University of Groningen

MoNuSAC2020

Verma, Ruchika; Kumar, Neeraj; Patil, Abhijeet; Kurian, Nikhil Cherian; Rane, Swapnil; Graham, Simon; Vu, Quoc Dang; Zwager, Mieke; Raza, Shan E.Ahmed; Rajpoot, Nasir

Published in:
Ieee transactions on medical imaging

DOI:
10.1109/TMI.2021.3085712

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2021

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Verma, R., Kumar, N., Patil, A., Kurian, N. C., Rane, S., Graham, S., Vu, Q. D., Zwager, M., Raza, S. E. A., Rajpoot, N., Wu, X., Chen, H., Huang, Y., Wang, L., Jung, H., Thomas Brown, G., Liu, Y., Liu, S., Jahromi, S. A. F., ... Sethi, A. (2021). MoNuSAC2020: A Multi-Organ Nuclei Segmentation and Classification Challenge. *Ieee transactions on medical imaging*, *40*(12), 3413-3423. https://doi.org/10.1109/TMI.2021.3085712

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 12-10-2022

IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 40, NO. 12, DECEMBER 2021

MoNuSAC2020: A Multi-Organ Nuclei Segmentation and Classification Challenge

Ruchika Verma, Neeraj Kumar, Abhijeet Patil, Nikhil Cherian Kurian, Swapnil Rane,
Simon Graham, Quoc Dang Vu, Mieke Zwager, Shan E. Ahmed Raza, Nasir Rajpoot, Xiyi Wu,
Huai Chen, Yijie Huang, Lisheng Wang, Hyun Jung, G. Thomas Brown, Yanling Liu, Shuolin Liu,
Seyed Alireza Fatemi Jahromi, Ali Asghar Khani, Ehsan Montahaei, Mahdieh Soleymani Baghshah,
Hamid Behroozi, Pavel Semkin, Alexandr Rassadin, Prasad Dutande, Romil Lodaya, Ujjwal Baid,
Bhakti Baheti, Sanjay Talbar, Amirreza Mahbod, Rupert Ecker, Isabella Ellinger, Zhipeng Luo, Bin Dong,
Zhengyu Xu, Yuehan Yao, Shuai Lv, Ming Feng, Kele Xu, Hasib Zunair, Abdessamad Ben Hamza,
Steven Smiley, Tang-Kai Yin, Qi-Rui Fang, Shikhar Srivastava, Dwarikanath Mahapatra,
Lubomira Trnavska, Hanyun Zhang, Priya Lakshmi Narayanan, Justin Law, Yinyin Yuan,
Abhiroop Tejomay, Aditya Mitkari, Dinesh Koka, Vikas Ramachandra,
Lata Kini, and Amit Sethi

***Abstract*—Detecting various types of cells in and around the tumor matrix holds a special significance in characterizing the tumor micro-environment for cancer prognostication and research. Automating the tasks of detecting, segmenting, and classifying nuclei can free up the pathologists' time for higher value tasks and reduce errors due to fatigue and subjectivity. To encourage the computer vision research community to develop and test algorithms for these tasks, we prepared a large and diverse dataset of nucleus boundary annotations and class labels. The dataset has over 46,000 nuclei from 37 hospitals, 71 patients, four organs, and four nucleus types. We also organized a challenge around this dataset as a satellite event at the International Symposium on Biomedical Imaging (ISBI) in April 2020. The challenge saw a wide participation from across the world, and the top methods were able to match inter-human concordance for the challenge metric. In this paper, we summarize the dataset and the key findings of the challenge, including the commonalities and differences between the methods developed by various participants. We have released the MoNuSAC2020 dataset to the public.**

***Index Terms*—Multi-organ dataset, nucleus classification, computational pathology, instance segmentation, panoptic quality.**

I. INTRODUCTION

A TUMOR'S microenvironment (TME) is characterized by the presence and spatial organization of different types of cells in and around a tumor. To assess the role of the TME milieu in tumor initiation, development, invasion, and outcome, it is important to accurately identify different types of cells. For example, while the *epithelial cells* are important because many cancers originate there, the tumor infiltrating *lymphocytes* (TILs) are associated with the outcomes of various cancers [1], [2]. Similarly, the tumor associated *macrophages* (TAMs) and *neutrophils* also influence diverse processes such as angiogenesis, neoplastic cell mitogenesis, antigen presentation, matrix degradation, and cytotoxicity in various tumors [3].

Although hematoxylin and eosin (H&E) stained tissue slides are often used for the primary diagnosis due to their low cost, they are not very good for differentiating nuclei of different types of cells. For instance, epithelial nuclei within the tumor vary in their appearance across organs, and these can appear small and dense, just like lymphocytes, within a necrotic tissue. Macrophages can be mistaken for the other cells that they may have just started digesting. The multi-lobular structure of neutrophils can be missed depending on their angle of rotation. More importantly, the manual identification of individual cells takes substantial time and is prone to inter- and intra-observer variability. Thus, automated detection of nuclei of different types of cells using artificial intelligence (AI) can reduce time, costs, and errors to set up for the analysis of the TME.

To build generalizable and robust algorithms for segmenting and recognizing nuclei of various types of cells, the computer vision community needs a large, diverse, curated, and annotated dataset. Previously, several public datasets have been released for nucleus segmentation, but the annotated nuclei were either not further classified into different types [4]–[8], or not sourced from multiple organs [8], or not curated to the level of individual nuclei [9]. We, therefore, prepared the multi-organ nucleus segmentation and classification (MoNuSAC) dataset with over 46,000 hand-annotated nuclei covering 71 patients, 31 hospitals, four organs, and four cell types. Our focus was on identifying epithelial and three

Manuscript received April 18, 2021; accepted May 23, 2021. Date of publication June 4, 2021; date of current version November 30, 2021. The work of Simon Graham, Shan E. Ahmed Raza, and Nasir Rajpoot was supported by the Data to Early Diagnosis and Precision Medicine Strand of the Government's Industrial Strategy Challenge Fund, managed and delivered by U.K. Research and Innovation (UKRI) and the authors are a part of the PathLAKE Digital Pathology Consortium. The work of Amirreza Mahbod was supported by the Austrian Research Promotion Agency (FFG) under Grant 872636. *(Corresponding author: Ruchika Verma.)*

Please see the Acknowledgment section of this article for the author affiliations.

This article has supplementary downloadable material available at https://doi.org/10.1109/TMI.2021.3085712, provided by the authors.

Digital Object Identifier 10.1109/TMI.2021.3085712

types of immune cells that are important for characterizing the TME. We sourced the raw images from The Cancer Genome Atlas (TCGA) repository [10].

Based on the dataset, the computer vision community was invited to participate in the MoNuSAC2020 challenge that featured as a satellite event in the International Symposium on Biomedical Imaging (ISBI) 2020. The challenge webpage was hosted at grand-challenges.org – a well-known host and portal for biomedical image analysis challenges [11]. Thirteen teams from 11 countries successfully completed the challenge. Post-challenge submissions were opened for another month after the challenge workshop due to a popular demand and the COVID-19 situation, which attracted four additional submissions and two resubmissions. Computational credits for the top ten teams to refine their techniques and a GPU prize for the challenge winner were sponsored by the Nvidia Corporation.

After the challenge, we confirmed that the top few techniques compared favorably with inter-human agreement for the challenge metric. We also checked that ensembling the top few techniques led to only a modest improvement in accuracy. Thus, we hope that the challenge dataset and competing algorithms will be used in the near future to characterize the TME for cancer prognostication and immune response assessment. With this paper, we have released the details of the competing techniques and MoNuSAC training and testing dataset to the public under the creative commons license (CC BY-NC-SA 4.0) [11].

We revisit other data sources and techniques for nucleus segmentation and classification in Section II. We describe the preparation of MoNuSAC2020 training and testing datasets, and the choice of evaluation metric in Section III. We summarize the techniques developed by the challenge participants and their results in Section IV, post-challenge experiments in Section V, and the conclusions in Section VI.

II. BACKGROUND AND PRIOR WORK

A. Nucleus Segmentation and Classification

Earlier nucleus segmentation algorithms largely used non-machine learning approaches, such as watershed segmentation, morphological operations, color-based thresholding, and variants of active contours [7], [12]–[15]. However, these methods failed to generalize across diverse nuclear morphologies present in multiple organs and tissue types. Furthermore, these algorithms did not perform well for segmenting touching and overlapping nuclei [5]. Due to these reasons several machine learning algorithms were developed for accurate and generalizable nucleus segmentation. Earlier machine learning approaches focused on the extraction of hand-crafted features based on color, intensity, and texture from H&E stained tissue images to distinguish between nuclear and non-nuclear pixels using traditional algorithms such as random forests, support vector machines, etc. [16]–[18]. However, feature selection for such algorithms relied on domain knowledge and such algorithms also used trial-and-error for improving nucleus segmentation performance. Additionally, these methods treated simultaneous nucleus segmentation and classification as two distinct decoupled tasks. More detailed reviews of these nucleus segmentation algorithms appear in [4], [5], [19], [20].

The advent of large annotated datasets and supervised deep learning models overcame the limitations of aforementioned hand-crafted feature based approaches. As algorithms based on deep learning that scale performance more effectively to the availability of labeled data started outperforming prior techniques, the focus of computer vision research shifted from engineering of features to that of neural network architectures and loss functions. For nucleus segmentation, there has been a recent interest in deep neural net architectures inspired from the fully convolutional neural networks (FCN) [5], [8], [21], [22]. These architectures involve the use of encoder-decoder blocks that could systematically transfer features from multiple scales and levels for efficient image segmentation. The U-Net model, a variant of FCN architecture, has given promising results for nucleus and other segmentation tasks [23]. The U-Net model has additional skip connections compared to the FCN for spatially refined semantic segmentation, when touching objects of the same kind need not be recognized as separate entities. Some modifications to the basic U-Net architectures, such as U-Net++ [24], utilize enhanced encoder-decoder architectures and loss functions to improve the *instance segmentation* of individual (possibly crowded) nuclei.

More robust architectures have also been proposed such as W-Net, which uses a two-step procedure for contour detection and segmentation map generation for improved nucleus segmentation [25]. Some of the other techniques that make use of the nuclear contours for effective instance segmentation are based on a unified multi-level FCN framework with contour map generation for separating overlapping nuclei instances [26], and a contour-seed pairs learning for robust nucleus segmentation [27]. Another set of popular techniques adopt mask-RCNNs [28] from the general vision literature for robust instance level nucleus segmentation [29]–[31]. These methods make use of an additional region proposal step to identify potential nucleus locations and following it up with a segmentation step.

For nucleus classification, usually a patch centered around a nucleus is input to a neural network, which makes detection or segmentation a prerequisite task. For example, SC-CNN uses a novel neighboring ensemble predictor to label individual nuclei that follows detection by a spatially constrained convolutional neural network [32], while Hover-Net performs simultaneous nucleus segmentation and classification using horizontal and vertical distance maps to resolve and separate clustered nuclei [33]. Recently, mask-RCNN has also been used for the joint segmentation and classification of various types on nuclei in H&E stained tissue images of lung adencarcinoma [34]. An exhaustive list of other deep learning based nucleus segmentation and classification techniques can be found in [20].

B. Annotated Nuclear Datasets

The advent of deep learning for computer vision has also shifted the focus of research to the preparation of large labeled datasets such as ImageNet [35]. The computational

TABLE I
DATASETS OF ANNOTATED NUCLEI IN H&E IMAGES

Dataset	Nuclei	Label Type	Organs
Gelasca *et. al.* [37]	1,895	Semantic	1
Irshad *et. al.* [6]	2,532	Semantic	1
Naylor *et. al.* [8]	4,022	Semantic	1
Kainz *et. al.* [38]	4,205	Detection	1
CPM-17 [39]	7,570	Instance	1
Wienert *et. al.* [7]	7,921	Instance	5
Janowczyk *et. al.* [40]	12,000	Semantic	1
MoNuSeg [5]	21,623	Instance	7
CRCHisto [32]	22,444	Detection + Classification	1
CoNSeP [33]	24,319	Instance + Classification	1
PanNuke [9]	205,343	Instance + Classification	19
MoNuSAC	46,909	Instance + Classification	4

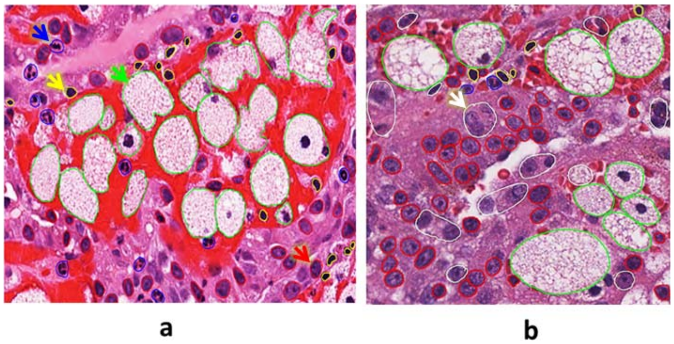

Fig. 1. Sample annotations: (a) A sub-image from the training dataset with boundaries of different types of nuclei shown in different colors – epithelial cells in red, lymphocytes in yellow, macrophages in green, and neutrophils in red, and (b) a test image with additional white boundaries for ambiguous regions. Some boundaries are pointed out for the reader by arrows of the same colors.

pathology community has been somewhat slow in adopting the trend of releasing large curated datasets for deep learning, primarily because of the expensive and time consuming data curation process, privacy concerns, and approvals required for publicly releasing de-identified human tissue datasets. The noteworthy publicly available datasets with nuclear annotations in H&E stained slides are listed in Table I. These datasets, with the exception of PanNuke [9], either cover only one organ or do not provide nucleus type labels, which is insufficient for training neural networks that generalize well for simultaneous nuclear segmentation and classification. And, although PanNuke seems to bridge this gap, its annotations were collected using a semi-automatic strategy. Thus, there is a possibility of nucleus annotation errors because it cannot be guaranteed that every nucleus was visually inspected during the pathologist verification. Furthermore, the PanNuke dataset does not contain separate labels for lymphocytes, macrophages and neutrophils, but it combined them into a single “inflammatory cells” category [9]. Large-scale regional annotations of different components of tissues are also available in the Atlas of Histopathology [36], although it lacks the individual nucleus-level annotations. Thus, there was a clear need for a curated multi-organ dataset of different nucleus types, especially one that differentiated between various types of immune cells given their importance in characterizing TME.

III. DATASET AND CHALLENGE RULES

The training and testing datasets, and the rules of the challenge were prepared to meet the objective of providing an open, large, usable, annotated, and curated dataset for the development and comparison of robust algorithms for nucleus instance segmentation and classification. For usability, we sampled multiple organs and concentrated on nucleus types important for a wide spectrum of studies on tumor initiation and progression, i.e., epithelial cells, lymphocytes, neutrophils, and macrophages [1]–[3].

A. Training Dataset

The training set was prepared by cropping whole slide images (WSIs) of 46 patients from 32 hospitals downloaded from The Cancer Genome Atlas (TCGA) data portal [10]. WSIs were scanned by the TCGA contributors at a 40× magnification. Cropping WSIs ensured that the annotations were diverse and of high quality as we could sample more slides and nucleus-rich regions. This also encouraged wider participation in the challenge due to the computational manageability of cropped images.

Annotations were performed using Aperio ImageScope® application, on a 25” monitor with a 200× digital magnification such that each image pixel occupied 5 × 5 screen pixels to ensure clear visibility for annotating nuclear boundaries with a laser mouse. We adopted the previously established protocol for manual nuclear boundary annotations as described in [4]. In addition to nuclear boundary annotations, we also provided nucleus class labels for each annotated nuclei included in MoNuSAC2020 dataset. For overlapping nuclei, each multi-nuclear pixel was assigned to the nucleus that appeared to be on top. Each type of nucleus was annotated using a unique colored marker. An example of the training set annotations is shown in Figure 1(a). Annotations were saved as XML files and the code for reading annotations was provided to the participants [11].

The annotators were engineering graduate students and the quality control was performed by an expert pathologist with several years of experience in analyzing tissue sections. Specifically, the pathologist reviewed the annotations of each nucleus type to point out missed nuclei, false nuclei, mislabeled nuclei, and nuclei with wrong boundaries. Nuclear annotations were iteratively revised based on the pathologist’s feedback, until less than 1% nuclei of any type had any of these errors. The distribution of annotated nuclei by nucleus type and organ in the training data is shown in Table II.

B. Testing Dataset

The testing data were prepared using the same procedure as that for the training data, except that it was sourced from 25 patients that did not overlap with the training data. This data was taken from 19 hospitals out of which 14 were shared with the training dataset. Additionally, the testing data also contained annotations for *ambiguous* regions. The ambiguous

TABLE II

COMPOSITION OF THE MONUSAC2020 TRAINING AND TESTING DATA BY ORGAN AND NUCLEUS TYPE (WITH THE NUMBER OF PATIENTS SHOWN IN PARENTHESIS)

Training Set					
Nucleus Type	**Breast** (N=10)	**Kidney** (N=12)	**Lung** (N=10)	**Prostate** (N=14)	**Total** (N=46)
Epithelial	4,566	3,547	2,981	3,445	**14,539**
Lymphocytes	4,689	4,126	3,018	3,821	**15,654**
Macrophoges	147	102	186	152	**587**
Neutrophils	105	149	142	235	**631**
					31,411

Testing Set					
Nucleus Type	**Breast** (N=5)	**Kidney** (N=7)	**Lung** (N=7)	**Prostate** (N=6)	**Total** (N = 25)
Epithelial	1,377	2,248	1,489	2,099	**7,213**
Lymphocytes	1,833	2,141	1,674	2,158	**7,806**
Macrophoges	49	84	53	121	**307**
Neutrophils	26	56	36	54	**172**
Ambiguous	908	222	764	509	**-**
					15,498

regions were not used for computing the evaluation metric for ranking the participants because these regions either had very faint nuclei with unclear boundaries, or the annotators were unsure of their true class. Participants were provided with the images and the annotations of the ambiguous regions but not the four nucleus types, while the latter were exclusively used to compute the challenge metric by the MoNuSAC2020 organizers. An example of the test set annotations with ambiguous regions in white is shown in Figure 1(b). The distribution of testing dataset across organs and nucleus types is shown in Table II.

The patient and hospital details for the training and the testing data are available in the supplementary material. With this paper, we have publicly released the complete training and testing datasets of MoNuSAC2020 including the testing set annotations, under the creative commons license (CC BY-NC-SA 4.0), on the challenge webpage [11].

C. Challenge Metric and Leaderboard

To evaluate both the semantic segmentation and the instance classification accuracy of the challenge participants using a single metric, the average of class-specific panoptic quality (a-PQ) was used [41]. For the i^{th} image and the c^{th} class of nuclei, the class-specific PQ was computed as the sum of intersection-over-union (IoU) of the matching pairs of the detected nuclei $p_c^{i,j}$, indexed by j, and the ground truth nuclei $g_c^{i,k}$, indexed by k. That is,

$$PQ_c^i = \frac{\sum_{(p_c^{i,j}, g_c^{i,k}) \in \{TP_c^i\}} IoU(p_c^{i,j}, g_c^{i,k})}{|TP_c^i| + \frac{1}{2}|FP_c^i| + \frac{1}{2}|FN_c^i|}, \qquad (1)$$

where TP_c^i, FP_c^i, and FN_c^i are respectively the sets of true positive, false positive, and false negative nuclei for the i^{th} image and the c^{th} class, and $|.|$ represents set cardinality. To be considered a match for inclusion in TP_c^i, the segmented and the ground-truth nucleus pair $\{p_c^{i,j}, g_c^{i,k}\}$ should have an $IoU > 0.5$, which would make the matching bidirectionally unique. For the i^{th} test image, arithmetic mean of the class-specific PQ_c^i over the number of classes (C) in that image gave the image-level PQ score, $PQ^i = \frac{1}{C}\sum_{c=1}^{C} PQ_c^i, \ \ for \ C \in \{1, 2, 3, 4\}$.

The value of PQ ranges between 0 to 1, where 0 means all nucleus instances were segmented incorrectly (no-overlap between the ground truth and segmented nuclei) at the pixel-level and each segmented nuclei was also incorrectly labeled, while 1 means perfect segmentation and classification. Participants submitted a separate output file for each of the 25 test images. PQ^i was assigned 0 if the file for the i^{th} test image was not submitted. Arithmetic mean of the 25 PQ^i scores formed the final average panoptic quality (a-PQ) score for ranking the participants.

Each participating team was allowed two submissions. Based on their first submissions, the organizers released the *preliminary leaderboard* on the challenge webpage [11] three weeks before the challenge workshop at ISBI 2020. The top ten teams on the preliminary leaderboard were given access to a cloud instance with two V100 GPUs (16GB VRAM each) for a week by the Nvidia Corporation to refine their models for the final submission. The *challenge leaderboard* for the rankings at the time of the challenge workshop at ISBI 2020 is shown in Table III (with a prefix L). The challenge workshop was organized virtually as a part of ISBI 2020, where the participants presented their algorithms (recordings are available at [11]). The winner of the challenge was awarded a TITAN V GPU from Nvidia Corporation. Keeping in view the COVID-19 situation and requests from the challenge workshop attendees, the challenge was re-opened for another month and the post-challenge participants were allowed only one submission. The post-challenge rankings are also shown in Table III (with a prefix PL).

Table III also shows the 95% confidence intervals (95% CIs) around a-PQ for each participant. It is evident that most of the techniques (especially, the top three techniques) have narrow CIs, indicating that the ranking based on a-PQ is stable. Fine-grained results are shown in Supplementary Table S2, where it is apparent that the ranking trends for the overall a-PQ are held at the level of organs and nuclear types as well. This means that nucleus segmentation and classification for some organs and certain types of nuclei were relatively

TABLE III

A SUMMARY OF FEATURES OF THE TECHNIQUES ON THE MONUSAC2020 CHALLENGE LEADERBOARD (*L*) AND POST-CHALLENGE LEADERBOARD (*PL*)

Rank	Team Name	a-PQ (95% CI)	Pre-Processing		Data Augmentation									Model and Architecture									Loss				Post-Processing					
			Color normalization	Range standization	Rotation	Flip	Affine transform	Scaling	Shear	Elastic transform	Noise addition	Color jitter	Blurring/sharpen	U-Net	FPN	FCN	ResNet	VGG-Net	Hover-Net	EfficientNet	DenseNet	Distance map	Cross entropy	Dice loss	Focal loss	Perception loss	Gaussian smoothing	Watershed segmentation	Morphological operations	Mean of predictions	CRFs	Thresholding
Challenge Leaderboard																																
$L1$	TIA-Lab	0.612 (0.595-0.629)		✓	✓	✓		✓	✓			✓	✓	✓		✓	✓		✓				✓	✓								
$L2$	SJTU-426	0.579 (0.560-0.598)	✓	✓	✓	✓	✓	✓		✓				✓			✓	✓					✓			✓			✓	✓		
$L3$	IVG	0.508 (0.490-0.527)			✓	✓		✓		✓		✓		✓			✓						✓							✓		
$L4$	LSL000UD	0.497 (0.475-0.518)			✓	✓		✓		✓	✓	✓		✓							✓		✓	✓								
$L5$	SharifHooshPardaz	0.481 (0.463-0.499)	✓		✓	✓		✓		✓	✓	✓	✓		✓						✓		✓					✓	✓	✓	✓	
$L6$	xperience.ai	0.449 (0.425-0.473)			✓	✓		✓		✓	✓	✓	✓	✓										✓				✓	✓			
$L7$	TeamTiger	0.426 (0.408-0.444)	✓												✓		✓			✓												
$L8$	Amirreza Mahbod	0.389 (0.375-0.403)		✓		✓		✓				✓		✓								✓		✓	✓		✓	✓	✓			✓
$L9$	DeepBlueAI	0.337 (0.316-0.357)				✓		✓				✓		✓			✓						✓									
$L10$	Debut-Kele	0.263 (0.244-0.282)			✓		✓							✓																		
$L11$	The great backpropagator	0.184 (0.149-0.183)												✓						✓					✓			✓				
$L12$	StevenSmiley	0.166 (0.149-0.183)			✓	✓		✓				✓		✓									✓				✓					✓
$L13$	NUKMLMA	0.149 (0.139-0.160)			✓	✓																								✓		
Post-challenge Leaderboard																																
$PL1$	Amirreza Mahbod	0.561 (0.544-0.577)		✓		✓		✓				✓		✓								✓	✓	✓			✓	✓	✓			✓
$PL2$	IIAI	0.549 (0.529-0.569)			✓	✓	✓	✓	✓							✓	✓		✓			✓	✓									
$PL3$	SharifHooshPardaz	0.536 (0.519-0.553)	✓		✓	✓		✓		✓	✓	✓	✓		✓					✓			✓					✓	✓	✓		
$PL4$	Trnal	0.454 (0.431-0.477)			✓	✓								✓									✓									✓
$PL5$	Cp&ig	0.399 (0.379-0.419)		✓		✓								✓									✓						✓			✓
$PL6$	Onward	0.389 (0.367-0.411)	✓											✓		✓	✓		✓			✓						✓				✓

difficult (e.g., kidney and neutrophils) or easy (e.g., prostate and epithelial nuclei) for all techniques. Nuclei of neutrophils are quite difficult to segment because of high heterogeneity in their shapes and sizes compared to other types of nuclei. This heterogeneity stems from the fact that neutrophils are amorphous and change their shape (from spherical to amoeba-like) once activated and extend pseudopods while hunting for antigens [1], [2].

IV. COMPETING TECHNIQUES AND OBSERVATIONS

In this section we summarize the techniques developed by MoNuSAC2020 participants who appeared on the challenge and post-challenge leaderboards, as shown in Table III. We first describe the key trends in the use of pre-processing, data augmentation, deep learning, and post-processing techniques. We then summarize individual techniques. More details about each technique are available in the supplementary material and under the "manuscripts" tab on the challenge webpage [11].

A. Pre-Processing and Data Augmentation

The performance of deep learning models with relatively fewer parameters can be improved by data normalization that reduces the unwanted variations (e.g., in image contrast) in the input data. On the other hand, models with relatively more parameters benefit from data augmentation that synthetically introduces variations (e.g., color augmentation) to help the model generalize over those variations. Both trends were observed in the pre-processing techniques of the entries submitted to MoNuSAC2020, as shown in Table III.

Color and intensity transformations were the most common data pre-processing techniques seen in the challenge. For example, five teams used color normalization to reduce the color variations, which could have arisen due to the use of different scanners and reagent concentrations in H&E stained tissue images. Out of these, four teams used the Structure Preserving Color Normalization (SPCN) [42]. On the other hand, ten teams used color jitter for data augmentation.

Among geometric data augmentation techniques, rigid transformations of images – such as rotations (especially, by multiples of 90 degrees), horizontal/vertical flips – were the most common. Additionally, twelve teams used image rescaling, and six teams also used elastic transformations.

B. Specification of the Learning Task

Given the images and the target nuclear boundary masks in the training set, the participants formulated the learning problem in one of the following three ways: (1) binary segmentation, (2) ternary segmentation, or (3) distance map estimation. In binary segmentation, the learning task is formulated as the segmentation of foreground nuclear pixels from the non-nuclear background, while the task of separating overlapping nuclei is accomplished using post-processing techniques. An additional class of boundary pixels is used in ternary segmentation to facilitate the separation of touching and overlapping nuclei [22]. Distance map based algorithms estimate how far a nuclear pixel is from the centroid of a nucleus [22]. The learning task is then formulated as a per-pixel regression problem in contrast to the former two formulations. Four teams, including the challenge winner, employed a variation of the distance map regression approach to estimate the horizontal and vertical distance maps separately.

C. Model Architectures and Loss Functions

All participants used deep convolutional neural networks to segment and classify the nuclei. Fourteen teams used variants of U-Net [23] with different base architectures including ResNet [43], FCN [21], Hover-Net [33], and DenseNet [44], while three teams used feature pyramid network (FPN) [45] for semantic segmentation and EfficientNetB7 [46] as an encoder, as shown in Table III.

For the loss function, pixel-wise cross entropy loss, Dice loss [47], or a hybrid of the two losses were the most common ones used for the challenge, as shown in Table III.

D. Post-Processing

Post-processing was used by several teams for the separation of overlapping nuclei. Seven teams used watershed segmentation (WS) for separating touching nuclei in binary segmentation masks. Additionally, a few teams used a neural network to predict the location markers for individual nuclei. Markers were based on either maps of the nucleus centers, or maps of the distance from or the direction towards the nearest nucleus center, as shown in Table III.

Other common uses of post-processing steps were the suppression of small stray or weak detections using Gaussian smoothing or morphological operations, and averaging results from the augmented versions of a given test image.

E. Description of the Techniques

While each technique is described in detail in the supplementary material and in the respective manuscripts on the challenge webpage [11], we briefly describe the key features of the techniques that appear on the challenge and the post-challenge leaderboards. The teams are ordered as per their final ranks on the challenge leaderboard (L1 is the winner) in Table III. Two teams that made it to the challenge leaderboard – Amirreza Mahbod and SharifHooshPardaz – revised their models for the post-challenge submission as well. Visual samples of the segmentation and classification results of the top-five techniques across the two leaderboards are shown in Figure 2.

1) L1, TIA-Lab: Patches of size 256×256 were sampled, normalized in range $[0, 1]$, and augmented using random affine transform, rotation, blur, and color jitter. Horizontal and vertical distance maps of nuclear pixels were generated to separate clustered nuclei. Hover-Net [33] architecture, inspired by U-Net [23], was used as the learning model. Resnet50 [48] was used for feature extraction as an encoder followed by three FCN decoders [21], which were used for (a) binary segmentation, (b) prediction of the distance maps,

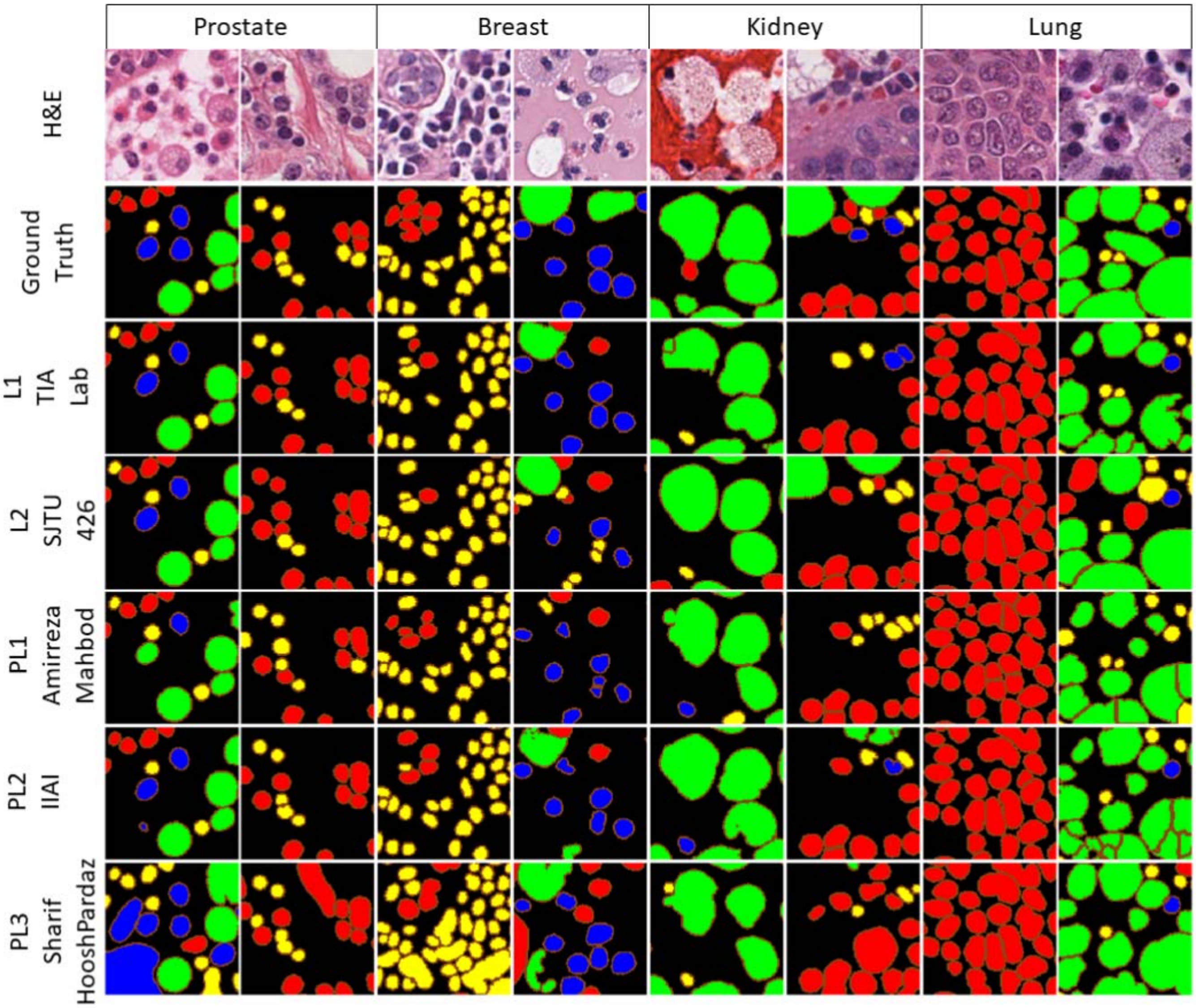

Fig. 2. Test sub-images taken from different organs exemplifying challenges of working with varied nuclear appearances, types, and crowding patterns are shown in columns. Original H&E images, ground truth masks, and segmentation results from the top-five techniques from the challenge and the post-challenge leaderboards are shown in rows. To highlight the separation of touching and overlapping nuclear instances belonging to the same class, the nuclear boundaries are highlighted in brown. Annotations and segmentations of each cell class are shown using matching colors.

and (c) nucleus classification, respectively. A combination of cross entropy, Dice loss, and mean square error (MSE) were used to train the Hover-Net. The initial model weights were obtained by training on PanNuke dataset [9] to improve results over random weight initialization.

2) L2, SJTU 426: Reinhard color normalization [49] followed by an extensive data augmentation using random scaling, horizontal and vertical flips, affine and elastic transformation, random rotations, and z-score normalization of each RGB channel was performed on input images. A two-level hierarchical deep neural network was trained. U-Net [23], with Resnet34 architecture, was used to predict ternary probability maps for pixels inside, on the boundary of, and outside the nuclei, coupled with a combination of weighted cross-entropy and perceptual losses. Individual nucleus instances thus segmented were used to train another U-Net, with VGG16 [50] architecture, to classify the nuclei using cross entropy loss.

3) L3, IVG: Patches of size 256 × 256 were sampled and augmented with RGB shift, color channel shuffle, random color variations, affine and elastic transformations, horizontal and vertical flips, and 90° rotations using the Albumentation library [51]. A U-Net [23] architecture with a Resnet50 [48] encoder and squeeze-and-excitation blocks [52] was trained using a combination of weighted cross-entropy and Lovasz-softmax loss [53]. Test-time augmentation was also used and the final results were obtained through averaging.

4) L4, LSL000UD: Patches of size 256 × 256 were sampled and augmented using random cropping, color jitter, random flips, rotation, scaling, elastic transforms, and an addition of speckle and Gaussian noise. A two stage approach was developed where a U-Net [23] model with DenseNet121 [44] architecture was used for nuclei boundary prediction, followed by three separate U-Nets trained for the classification of various class combinations. Both MoNuSeg [5] and

MoNuSAC2020 [11] data were used for training the boundary prediction stage with binary cross-entropy, while only the MoNuSAC2020 [11] data were used for pixel classification with Dice loss.

5) L5 and PL3, SharifHooshPardaz: Images were color normalized using SPCN [42], and data augmentation was done using random flips, rotation, translation, scaling, elastic transformation, addition of random values to hue and saturation channels, brightness and contrast adjustments, and Gaussian blurring. FPN [45] for semantic segmentation with the EfficientNetB7 architecture [46] was trained to segment six classes: background, four cell types, and the boundary pixels. The model was pre-trained on ImageNet [35] using the weighted cross-entropy loss. Test-time augmentation was also used and the final results were obtained by averaging followed by watershed segmentation. After watershed segmentation, a conditional random field (CRF) [54] was used as another post-processing step. However, in the post-challenge submission, the CRF was removed, which led to an improvement in the accuracy.

6) L6, xperience.ai: Data were augmented using random flips, shifts, scaling, rotations, noise addition, and random color variations. A combination of U-Net and dual path networks (DPN) [55] was trained with Dice loss and the predicted outcomes were post-processed using the watershed segmentation. The submitted model was compared with LinkNet [56], PSPNet [57], SAUNet [58] and the traditional UNet [23], but a combination of U-Net and DPN gave the best results.

7) L7, TeamTiger: Patches of size 256 × 256 were sampled and color normalized using SPCN [42]. Two FPNs [45] were trained, one with an EfficientNetB7 [46] to segment epithelial and lymphocytes, while other using ResNet34 [48] to segment macrophages and neutrophils.

8) L8 and PL1, Amirreza Mahbod: Patches of size 256 × 256 were sampled and augmented using random flips, crops, scaling, and color augmentation through contrast and brightness alteration. Two separate U-Nets were trained for binary segmentation and distance map prediction for nuclei instance segmentation. Gaussian smoothing and watershed algorithms were applied on the average output of the two U-Nets to obtain the segmented nucleus instances, which were classified using a third U-Net.

In their post-challenge submission, they used only one U-Net for nucleus instance segmentation, but the segmentation U-Net was given two decoder units – one for semantic segmentation and another one for distance map prediction. The use of two decoders in the segmentation U-Net led to an improvement in their performance.

9) L9, DeepBlueAI: Data were augmented using random crops, scaling, flips, and color jitter. A hyper column and context attention based U-Net (HCCA-UNet), which combines hypercolumn [59] and context attention [60] modules, based on ResNet-101 backbone was trained for simultaneous segmentation and classification of nuclear instances using a cross entropy loss.

10) L10, Debut Kele: Data were augmented using random crops, affine transformations, and rotations. A U-Net was trained for nucleus instance segmentation. For cell-type classification, an XGBoost [61] model was trained using local statistical features extracted from the segmented instances.

11) L11, The Great Backpropagator: Patches of size 96 × 96 were used to train a U-Net with EfficientNet-B3 architecture [46] to segment and classify nuclei. The model was initialized with weights that were pre-trained on ImageNet [35]. A combination of focal loss [62] and Jaccard index was used to optimize the model. Median filtering and watershed algorithm were used for post-processing the segmented masks.

12) L12, Steven Smiley: Patches were augmented using random rotations, flips, scaling, and color jitter. Four U-Net models were trained – one each for the segmentation of each nucleus type. The final softmax layer of the U-Nets were replaced with sigmoid layers and the models were optimized for the binary cross-entropy loss.

13) L13, NUKMLMA: Data were augmented using random crops, rotations, and flips. A pyramid scene parsing network (PSPNet) [57] was trained for nucleus segmentation and classification. While testing, the images were augmented using 90° rotations and flips. Results were averaged across the test-time augmented outputs.

14) PL2, IIAI: Patches of size 540 × 540 were sampled and augmented using color jitter, Gaussian and median blurring, affine transformations including random rotations, scaling, shear, translation, vertical and horizontal flips, and center cropping. Hover-Net architecture with Preact-ResNet50 encoder [33] was trained with cross-entropy loss on multiple datasets (CPM17 [39], Kumar [22], and CoNSep [33]) before fine-tuning on the MoNuSAC data [11].

15) PL4, Trnal: Patches of size 128 × 128 pixels were sampled and augmented using random rotations and flips. Four U-Nets, one for each cell-type, with exponential linear activation units were trained for segmentation. For a given patch, each pixel was labelled based on the maximum probability across the four class-specific U-Nets. Augmentation and result averaging was performed at test time.

16) PL5, Cp&ig: Pixel intensities were normalized in the range $[0, 1]$ and the data were augmented by horizontal and vertical flips. A novel fully convolutional multi-class distance regularized dense inception network (m-DRDIN) was proposed to segment and classify nuclei instances. m-DRDIN contains one encoder branch with four inception blocks and two decoder branches, each with three inception blocks. The first decoder branch caters to the instance segmentation while the second one gives multi-class labels. The model used the boundary weight map to regularize the network and to learn to identify the boundary pixels. A combination of weighted binary cross-entropy and boundary-aware weight map loss function was used for training.

17) PL6, Onward: Images were color normalized by SPCN [42] and a variant of U-Net, U-HoverNet, with two decoders was trained to segment and classify nuclei. One decoder produced the class-specific segmentation outputs and the other produced the horizontal and vertical distance maps that were used by the watershed algorithm to improve the class-specific segmentation maps obtained from the first decoder.

V. ADDITIONAL EXPERIMENTS

After the challenge, we checked how the top few techniques compared to inter-human agreement, and whether the results can be further improved by ensembling the top few techniques.

A. Comparison With Inter-Human Agreement

We re-annotated the test images and computed their a-PQ with the previous annotations. The re-annotation protocol was identical to the one used for creating the training set of MoNuSAC2020, but was done by a separate student annotator who was blinded to the previous test-set annotations. The a-PQ between new and old manual annotations across the test images was 0.594 (95% CI: 0.580-0.608) to which the a-PQ of the top few techniques compare favorably. This suggests that for multi-class nucleus segmentation in H&E images, machine performance is at par with human performance if the image quality is as good as the one used in this challenge.

B. Ensemble of the Top-Three Techniques

Unlike ensembling of semantic segmentation, where class probabilities or decisions can be averaged for each pixel location, ensembling of multi-class instance segmentation results is not straightforward, and we did not find any usable solution to this problem. We have proposed our own approach for ensembling segmentation results earlier [5], which we modified to generate the ensemble output of instances segmented by the top-three techniques ($L1$, $L2$, and $PL1$ in Table III) as follows. We created the ensemble nuclear masks separately for each cell-type. For example, first we obtained the segmented masks of epithelial nuclei of top three techniques and looped over the epithelial instances to identify the corresponding instances from the other two techniques based on the number of overlapping pixels. An ensemble instance was thus computed through pixel-level majority voting of the three corresponding instances. Once we looped over all epithelial nuclei instances predicted by the rank 1 technique, to incorporate the epithelial nuclei instances missed so far, we looped over all epithelial nuclei instances of rank 2 technique that did not find an overlap with those of the rank 1 technique. We used the aforementioned ensemble method separately for other cell-types including lymphocytes, neutrophils, and macrophages. Finally, we aggregated the ensemble masks of individual cell-types to obtain the multi-class nuclei instance masks. This ensembling method gave an overall a-PQ of 0.620 (95% CI: 0.607-0.635), which is only marginally better than the individual results of the top-three teams.

VI. CONCLUSION AND DISCUSSION

In this paper, we described the dataset and the organization of the MoNuSAC2020 challenge [11] for instance segmentation and classification of nuclei in H&E-stained cancer tissue images. The major trends observed in the development of techniques for this problem can be summarized as follows. Firstly, fully convolutional architectures inspired by the U-Net [23] and FCN [21] seem to be useful for segmentation due to their efficient use of convolutional parallelism for not only the input but also the output. Secondly, heavy data augmentation was also a common step used by several teams as it helps make the training data diverse enough to model the distribution of the testing data as well. Thirdly, an appropriate combination of the semantic loss (e.g. Dice loss) and instance label loss (e.g. cross-entropy) worked best for training neural networks for nucleus segmentation and classification task. Finally, a few participants also developed novel and unique neural architectures, such as m-DRDIN (team Cp&ig).

Overall, the results were quite satisfactory, both qualitatively (as shown in Figure 2) and quantitatively (as shown in Table III and Section V-A). Some difficult examples of macrophages and neutrophils, which were under-represented in the training set, were also accurately segmented and labeled by the top participants. Additionally, their performance matched that of inter-human agreement. Ensembling the top three techniques did not improve the results significantly, perhaps because of the limitations of human subjectivity in labeling and annotating nuclei in particular, and images in general. For instance, the rotation angle of a neutrophil may hide its characteristic tell-tale multi-lobular structure. Additionally, in H&E stained tissue images, the nuclear boundaries are not easily visible below the cytoplasm of another cell for overlapping nuclei. Finally, the annotation strokes can also be imprecise, and therefore, some level of noise in manual annotations is to be expected. Hence, to handle the manual labeling noise, robust machine learning techniques should be developed [63], [64].

While improvements in nucleus segmentation and classification will indeed come in the next few years, we hope that the MoNuSAC2020 data released with this paper [11] will accelerate the development of such algorithms. Additionally, by using the techniques developed by the challenge participants, researchers will be able to plan studies to compare the tumor microenvironments of various cohorts of cancer patients. For instance, quantitative investigations into the intensity and proximity of immune response and their association with disease progression can now be studied, as can be the role of neutrophils and macrophages in tumor progression [1]–[3]. We also hope that the challenge serves as a template for others to release carefully annotated and curated datasets for building algorithms for medical image assessment to support the future downstream analysis such as understanding disease pathobiology, prognostication, and treatment planning.

ACKNOWLEDGMENT

Ruchika Verma, Neeraj Kumar, Abhijeet Patil, Nikhil Cherian Kurian, Swapnil Rane, and Amit Sethi co-organized the challenge; all others contributed their algorithms and results. The organizers and participants of MoNuSAC2020 thank NVIDIA for providing GPU compute credits to top 10 participants and a GPU to the winner.

Ruchika Verma is with the Department of Biomedical Engineering, Case Western Reserve University, Cleveland, OH 44106 USA (e-mail: verma.ece.ruchika@gmail.com).

Neeraj Kumar is with the Department of Computing Science, University of Alberta, Edmonton, AB T6G 2R3, Canada, and also with the Alberta Machine Intelligence Institute, Edmonton, T5J 3B1, Canada (e-mail: neeraj.kumar.iitg@gmail.com).

Abhijeet Patil, Nikhil Cherian Kurian, and Amit Sethi are with the Department of Electrical Engineering, IIT Bombay, Mumbai 400076, India (e-mail: abhijeetptl5@gmail.com; nikhilcherian30@gmail.com).

Swapnil Rane is with the Department of Pathology, Tata Memorial Centre-ACTREC, HBNI, Mumbai 400012, India (e-mail: raneswapnil82@gmail.com).

Simon Graham, Quoc Dang Vu, Shan E. Ahmed Raza, and Nasir Rajpoot are with the Department of Computer Science, University of Warwick, Coventry CV4 7AL, U.K. (e-mail: S.Graham.1@warwick.ac.uk; dangvuquoc1993@gmail.com; shan.raza@warwick.ac.uk; n.m.rajpoot@warwick.ac.uk).

Mieke Zwager is with the Department of Pathology, University of Groningen, 9712 Groningen, The Netherlands (e-mail: mieke.zwager@warwick.ac.uk).

Xiyi Wu, Huai Chen, Yijie Huang, and Lisheng Wang are with the Department of Automation, Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: wuxiyi@sjtu.edu.cn; libaozhang@bnu.edu.cn).

Hyun Jung, G. Thomas Brown, and Yanling Liu are with the Advanced Biomedical Computing Science, Frederick National Laboratory for Cancer Research, National Cancer Institute, Frederick, MD 21702 USA.

Shuolin Liu is with the Department of Electrical Engineering and Automation, Anhui University, Hefei 230039, China.

Seyed Alireza Fatemi Jahromi, Ali Asghar Khani, Ehsan Montahaei, and Mahdieh Soleymani Baghshah are with the Department of Computer Engineering, Sharif University of Technology, Tehran 11365-8639, Iran.

Hamid Behroozi is with the Department of Electrical Engineering, Sharif University of Technology, Tehran 11365-8639, Iran.

Pavel Semkin and Alexandr Rassadin are with Xperience AI, 603155 Nizhny Novgorod, Russia.

Prasad Dutande, Romil Lodaya, Ujjwal Baid, Bhakti Baheti, and Sanjay Talbar are with the Center of Excellence in Signal and Image Processing, SGGS Institute of Engineering and Technology, Nanded 431606, India.

Amirreza Mahbod is with the Institute for Pathophysiology and Allergy Research, Medical University of Vienna, 1090 Vienna, Austria, and also with the Department of Research and Development, TissueGnostics GmbH, 1020 Vienna, Austria.

Rupert Ecker is with the Department of Research and Development, TissueGnostics GmbH, 1020 Vienna, Austria.

Isabella Ellinger is with the Institute for Pathophysiology and Allergy Research, Medical University of Vienna, 1090 Vienna, Austria.

Zhipeng Luo, Bin Dong, Zhengyu Xu, and Yuehan Yao are with DeepBlue Technology (Shanghai) Company, Ltd., Shanghai 200336, China.

Shuai Lv is with the School of Automation, Guangdong University of Technology, Guangzhou 510006, China.

Ming Feng is with the School of Electronic and Information Engineering, Tongji University, Shanghai 200092, China.

Kele Xu is with the National Key Laboratory of Parallel and Distributed Processing, National University of Defense Technology, Changsha 410073, China.

Hasib Zunair and Abdessamad Ben Hamza are with the Concordia Institute for Information Systems Engineering, Concordia University, Montreal, H4B 1R6, Canada.

Steven Smiley is with Site Risk Management Engineer, Jensen Hughes, Liverpool, NY 13212 USA.

Tang-Kai Yin and Qi-Rui Fang are with the Department of Computer Science and Information Engineering, National University of Kaohsiung, Kaohsiung 811, Taiwan.

Shikhar Srivastava and Dwarikanath Mahapatra are with the Inception Institute of Artificial Intelligence, Abu Dhabi, United Arab Emirates.

Lubomira Trnavska is with the Faculty of Informatics and Information Technology, Slovak University of Technology, 811 07 Bratislava, Slovakia.

Hanyun Zhang, Priya Lakshmi Narayanan, Justin Law, and Yinyin Yuan are with the Centre for Evolution and Cancer, Institute of Cancer Research, London SW7 3RP, U.K., and also with the Division of Molecular Pathology, Institute of Cancer Research, London SW7 3RP, U.K.

Abhiroop Tejomay, Aditya Mitkari, Dinesh Koka, Vikas Ramachandra, and Lata Kini are with Onward Health, Hyderabad 500032, India.

REFERENCES

[1] W. H. Fridman, L. Zitvogel, C. Sautès–Fridman, and G. Kroemer, "The immune contexture in cancer prognosis and treatment," *Nature Rev. Clin. Oncol.*, vol. 14, no. 12, pp. 717–734, Dec. 2017.

[2] J. J. Havel, D. Chowell, and T. A. Chan, "The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy," *Nature Rev. Cancer*, vol. 19, no. 3, pp. 133–150, Mar. 2019.

[3] B. Ruffell and L. M. Coussens, "Macrophages and therapeutic resistance in cancer," *Cancer Cell*, vol. 27, no. 4, pp. 462–472, Apr. 2015.

[4] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, and A. Sethi, "A dataset and a technique for generalized nuclear segmentation for computational pathology," *IEEE Trans. Med. Imag.*, vol. 36, no. 7, pp. 1550–1560, Jul. 2017.

[5] N. Kumar *et al.*, "A multi-organ nucleus segmentation challenge," *IEEE Trans. Med. Imag.*, vol. 39, no. 5, pp. 1380–1391, May 2020.

[6] H. Irshad *et al.*, "Crowdsourcing image annotation for nucleus detection and segmentation in computational pathology: Evaluating experts, automated methods, and the crowd," in *Proc. Symp. Biocomput.*, Nov. 2014, pp. 294–305.

[7] S. Wienert *et al.*, "Detection and segmentation of cell nuclei in virtual microscopy images: A minimum-model approach," *Sci. Rep.*, vol. 2, no. 1, p. 503, Dec. 2012.

[8] P. Naylor, M. Lae, F. Reyal, and T. Walter, "Segmentation of nuclei in histopathology images by deep regression of the distance map," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 448–459, Feb. 2019.

[9] J. Gamper *et al.*, "PanNuke dataset extension, insights and baselines," 2020, *arXiv:2003.10778*. [Online]. Available: https://arxiv.org/abs/2003.10778

[10] *The Cancer Genome Atlas (TCGA)*. Accessed: Aug. 1, 2019. [Online]. Available: http://cancergenome.nih.gov/

[11] *Multi-Organ Nuclei Segmentation and Classification Challenge (MoNuSAC) 2020*. Accessed: Aug. 11, 2020. [Online]. Available: https://monusac-2020.grand-challenge.org/

[12] X. Yang, H. Li, and X. Zhou, "Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and Kalman filter in time-lapse microscopy," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 53, no. 11, pp. 2405–2414, Nov. 2006.

[13] S. Ali and A. Madabhushi, "An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery," *IEEE Trans. Med. Imag.*, vol. 31, no. 7, pp. 1448–1460, Jul. 2012.

[14] M. Veta, P. J. van Diest, R. Kornegoor, A. Huisman, M. A. Viergever, and J. P. W. Pluim, "Automatic nuclei segmentation in H&E stained breast cancer histopathology images," *PLoS ONE*, vol. 8, no. 7, Jul. 2013, Art. no. e70221.

[15] Y. Al-Kofahi, W. Lassoued, W. Lee, and B. Roysam, "Improved automatic detection and segmentation of cell nuclei in histopathology images," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 4, pp. 841–852, Apr. 2010.

[16] H. Kong, M. Gurcan, and K. Belkacem-Boussaid, "Partitioning histopathological images: An integrated framework for supervised color-texture segmentation and cell splitting," *IEEE Trans. Med. Imag.*, vol. 30, no. 9, pp. 1661–1677, Sep. 2011.

[17] H. Chang *et al.*, "Invariant delineation of nuclear architecture in glioblastoma multiforme for clinical and molecular association," *IEEE Trans. Med. Imag.*, vol. 32, no. 4, pp. 670–682, Apr. 2013.

[18] M. Zhang, T. Wu, and K. M. Bennett, "Small blob identification in medical images using regional features from optimum scale," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 4, pp. 1051–1062, Apr. 2015.

[19] H. Irshad, A. Veillard, L. Roux, and D. Racoceanu, "Methods for nuclei detection, segmentation, and classification in digital histopathology: A review—Current status and future potential," *IEEE Rev. Biomed. Eng.*, vol. 7, pp. 97–114, 2014.

[20] E. Moen, D. Bannon, T. Kudo, W. Graf, M. Covert, and D. Van Valen, "Deep learning for cellular image analysis," *Nature Methods*, vol. 16, no. 12, pp. 1233–1246, 2019.

[21] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[22] N. Kumar *et al.*, "Convolutional neural networks for prostate cancer recurrence prediction," *Proc. SPIE*, vol. 10140, Mar. 2017, Art. no. 101400H.

[23] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput.-Assisted Intervent.* Munich, Germany: Springer, 2015, pp. 234–241.

[24] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Granada, Spain: Springer, 2018, pp. 3–11.

[25] X. Xia and B. Kulis, "W-Net: A deep model for fully unsupervised image segmentation," 2017, *arXiv:1711.08506*. [Online]. Available: https://arxiv.org/abs/1711.08506

[26] H. Chen, X. Qi, L. Yu, Q. Dou, J. Qin, and P.-A. Heng, "DCAN: Deep contour-aware networks for object instance segmentation from histology images," *Med. Image Anal.*, vol. 36, pp. 135–146, Feb. 2017.

[27] J. Song, L. Xiao, and Z. Lian, "Contour-seed pairs learning-based framework for simultaneously detecting and segmenting various overlapping cells/nuclei in microscopy images," *IEEE Trans. Image Process.*, vol. 27, no. 12, pp. 5759–5774, Dec. 2018.

[28] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2961–2969.

[29] A. O. Vuola, S. U. Akram, and J. Kannala, "Mask-RCNN and U-Net ensembled for nuclei segmentation," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 208–212.

[30] J. W. Johnson, "Automatic nucleus segmentation with mask-RCNN," in *Proc. Sci. Inf. Conf.* Las Vegas, NV, USA: Springer, 2019, pp. 399–407.

[31] Y. Liu, P. Zhang, Q. Song, A. Li, P. Zhang, and Z. Gui, "Automatic segmentation of cervical nuclei based on deep learning and a conditional random field," *IEEE Access*, vol. 6, pp. 53709–53721, 2018.

[32] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. J. Snead, I. A. Cree, and N. M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1196–1206, May 2016.

[33] S. Graham *et al.*, "Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images," *Med. Image Anal.*, vol. 58, Dec. 2019, Art. no. 101563.

[34] S. Wang *et al.*, "Computational staining of pathology images to study the tumor microenvironment in lung cancer," *Cancer Res.*, vol. 80, no. 10, pp. 2056–2066, May 2020.

[35] O. Russakovsky *et al.*, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[36] M. S. Hosseini *et al.*, "Atlas of digital pathology: A generalized hierarchical histological tissue type-annotated database for deep learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11747–11756.

[37] E. D. Gelasca, B. Obara, D. Fedorov, K. Kvilekval, and B. Manjunath, "A biosegmentation benchmark for evaluation of bioimage analysis methods," *BMC Bioinf.*, vol. 10, no. 1, p. 368, Dec. 2009.

[38] P. Kainz, M. Urschler, S. Schulter, P. Wohlhart, and V. Lepetit, "You should use regression to detect cells," in *Proc. Int. Conf. Med. Image Comput.-Assisted Intervent.* Munich, Germany: Springer, 2015, pp. 276–283.

[39] Q. D. Vu *et al.*, "Methods for segmentation and classification of digital microscopy tissue images," *Frontiers Bioeng. Biotechnol.*, vol. 7, p. 53, Apr. 2019.

[40] A. Janowczyk and A. Madabhushi, "Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases," *J. Pathol. Informat.*, vol. 7, no. 1, p. 29, 2016.

[41] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollar, "Panoptic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9404–9413.

[42] A. Vahadane *et al.*, "Structure-preserving color normalization and sparse stain separation for histological images," *IEEE Trans. Med. Imag.*, vol. 35, no. 8, pp. 1962–1971, Aug. 2016.

[43] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *Proc. 31st AAAI Conf. Artif. Intell.*, Feb. 2017, pp. 1–7.

[44] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.

[45] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2117–2125.

[46] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," 2019, *arXiv:1905.11946*. [Online]. Available: https://arxiv.org/abs/1905.11946

[47] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, Jul. 1945.

[48] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[49] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Comput. Graph. Appl.*, vol. 21, no. 4, pp. 34–41, Jul. 2001.

[50] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: https://arxiv.org/abs/1409.1556

[51] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: Fast and flexible image augmentations," *Information*, vol. 11, no. 2, p. 125, Feb. 2020.

[52] J. Hu, L. Shen, and G. Sun, "Squeeze- and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[53] M. Berman, A. R. Triki, and M. B. Blaschko, "The Lovász-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4413–4421.

[54] Y. Zhang and T. Chen, "Efficient inference for fully-connected CRFs with stationarity," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 109–117.

[55] Y. Chen, J. Li, H. Xiao, X. Jin, S. Yan, and J. Feng, "Dual path networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 4467–4475.

[56] A. Chaurasia and E. Culurciello, "LinkNet: Exploiting encoder representations for efficient semantic segmentation," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Dec. 2017, pp. 1–4.

[57] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, 2017, pp. 2881–2890.

[58] J. Sun, F. Darbehani, M. Zaidi, and B. Wang, "SAUNet: Shape attentive U-Net for interpretable medical image segmentation," 2020, *arXiv:2001.07645*. [Online]. Available: https://arxiv.org/abs/2001.07645

[59] B. Hariharan, P. Arbelaez, R. Girshick, and J. Malik, "Hypercolumns for object segmentation and fine-grained localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 447–456.

[60] Y. Yuan, L. Huang, J. Guo, C. Zhang, X. Chen, and J. Wang, "OCNet: Object context network for scene parsing," 2018, *arXiv:1809.00916*. [Online]. Available: https://arxiv.org/abs/1809.00916

[61] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794.

[62] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

[63] G. Patrini, A. Rozza, A. K. Menon, R. Nock, and L. Qu, "Making deep neural networks robust to label noise: A loss correction approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1944–1952.

[64] B. Han *et al.*, "Co-teaching: Robust training of deep neural networks with extremely noisy labels," 2018, *arXiv:1804.06872*. [Online]. Available: https://arxiv.org/abs/1804.06872