection, being consistent with our hypothesis that stop-gradient functioned in suppressing the redundant information introduced by the backpropagation of the detection branch. The activation map indicated that these features might relate to the false positives in tissue background. Additionally, both S-DRDIN and S-DRDIN-gradient showed activating signals for a larger number of cells than the DRDIN-classification, suggesting the advantages of additional information introduced by the parallel learning from detection and classification.

3.2. Cell classification evaluation

We compared performance for cell classification among different network and training strategies. Both the U-Net and DRDIN-classification were trained on the 7-channel binary masks described in subsection 2.2, with U-Net trained without additional weights assigned to the loss function. Classification for the CONCORDe-Net was performed using an SCCNN classifier (Sirinukunwattana et al., 2016) trained on single-cell patches extracted from the annotations. S-DRDIN achieved the highest score for weighted precision (0.870), weighted recall (0.872) and weighted F1-score (0.867) among all the five models, including U-Net, Concordent-Net and DRDIN-classification that were trained separately on detection and classification (Table 3). The training time of S-DRDIN was shorter than models trained on discrete tasks except for the U-Net, which was due to the difference in parameters (Table A.4). Predictive accuracy of CD8+, PD1+CD4-CD8- and PD1+CD4+CD8- was sub-optimal compared to that of the other cell type, which was likely due to the limitation of sample size (Table A.5). On the other hand, S-DRDIN-gradient also outperforms the models trained on separate tasks, suggesting that the incorporation of cell location into the training stage improves cell classification performance. Additionally, the employment of the pixel-weighted cross-entropy increased the weighted F1-score by at least 5% than the

baseline models, demonstrating the efficiency of pixel weighting strategy for reducing the negative influence of data imbalance. Examples of cell detection and classification results were shown in Figure A.3. Noted that S-DRDIN was the only model which successfully identified the rare PD1+CD4-CD8- cell in the region.

Table 3: Cell classification performance of different methods.

Methods	Weighted Precision	Weighted Recall	Weighted F1-score
U-Net	0.768 ± 0.006	0.795 ± 0.002	0.761 ± 0.006
CONCORDe-Net	0.797 ± 0.004	0.816 ± 0.009	0.778 ± 0.011
DRDIN-classification	0.839 ± 0.011	0.834 ± 0.009	0.829 ± 0.018
S-DRDIN-gradient	0.860 ± 0.015	0.860 ± 0.018	0.857 ± 0.021
S-DRDIN	$\mathbf{0.870 \pm 0.011}$	$\mathbf{0.872 \pm 0.009}$	$\mathbf{0.867 \pm 0.015}$

4. Discussion

Our study aims at designing a deep-learning method for detecting and classifying cells simultaneously in mIHC samples. The proposed model exploited the mutual dependency of the two highly relevant tasks, and introduced a novel stop-gradient approach to reduce the conflict between detection and classification features. In comparison with the conventional deep learning pipeline that processed the two tasks as separate steps, the proposed model directly predicts locations of cells of different types on a given mIHC image, therefore reduces the computation costs and speeds up the analysis for large scale datasets. Moreover, we demonstrated that the combination of stop-gradient operation, pixel-weighted cross-entropy loss and parallel learning from detection and classification results in higher precision for detection as well as better performance for classification.

This study was limited by the small dataset of 3674 single cell annotations with an extreme class imbalance. In future work, we propose to evaluate the generalization of the model on larger multiplex image datasets from different sources. Also, additional modifications can be applied to the model architecture to improve the detection recall. Once fully developed, the method will be able to accelerate the image analysis for large cohorts and promote the understanding of the spatial relationship between diverse components of the tumour immune microenvironment.

5. Conclusion

We presented a network S-DRDIN for simultaneous cell detection and classification in mIHC images. With the aid of stop-gradient and a loss function accounted for class proportions, the proposed model achieved an F1 score of 0.835 in cell detection, and a weighted F1 score of 0.867 in classification, which were 1.9% and 3.8% higher than the model trained separately on individual tasks.

References

Shahira Abousamra, Danielle Fassler, Le Hou, Yuwei Zhang, Rajarsi Gupta, Tahsin Kurc, Luisa F. Escobar-Hoyos, Dimitris Samaras, Beatrice Knudson, Kenneth Shroyer, Joel Saltz, and Chao Chen. Weakly-supervised deep stain decomposition for multiplex ihc images. volume 2020-April, pages 481–485. IEEE Computer Society, 4 2020. ISBN 9781538693308. doi: 10.1109/ISBI45749.2020.9098652.

Sami Blom, Lassi Paavolainen, Dmitrii Bychkov, Riku Turkki, Petra Mäki-Teeri, Annabrita Hemmes, Katja Välimäki, Johan Lundin, Olli Kallioniemi, and Teijo Pellinen. Systems pathology by multiplexed immunohistochemistry and whole-slide digital image analysis. *Scientific Reports*, 7(1):1–13, 12 2017. ISSN 20452322. doi: 10.1038/s41598-017-15798-4.

Ting Chen and Chukka Srinivas. Group sparsity model for stain unmixing in brightfield multiplex immunohistochemistry images. *Computerized Medical Imaging and Graphics*, 46:30–39, 12 2015. ISSN 18790771. doi: 10.1016/j.compmedimag.2015.04.001.

Angela Diana, Lai Mun Wang, Zenobia D'Costa, Paul Allen, Abul Azad, Michael A. Silva, Zahir Soonawalla, Stanley Liu, W. Gillies McKenna, Ruth J. Muschel, and Emmanouil Fokas. Prognostic value, localization and correlation of pd-1/pd-l1, cd8 and foxp3 with the desmoplastic stroma in pancreatic ductal adenocarcinoma. *Oncotarget*, 7(27):40992–41004, 7 2016. ISSN 19492553. doi: 10.18632/oncotarget.10038.

Rahul Duggal, Anubha Gupta, Ritu Gupta, and Pramit Mallick. Sd-layer: Stain deconvolutional layer for cnns in medical microscopic imaging. volume 10435 LNCS, pages 435–443. Springer Verlag, 9 2017. ISBN 9783319661780. doi: 10.1007/978-3-319-66179-7_50.

Danielle J. Fassler, Shahira Abousamra, Rajarsi Gupta, Chao Chen, Maozheng Zhao, David Paredes, Syeda Areeha Batool, Beatrice S. Knudsen, Luisa Escobar-Hoyos, Kenneth R. Shroyer, Dimitris Samaras, Tahsin Kurc, and Joel Saltz. Deep learning-based image analysis methods for brightfield-acquired multiplex immunohistochemistry images. *Diagnostic Pathology*, 15(1):100, 7 2020. ISSN 17461596. doi: 10.1186/s13000-020-01003-0.

Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. Technical report, 3 2010.

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*, 58:101563, 12 2019. ISSN 13618423. doi: 10.1016/j.media.2019.101563.

Yeman Brhane Hagos, Priya Lakshmi Narayanan, Ayse U. Akarca, Teresa Marafioti, and Yinyin Yuan. Concorde-net: Cell count regularized convolutional neural network for cell detection in multiplex immunohistochemistry images. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11764 LNCS:667–675, 8 2019.

Yeman Brhane Hagos, Catherine SY Lecat, Dominic Patel, Lydia Lee, Thien-An Tran, Manuel Rodriguez Justo, Kwee Yong, and Yinyin Yuan. Cell abundance aware deep

learning for cell detection on highly imbalanced pathological data. *Proceedings - International Symposium on Biomedical Imaging*, 2021-April:1438–1442, 2 2021.

H. Halse, A. J. Colebatch, P. Petrone, M. A. Henderson, J. K. Mills, H. Snow, J. A. Westwood, S. Sandhu, J. M. Raleigh, A. Behren, J. Cebon, P. K. Darcy, M. H. Kershaw, G. A. McArthur, D. E. Gyorki, and P. J. Neeson. Multiplex immunohistochemistry accurately defines the immune context of metastatic melanoma. *Scientific Reports*, 8(1): 11158, 12 2018. ISSN 20452322. doi: 10.1038/s41598-018-28944-3.

Diederik P. Kingma and Jimmy Lei Ba. Adam: A method for stochastic optimization. International Conference on Learning Representations, ICLR, 12 2015.

Amal Lahiani, Jacob Gildenblat, Irina Klaman, Nassir Navab, and Eldad Klaiman. Generalizing multistain immunohistochemistry tissue segmentation using one-shot color deconvolution deep neural networks. *IET Image Processing*, 13(7):1066–1073, 5 2018.

Steve Lu, Julie E. Stein, David L. Rimm, Daphne W. Wang, J. Michael Bell, Douglas B. Johnson, Jeffrey A. Sosman, Kurt A. Schalper, Robert A. Anders, Hao Wang, Clifford Hoyt, Drew M. Pardoll, Ludmila Danilova, and Janis M. Taube. Comparison of biomarker modalities for predicting response to pd-1/pd-l1 checkpoint blockade: A systematic review and meta-analysis. *JAMA Oncology*, 5(8):1195–1204, 8 2019. ISSN 23742445. doi: 10.1001/jamaoncol.2019.1549.

Priya Lakshmi Narayanan, Shan E.Ahmed Raza, Allison H. Hall, Jeffrey R. Marks, Lorraine King, Robert B. West, Lucia Hernandez, Naomi Guppy, Mitch Dowsett, Barry Gusterson, Carlo Maley, E. Shelley Hwang, and Yinyin Yuan. Unmasking the immune microecology of ductal carcinoma in situ with deep learning. *npj Breast Cancer*, 7(1):1–14, 12 2021. ISSN 23744677. doi: 10.1038/s41523-020-00205-5.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. volume 9351, pages 234–241. Springer Verlag, 5 2015. ISBN 9783319245737. doi: 10.1007/978-3-319-24574-4_28.

Korsuk Sirinukunwattana, Shan E.Ahmed Raza, Yee Wah Tsang, David R.J. Snead, Ian A. Cree, and Nasir M. Rajpoot. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE Transactions on Medical Imaging*, 35(5):1196–1206, 5 2016. ISSN 1558254X. doi: 10.1109/TMI.2016.2525803.

Tzu Hsi Song, Victor Sanchez, Hesham Eidaly, and Nasir M. Rajpoot. Simultaneous cell detection and classification in bone marrow histology images. *IEEE Journal of Biomedical and Health Informatics*, 23(4):1469–1476, 7 2019. ISSN 21682208. doi: 10.1109/JBHI.2018.2878945.

Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. volume 07-12-June-2015, pages 1–9. IEEE Computer Society, 10 2015. ISBN 9781467369640. doi: 10.1109/CVPR.2015.7298594.

Nastaran Zahir, Ruping Sun, Daniel Gallahan, Robert A. Gatenby, and Christina Curtis. Characterizing the ecological and evolutionary dynamics of cancer. *Nature Genetics*, 52 (8):759–767, 8 2020. ISSN 15461718. doi: 10.1038/s41588-020-0668-4.

Appendix A.

Table A.4: Training time of different methods for 100 epochs.

Methods	Training time
U-Net-detection+classification	0.09 hours
Concordenet-detection+classification	0.73 hours
DRDIN-detection+classification	0.16 hours
S-DRDIN	0.12 hours

Table A.5: Cell classification performance and detection recall of the proposed S-DRDIN for different cell types. Bold values highlight scores lower than 0.8.

Cell types	Precision	Recall	F1-score	Detection Recall
CD8+	0.833 ± 0.025	$\mathbf{0.683 \pm 0.069}$	$\mathbf{0.749 \pm 0.032}$	$\mathbf{0.731 \pm 0.113}$
CD4+FOXP3-	0.910 ± 0.059	0.954 ± 0.014	0.931 ± 0.024	$\mathbf{0.525 \pm 0.103}$
CD4+FOXP3+	0.967 ± 0.002	0.974 ± 0.019	0.970 ± 0.009	0.865 ± 0.017
PD1+CD4-CD8-	$\mathbf{0.675 \pm 0.035}$	$\mathbf{0.750 \pm 0.039}$	$\mathbf{0.711 \pm 0.037}$	0.833 ± 0.034
PD1+CD4-CD8+	0.848 ± 0.032	0.911 ± 0.018	0.877 ± 0.009	0.814 ± 0.062
PD1+CD4+CD8-	$\mathbf{0.646 \pm 0.065}$	$\mathbf{0.468 \pm 0.343}$	$\mathbf{0.514 \pm 0.264}$	$\mathbf{0.527 \pm 0.019}$

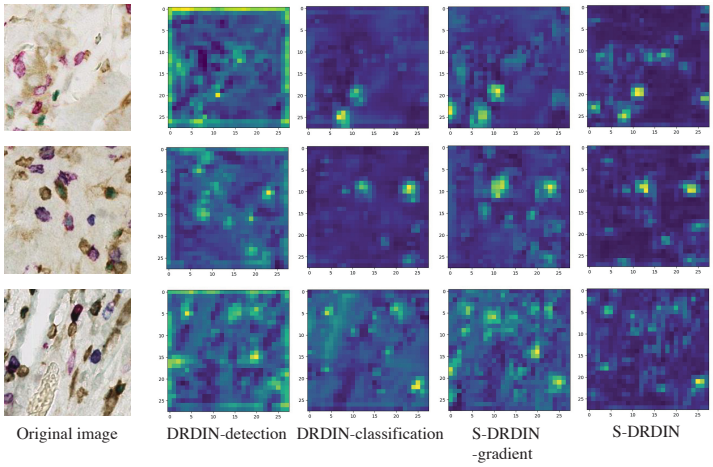

Figure A.2: A qualitative comparison of feature maps generated by the last layer of the encoder of different models. Yellow indicates high value.

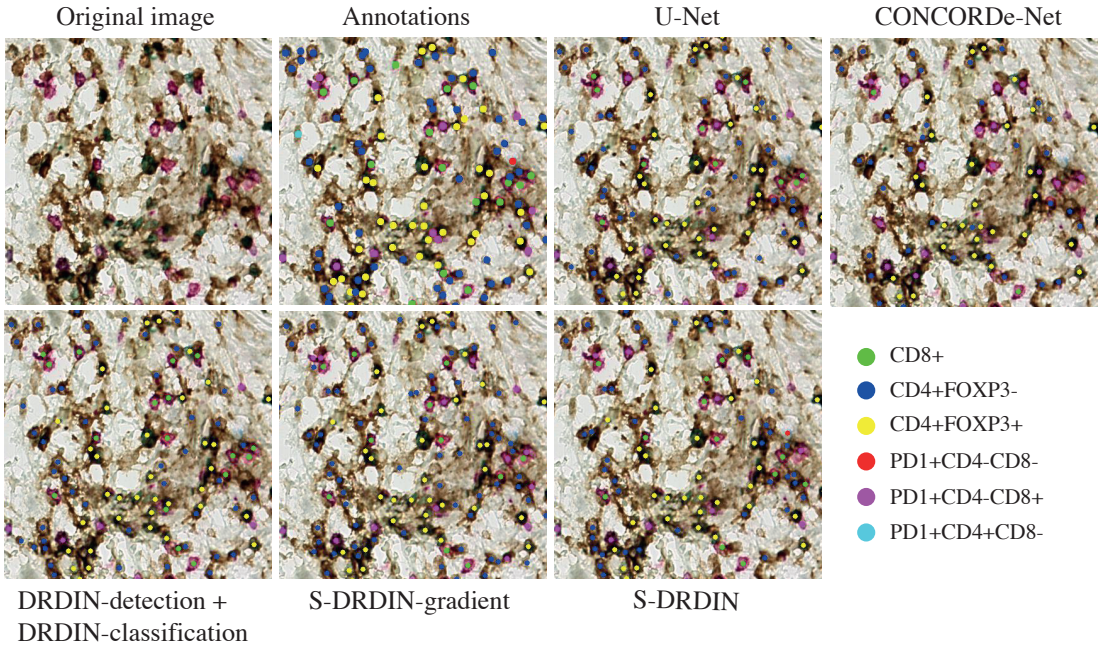

Figure A.3: Examples of cell detection and classification outputs from different methods.