

DeepCIS: An end-to-end Pipeline for Cell-type aware Instance Segmentation in Microscopic Images

Nabeel Khalid*, Mohsin Munir*, Christoffer Edlund†, Timothy R Jackson‡,
Johan Trygg§, Rickard Sjögren‡§, Andreas Dengel*¶, Sheraz Ahmed*

*German Research Center for Artificial Intelligence (DFKI) GmbH, Kaiserslautern 67663, Germany

firstname.lastname@dfki.de

†Sartorius Corporate Research, Sweden

‡Sartorius, BioAnalytics, Royston, United Kingdom

firstname.lastname@Sartorius.com

§Computational Life Science Cluster (CLiC), Umeå University, Sweden

¶Technische Universität Kaiserslautern, Kaiserslautern 67663, Germany

Abstract—Accurate cell segmentation in microscopic images is a useful tool to analyze individual cell behavior, which helps to diagnose human diseases and development of new treatments. Cell segmentation of individual cells in a microscopic image with many cells in view allows quantification of single cellular features, such as shape or movement patterns, providing rich insight into cellular heterogeneity. Most of the cell segmentation algorithms up till now focus on segmenting cells in the images without classifying the culture of the cell in the images. Discrimination among cell types in microscopic images can lead to a new era of high-throughput cell microscopy. Multiple cell types in co-culture can be easily identified and studying the changes in cell morphology can lead to many applications such as drug treatment. To address this gap, DeepCIS is proposed to detect, segment, and classify the culture of the cells and nucleus in the microscopic images. We have used the EVICAN60 dataset which contains microscopic images from a variety of microscopes having numerous cell cultures, to evaluate the proposed pipeline. To further demonstrate the utility of the DeepCIS, we have designed various experimental settings to uncover its learning potential. We have achieved a mean average precision score of 24.37% for the segmentation task averaged over 30 classes for cell and nucleus.

Index Terms—biomedical, healthcare, deep learning, cell-type segmentation, nucleus-type segmentation, cell-type classification

I. INTRODUCTION

Segmenting individual cells enables detailed studies of individual cell behavior, for instance, cell migration that is a central process behind many physiological and pathological processes like tissue maintenance, development, and healing. Microscopic images suffer from certain challenges including low contrast as well as irregularly shaped and overlapping cells as compared to natural images, making the segmentation process challenging. Convolutional neural network-based approaches are showing promising results in microscopic image analysis [1]. Even with the great advancements in the field of deep learning, there are no methods that classify the culture of each cell in the image in addition to the task of detection and segmentation. Chan et al. [2] proposed a deep learning

algorithm for cell segmentation. Ali et al. [3] proposed a deep learning-based algorithm for nucleus segmentation in brightfield cell microscopy images. These approaches don't differentiate between cell and nucleus types. Discrimination among cell types can lead to distinguishing cells in a co-culture environment. This can lead to studying morphological changes among different cell cultures which can, in turn, lead to drug treatment. Unlike other cell segmentation dataset which doesn't differentiate between different cell cultures, EVICAN60 [4] dataset differentiates cells among 30 different cell cultures.

In this study, we have proposed parameters for features extraction and anchor sizes to detect, segment, and classify cells and nucleus on the EVICAN60 [4] dataset. We propose a novel pipeline DeepCIS – an end-to-end pipeline for Culture Dependent Cell and Nucleus Segmentation. The main contribution of this study is as follows:

- An end-to-end cell and nucleus segmentation and culture classification pipeline based on Cascade Mask R-CNN [5] with ResNeSt [6] backbone.
- Extensive evaluation of proposed method on EVICAN60 [4] dataset from different perspectives.

II. METHODOLOGY

Fig. 1 provides a system overview diagram of the proposed pipeline. The proposed pipeline is divided into three blocks.

A. Backbone Network

The purpose of the backbone network is to extract feature maps from the input image at different scales. The backbone network of the proposed methodology is composed of Feature Pyramid Network (FPN) [7] and ResNeSt-200 [6]. FPN operates on the bottom-up pathway, top-down pathway, and lateral connection to combine the low resolution, semantically strong features with high resolution semantically weak features. The bottom-up pathway uses a normal feed-forward CNN architecture which computes the hierarchy of features consisting of feature maps at various scales. As we move up the convolution

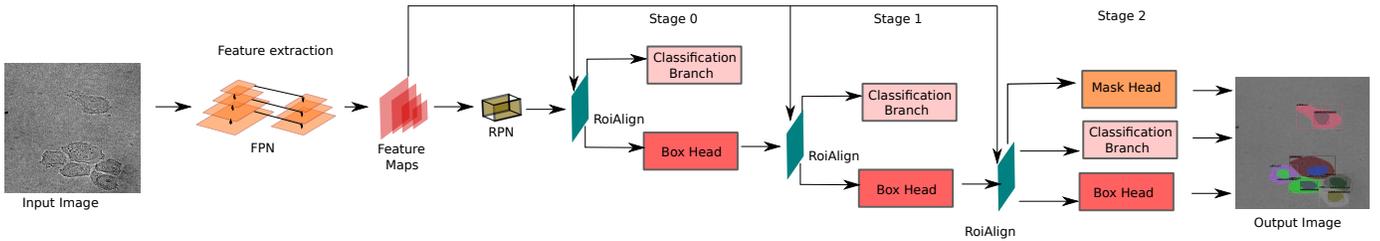


Fig. 1: System overview diagram of DeepCIS. Input image is passed to the proposed pipeline and the output image with detection and segmentation is produced.

layers, more high-level structures are detected. ResNeSt-200 [6] is used with deformable convolution [8] as a feed-forward CNN architecture in the bottom-up pathway of our approach. ResNeSt is composed of a modular split attention block that enables attention across feature map groups which helps different network branches to capture cross-feature interactions and learn diverse representations. The blocks are then stacked according to ResNet-style. Deformable convolution is being used instead of conventional convolution because it suits best for our application in terms of cell deformity [9]. The output of each convolution layer of ResNeSt is used in the top-down pathway which constructs a higher resolution layer from the semantic rich layer.

B. Region Proposal Network

After the extraction of multi-scale features from the backbone network, these features are further passed onto a Regional Proposal Network (RPN) [10]. The purpose of RPN is to detect regions that contain objects and match them to the groundtruth. This process is performed by generating anchor boxes on the input image. The generated anchors are then matched to the groundtruth by taking Intersection over Union (IoU) between anchors and ground truth. If IoU is larger than the defined threshold, i.e. 0.7, the anchor is linked to one of the ground truth boxes and assigned to the foreground. If the IoU is greater than 0.3, it is marked as background and ignored otherwise.

C. Prediction Head

After the successful generation of proposals, the next block in our pipeline is the prediction head. At the prediction head, we have groundtruth boxes, proposal boxes from RPN, and feature maps from FPN. The job of the prediction head is to predict the class, bounding box, and binary mask for each region of interest. Cascade Mask R-CNN [5] is used as the prediction head, which is an extension of Cascade R-CNN by adding a mask branch to the cascade. The output of one stage of cascade is used as input for training the next stage. In our methodology, we are adding the segmentation branch at the last stage of the Cascade R-CNN.

III. DATASET

We have used EVICAN60 [4] dataset in this study. The EVICAN60 dataset consists of 60 classes (30 classes for cell and nucleus each). A total of 52,959 instances are present in

the 4,640 partially annotated images in the training and validation dataset. The test set contains 1,057 instances in 98 fully annotated images. Based on the image quality characteristics, the test set is divided into three difficulty levels as defined in [4]. A total of 374, 356, and 327 instances are presented in 33, 33, and 32 images for difficulty levels 1, 2, and 3 respectively.

IV. EXPERIMENTAL SETUPS

Three different experimental settings have been designed and explained in this section to analyze the performance of the proposed pipeline. These experimentations manifest the impact and limitations of culture-dependent cell and nucleus detection and segmentation when treated individually and collectively. In the first two settings namely Culture-dependent cell segmentation and Culture-dependent nucleus segmentation, cell and nucleus detection and segmentation are performed separately, whereas, in the third experimental setting, namely cell Culture-dependent cell and nucleus segmentation, cell and nucleus detection and segmentation are carried out collectively. We leveraged transfer learning to train DeepCIS by using MS-COCO pre-trained model [7] for all the settings. Training for all the experiments is performed with a base learning rate of 0.02 and momentum of 0.9. The anchor sizes for each setting were set after careful consideration of the cells and nucleus pixel area in the images. Anchors aspect ratios were set to 0.25, 0.5, 1, 2, 4 for all the settings.

Evaluation Metrics: We are following the standard COCO evaluation protocol [7] to evaluate the performance of the proposed pipeline. Mean average precision (mAP) is reported at different IoU thresholds i.e., mAP50 and mAP75, and on three different area ranges i.e. mAPs (small), mAPm (medium), and mAPl (large).

A. Experimental Setting 1: Culture-dependent cell segmentation

In this experimental setting, the objective is to detect and segment only cells for the 30 different cell cultures in the EVICAN60 [4] dataset. There are 21,106 cell instances in the training dataset, 5,317 cell instances in the validation dataset, and 525 cell instances in the test set. In the training dataset, the most number of images are present in the cell type C2C12 i.e., 264, and the most number of labeled instances for cell culture HT29 i.e., 1,308, and only 30 cell instances are present for cell cultures SK-BR-3 and Colo. For the validation set, there are

no labeled instances for cell cultures HeL299 and HCT116. In the evaluation dataset, there are no images present for the cell cultures Caco-2, hMSC, CAKI-2, MCC26, Colo, DLD-1, and HT1080 for difficulty level 1. For difficulty level 2, there are no images for cultures HeLa, 769p, RKO, T47D, PC-3, SH-SY5Y, and SW-480. Cell cultures DU-145, MDA MB, CHO, 786-O, Colo, MCF-7, NIH-3T3, and SK-BR-3 don't have any images in the difficulty level 3 category of the evaluation data set. The anchor sizes for this setting were set to 2, 6, 17, 31, 64, 127, 256, 512, 1024 pixels after analyzing the histogram of cell pixel areas in the training set. The checkpoint at 1,500 iterations was chosen for evaluation. Table I shows the overall evaluation results for this setting. The mAP here is the mean of AP over all 30 classes of cells. The mAP for difficulty level 2 is nearly 0% for both detection and segmentation tasks for this setting. For difficulty level 1, the mAP is around 30% for both detection and segmentation tasks. For difficulty level 3, the mAP is around 10% for both detection and segmentation tasks.

B. Experimental Setting 2: Culture-dependent nucleus segmentation

In this experimental setting, the objective is to detect and segment the nucleus only for the 30 different cell cultures in the EVICAN60 [4] dataset. There are 21,211 nucleus instances in the training dataset, 5,325 nucleus instances in the validation dataset, and 525 nucleus instances in the test set. The evaluation set doesn't have any images for 8, 7, and 9 classes for difficulty levels 1, 2, and 3 respectively. The anchor sizes of 2, 6, 8, 12, 17, 24, 31, 64, 127, 256 pixels were chosen for this experiment. The checkpoint at 2,000 iterations was chosen for evaluation. The overall evaluation scores averaged over 30 classes of the nucleus are given in Table I. The mAP for the nucleus is very low as compared to that for cells.

C. Experimental Setting 3: Culture-dependent cell and nucleus segmentation

In this experimental setting, 30 classes per cell and 30 classes per nucleus are present, which correspond to a total of 60 classes. The anchor sizes used in experimental setting 1 suit well for this setting. The checkpoint at 3,500 iterations was chosen for the evaluation. Table I shows the mean AP scores for all 60 classes. For difficulty level 1, an mAP of 24.37% for both detection and segmentation tasks is achieved which is almost 3 times the mAP we get for difficulty level 3.

V. ANALYSIS AND DISCUSSION

In experimental setting 1, segmentation mAP of 30.31% is achieved for difficulty level 1. The per-class segmentation AP scores for difficulty level 1 indicate that the highest score of 83.90% is recorded against the cell culture SW-480. The mAP scores drop with the decline in image quality characteristics, such as increased cell-cell contact and invisible cell outlines, as mentioned in [4] for the difficulty levels 2 and 3. The segmentation score for nucleus in experimental setting 2 is

nearly half of the cell segmentation score in experimental setting 1. Similar to experimental setup 1, we get an AP score of 0% for difficulty level 2. For experimental setting 3, an mAP of 24.37% is achieved for both detection and segmentation averaged over all the 60 classes. It was observed that the per class score for cell and nucleus culture segmentation increases for the experimental setting 3 where all the 30 cell and nucleus classes are trained collectively which is due to more training samples per class. The main reason behind these low scores is the number of images and the number of instances per class in all the difficulty levels. Difficulty level 1 doesn't contain any image or instances for 8 cell cultures, 11 classes with only 1 image, and the rest of 11 classes with just 2 images. The same trend is seen across the other two difficulty levels too.

Fig. 2 shows the inference results on some samples where the proposed pipeline performed adequately and inadequately. The AP50 on top of every sub-image represents the segmentation mAP score at IoU threshold of 0.5. The predictions with detection scores above 0.5 are shown in the inference results. The difficulty level of each sample is represented by the glow of green, blue and red for difficulty levels 1, 2, and 3 respectively.

For the culture-dependent cell segmentation in the adequate column, we get an AP50 score of 100% with the correct cell culture predicted i.e., SW480. For the nucleus segmentation experiment sample in the adequate column, we get an AP50 score of 100% with most of the nucleus instances correctly predicted. A similar trend can be seen for the last sample in the adequate column for the third experiment. All the cell and nucleus instances are correctly segmented and classified as belonging to PC-3 cell culture.

In the inadequate column results in Fig 2, our model failed to predict 3 cell instances for the culture-dependent cell segmentation. Similarly, for the second experiment, the model fails to predict one nucleus instance but the rest of the instances are correctly segmented with the correct cell culture. In the last experimental setting, our model is successful in predicting just two cell instances, with no nucleus instances detected. The class of these two cell instances is predicted to be HeL299, whereas in actual they belong to the cell culture FADU, hence the AP50 score of 0%.

VI. CONCLUSION

In this study, we have proposed a novel pipeline for cell and nucleus detection and segmentation alongside the identification of the class for each cell and nucleus instance. The average precision scores are still very low which is mostly because the EVICAN60 dataset used for training is partially annotated, where all the cells and nucleus instances are not labeled. In addition to that, the distribution of the images in the evaluation set doesn't provide the best platform to judge the proposed pipeline. Still, the proposed pipeline performs relatively well for detecting and segmenting cell and nucleus instances. We believe that with a more balanced dataset, the results can be improved further.

TABLE I: Detection and segmentation results for the 3 experimental settings

Difficulty	Culture-Dependent Cell Segmentation											
	mAP		mAP50		mAP75		mAPs		mAPm		mAPI	
1	29.83	30.31	41.09	41.09	33.51	33.38	15.15	15.15	27.89	27.91	36.34	37.17
2	0.04	0.07	0.18	0.18	0.00	0.00	0.27	0.12	0.08	0.12	0.00	0.00
3	9.34	9.67	16.70	17.39	8.48	9.80	0.00	0.00	12.39	11.23	17.63	18.09
Culture-Dependent Nucleus Segmentation												
1	16.49	16.12	27.11	27.02	20.12	18.31	11.58	10.84	26.98	28.69	0.00	0.00
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	4.29	3.69	2.97	8.66	2.97	8.66	2.75	2.15	9.96	9.67	10.68	11.96
Culture-Dependent Cell and Nucleus Segmentation												
1	24.37	24.37	37.56	37.37	25.63	26.50	11.99	11.36	30.66	30.37	33.79	35.31
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	8.97	8.11	14.68	14.61	9.64	8.47	3.60	3.64	10.51	8.96	24.07	21.57

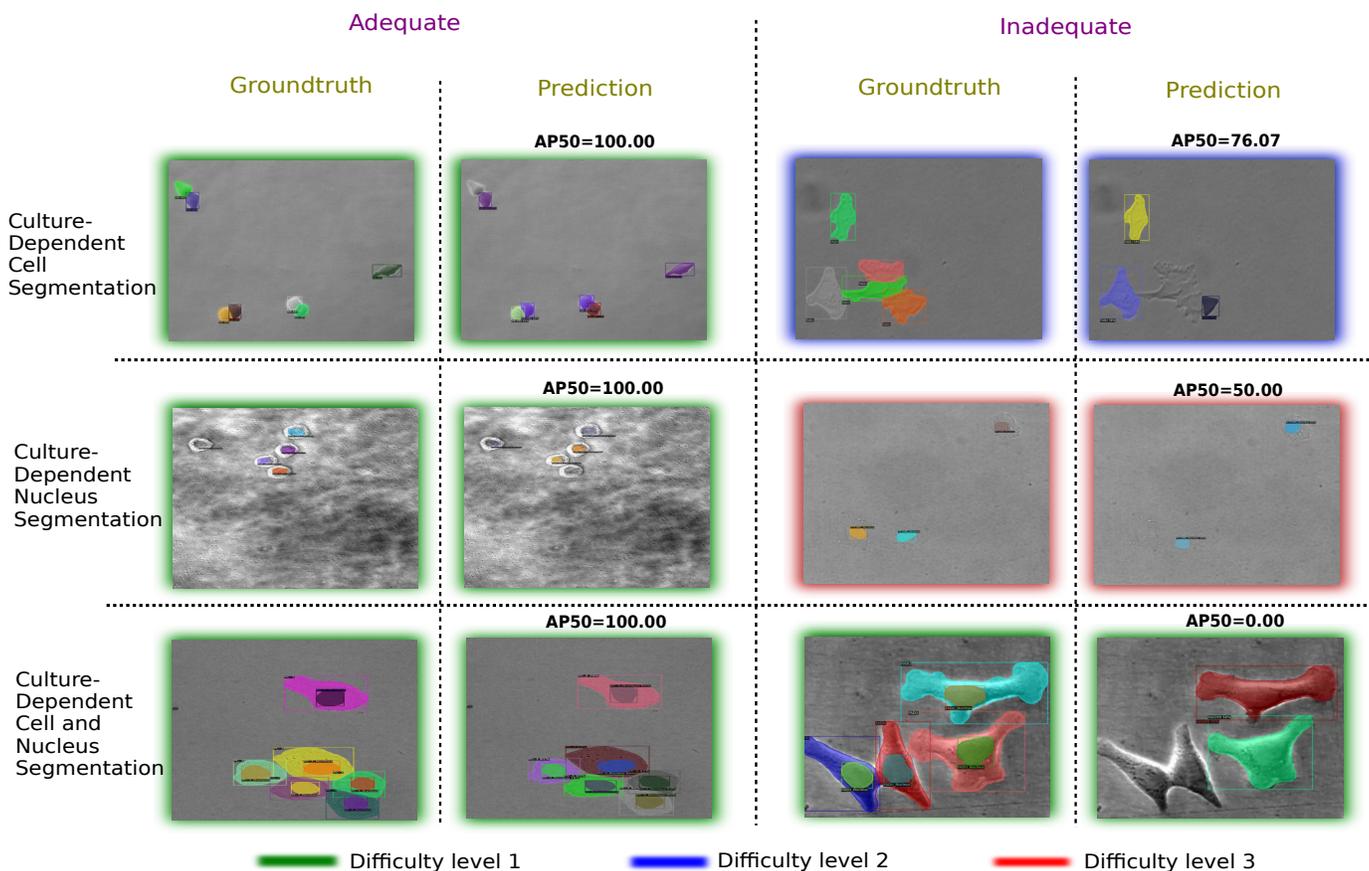


Fig. 2: Inference results of some samples where DeepCIS performed adequately and inadequately. The glow around the image represents to which difficulty level the test image belongs to (green, blue and red for difficulty level 1, 2 and 3 respectively).

REFERENCES

- [1] E. Moen, D. Bannon, T. Kudo, W. Graf, M. Covert, and D. Van Valen, "Deep learning for cellular image analysis," *Nature methods*, 2019.
- [2] S. Chan, C. Huang, C. Bai, W. Ding, and S. Chen, "Res2-unext: a novel deep learning framework for few-shot cell image segmentation," *Multimedia Tools and Applications*, pp. 1–14, 2021.
- [3] M. A. Ali, O. Misko, S.-O. Salumaa, M. Papkov, K. Palo, D. Fishman, and L. Parts, "Evaluating very deep convolutional neural networks for nucleus segmentation from brightfield cell microscopy images," *SLAS DISCOVERY: Advancing the Science of Drug Discovery*, p. 24725552211023214, 2021.
- [4] M. Schwendy, R. E. Unger, and S. H. Parekh, "Evcin—a balanced dataset for algorithm development in cell and nucleus segmentation," *Bioinformatics*, 2020.
- [5] Z. Cai and N. Vasconcelos, "Cascade r-cnn: Delving into high quality object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018.
- [6] H. Zhang, C. Wu, Z. Zhang, Y. Zhu, Z. Zhang, H. Lin, Y. Sun, T. He, J. Mueller, R. Manmatha *et al.*, "Resnest: Split-attention networks," *arXiv preprint arXiv:2004.08955*, 2020.
- [7] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014.
- [8] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, "Deformable convolutional networks," in *Proceedings of the IEEE international conference on computer vision*, 2017.
- [9] M. Zhang, X. Li, M. Xu, and Q. Li, "Automated semantic segmentation of red blood cells for sickle cell disease," *IEEE Journal of Biomedical and Health Informatics*, 2020.
- [10] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *arXiv preprint arXiv:1506.01497*, 2015.