MULTISTAR: INSTANCE SEGMENTATION OF OVERLAPPING OBJECTS WITH STAR-CONVEX POLYGONS

Florin C. Walter, Sebastian Damrich, Fred A. Hamprecht

HCI/IWR, University of Heidelberg, Germany

ABSTRACT

Instance segmentation of overlapping objects in biomedical images remains a largely unsolved problem. We take up this challenge and present *MultiStar*, an extension to the popular instance segmentation method *StarDist*. The key novelty of our method is that we identify pixels at which objects overlap and use this information to improve proposal sampling and to avoid suppressing proposals of truly overlapping objects. This allows us to apply the ideas of *StarDist* to images with overlapping objects, while incurring only a small overhead compared to the established method. *MultiStar* shows promising results on two datasets and has the advantage of using a simple and easy to train network architecture.

Index Terms— instance segmentation, overlapping objects, star-convex polygons, deep learning.

1. INTRODUCTION

Instance segmentation is the image analysis task of identifying distinct objects of the same category and assigning unique instance labels to the associated pixels. While difficult in itself, this problem becomes even more challenging when the objects are clustered and appear to overlap in the image projection. Many applications require segmentation of the full overlapping objects, which means that multiple instance labels need to be assigned to pixels in the object overlap.

A common biomedical application is the segmentation of cells in microscopy images, where objects can be dense and seemingly overlap with each other. Reliable automated processing of such images reduces the need for human experts.

In recent years, segmentation of overlapping objects has been addressed by a number of methods. [1] and [2] lift the label space to 3D and shear the segmentation masks, yielding a non-overlapping representation. [3] introduces an end-to-end approach with an encoder-decoder network, and [4] uses instance relation interaction. [5] presents a variational method for segmentation of near-circular shaped objects.

Instance segmentation of dense but non-overlapping objects has been demonstrated in a very effective and elegant way by *StarDist* [6][7], which parameterizes objects by star-convex polygons. With *MultiStar* we extend this method to images with overlapping objects, by additionally predicting
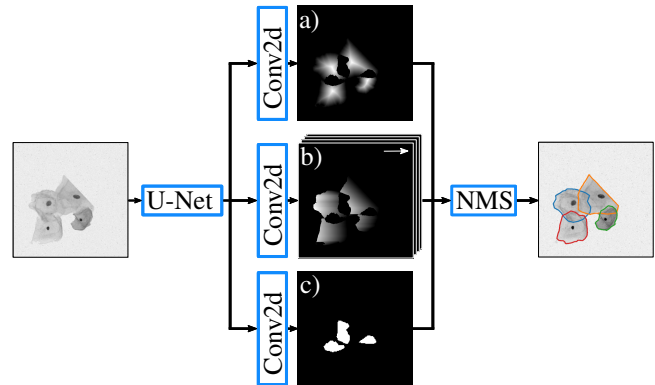

Fig. 1. The model predicts **a)** *Object Probability*, **b)** *Star Distances* (one of 32 directions shown) and **c)** *Overlap Probability*. Overlap areas are masked in a) and b). Incorporating the predicted overlap in the non-maximum suppression (NMS) allows to generate overlapping segmentations.

object overlap. In doing so, we can identify pixels at which a star-convex parameterization is unambiguously possible and account for the predicted overlap in the non-maximum suppression. The main strength of *MultiStar* is that while being a very simple and straight-forward modification of the successful *StarDist*, it greatly extends its application area.

2. METHOD

2.1. Review of *StarDist*

StarDist [6] is a proposal based method. For every pixel p it predicts a polygon to capture the pixel's object instance. The polygons are parameterized by the *Star Distances*, defined as the Euclidean distances from p to the polygon vertices, along a fixed number of equiangular radial directions. Proposals are generated at positions that are sampled according to the predicted *Object Probability*, defined inside objects as the normalized Euclidean distance transform and in the background as zero. The *Object Probability* furthermore acts as confidence score in the subsequent non-maximum suppression (NMS), which determines the final set of proposals representing the instance segmentation.

NMS requires setting a threshold on the intersection over union (IoU) between two proposals, above which the less confident proposal is suppressed. This threshold has to be low enough to avoid multiple detections of the same object. But a low threshold also prevents the final proposals from overlapping, even if the ground truth objects do overlap. *StarDist* therefore needs to trade off avoiding to detect false positives against detecting truly overlapping objects. This makes it difficult to determine the optimal IoU threshold in the NMS.

Another important point when considering images with object overlap is that *Star Distances* and *Object Probability* are ill-defined at overlap pixels, because it is not clear to which object they should refer.

Fig. 2. **a)** *StarDist* does not predict the object overlap (yellow), which leads to **b)** large IoU and thus **c)** suppression of the less confident proposal (dotted contour). **d)** *Multistar* predicts the overlap (red, hashed), **e)** excludes it from the intersection and due to the low IoU **f)** both proposals are accepted.

2.2. *MultiStar*

We solve both of these issues by additionally predicting the *Overlap Probability* P_{over} (Fig. 1 c)). Its ground truth value is 1 at pixels where at least two objects overlap and 0 elsewhere. The prediction can take intermediate values, reflecting different degrees of certainty.

Given the predicted *Object Probability* P_{obj}, we sample a proposal at pixel p with the thresholded probability

$$P_{proposal}(p) \propto P_{obj}(p) \cdot (1 - P_{over}(p)) \qquad (1)$$

as opposed to simply sampling from $P_{obj}(p)$. In the ideal case of correct overlap probabilities $P_{over}(p) \in \{0, 1\}$, this corresponds to sampling only at non-overlap pixels, where the proposal parameterization is well-defined. For intermediate values, the probability to sample at p is reduced according to how confidently the model predicts overlap at p.

In the NMS, *MultiStar* avoids suppressing detections of truly overlapping objects by excluding the predicted overlap from the intersection I of two proposals A and B

$$I \equiv \sum_{p \in A \cap B} (1 - P_{over}(p)) \qquad (2)$$

and computing the IoU with the usual definition of the union. The rationale here is that when two proposals overlap in an area where the model predicts object overlap (high P_{over}), the two proposals are probably correct detections of overlapping objects and not multiple detections of the same object. If computed with (2), the IoU in this case is low and neither of the two objects is suppressed (Fig. 2). Conversely, if two proposals overlap just slightly where the model does not predict any overlap, this indicates false positives. Hence, it is possible to choose a low IoU threshold without suppressing detections of truly overlapping objects.

MultiStar can detect an unlimited number of objects involved in the overlap. It can however not detect objects that are fully overlapped by other objects, because it does not allow to sample proposals at predicted overlap pixels.
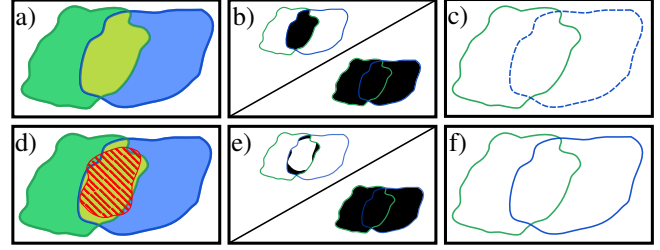
2.3. Model

We use a generic U-Net [8] with 3 output branches (Fig. 1), very similar to [6] but with one additional output branch for the *Overlap Probability* prediction.

The U-Net backbone has 5 levels with 16, 32, 64, 128 and 256 channels. Each down-/upsampling block consists of two 3×3 convolutions with Batch Normalization [9], ReLU activations and subsequent 2×2 max-pooling or upsampling. The U-Net output has 256 channels and ReLU activations.

We append three output branches for the three prediction features, consisting of a single convolutional layer each. In the *Object Probability* and *Overlap Probability* branches we use a single output channel with sigmoid activations. In the *Star Distances* branch we use 32 output channels for the 32 radial directions and ReLU activations.

2.4. Training

Inspired by [10], we optimize over the network parameters θ and the task uncertainties σ_i to minimize the regularized weighted sum of the three network output losses

$$
\begin{aligned}
L(\theta, \sigma_i) = \frac{1}{\sigma_{over}^2} L_{over}(\theta) + \frac{1}{\sigma_{obj}^2} L_{obj}(\theta) \\
+ \frac{1}{\sigma_{dist}^2} L_{dist}(\theta) + \log(\sigma_{over} \sigma_{obj} \sigma_{dist}).
\end{aligned}
\qquad (3)
$$

L_{over} and L_{obj} are binary cross-entropy losses. L_{dist} is the mean absolute difference between the predicted and true *Star Distances* with every pixel's contribution weighted by its true *Object Probability*. Pixels at which ground truth objects overlap are excluded from L_{obj} and L_{dist}. The model parameters and task uncertainties are jointly optimized with Adam [11] and learning rate 10^{-4}. We apply random flips, rotations and elastic deformations to augment the training data.

3. EXPERIMENTS

3.1. Evaluation metrics

We use the metrics described in [12]: For matched predictions and ground truth objects with dice coefficient greater than 0.7, we determine the average dice coefficient DC (higher is better) and pixel-based true positive and false positive rates TP_p and FP_p. The object-based false negative rate FN_o accounts for ground truth objects without matched prediction with dice coefficient greater than 0.7.

The above metrics do not penalize false positive detections. This is problematic, as an excessive amount thereof limits the usefulness of a segmentation result. Hence, we also compute the average precision $AP = \frac{TP}{TP+FP+FN}$ per image and average over all images. A pair of a prediction and a ground truth object is considered a true positive TP if its IoU is above a threshold τ, otherwise as a false positive FP.

3.2. Datasets and Results

The **OSC-ISBI** dataset contains images from the first and second *Overlapping Cervical Cytology Image Segmentation Challenge* [12][13]. As in [1-3], we train on the 945 synthetic images from [12] and the 8 training images from [13] and evaluate on the 9 test images from [13].

Our experiments show that the optimal choice of the threshold ρ on $P_{proposal}$ and the IoU threshold ν in the NMS depends on which metrics are evaluated. We obtain the best results with respect to DC, TP_p, FP_p and FN_o with $\rho = 0.2$ and $\nu = 0.5$ (*MultiStar1* in Table 1). Despite its good scores, this setting generates many false positive detections, see Fig. 3. By considering only AP with $\tau = 0.5$, we obtain the best results with $\rho = 0.1$ and $\nu = 0.2$ (*MultiStar2* in Table 1). Only FN_o is significantly worse in this setting than in *MultiStar1*. Nevertheless, qualitatively the segmentation results look more useful for a practitioner, see Fig. 3. This suggests that optimizing DC, TP_p, FP_p and FN_o alone might not be sufficient to identify a useful segmentation method for overlapping instances.

Note that *MultiStar* is trained using only the cytoplasm annotations, whereas [1][2][3] also use nuclei annotations. Remarkably, we still manage to achieve competitive DC, FN_o and TP_p scores, while only falling behind on the TP_p.

Our second dataset is based on the real microscopy images of cell nuclei from various organisms that were used in

	DC ↑	FN$_o$ ↓	TP$_p$ ↑	FP$_p$ ↓	AP ↑
ISOO$_{DL}$ [1]	$.86 \pm .07$	$.37 \pm .14$	$.90 \pm .11$	$.001 \pm .001$	-
ISOO$_{DL}$^{V2} [2]	$.90 \pm .08$	$.29 \pm .15$	$.90 \pm .11$	$.001 \pm .001$	-
Diskmask [3]	$.90 \pm .08$	$.22 \pm .13$	$.90 \pm .11$	$.001 \pm .001$	-
MultiStar1	$.86 \pm .07$	$.31 \pm .13$	$.83 \pm .10$	$.001 \pm .001$	$.21 \pm .06$
MultiStar2	$.85 \pm .07$	$.42 \pm .16$	$.82 \pm .10$	$.001 \pm .001$	$.47 \pm .13$

Table 1. Results on the test images of the **OSC-ISBI** dataset.
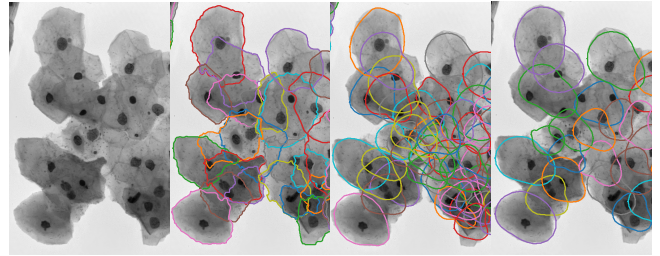

Fig. 3. Example image from the **OSC-ISBI** dataset. From left to right: original image, ground truth segmentation, *MultiStar1* segmentation, *MultiStar2* segmentation.

[6] to establish the performance of *StarDist*[1]. Since these images do not contain object overlap, we create the synthetic dataset **DSB-OV**[2] by randomly replicating, flipping, rotating and shifting objects in the images and adding up the intensities, such that at least 15% of the object pixels in every image are in the overlap of multiple objects.

Based on AP with different thresholds τ, we compare *MultiStar* with the pretrained *StarDist* model from [6] on **DSB-OV** in Table 2. We find the optimal values $\rho = 0.3$ and $\nu = 0.1$ for *MultiStar*, and $\rho = 0.5$ and $\nu = 0.5$ for *StarDist*. For all but the strictest τ *MultiStar* outperforms *StarDist*. As argued above, the optimal ν is significantly higher for *StarDist* than for *MultiStar*, since *StarDist* does not explicitly model overlapping instances. Even at this high ν, *StarDist* tends to merge overlapping objects, while *MultiStar* separates them much better, see Fig. 4. For *MultiStar* a low ν is enough to retain overlapping proposals while avoiding false positives.

τ	0.4	0.5	0.6	0.7	0.8
StarDist	$.62 \pm .15$	$.55 \pm .16$	$.46 \pm .16$	$.34 \pm .14$	$.19 \pm .10$
MultiStar	$.73 \pm .13$	$.66 \pm .16$	$.54 \pm .21$	$.33 \pm .22$	$.15 \pm .18$

Table 2. AP results of the pretrained *StarDist* and *MultiStar* with optimally chosen ρ and ν on the test images of the **DSB-OV** dataset.

4. DISCUSSION

We introduced a very simple but effective extension to the popular method *StarDist* [6], which extends its applicability to images with strongly overlapping objects. With a much simpler architecture and without exploiting cell nuclei annotation for training, we are only slightly below the current state of the art on the **OSC-ISBI** dataset. Compared to the original *StarDist*, we achieve a substantial boost in performance

---

[1] github.com/mpicbg-csbd/stardist/releases/download/0.1.0/dsb2018.zip subset of the *Kaggle Data Science Bowl 2018* dataset available at kaggle.com/c/data-science-bowl-2018

[2] Code for our model and the DSB-OV dataset is publicly available at https://github.com/hci-unihd/MultiStar.

Authorized licensed use limited to: Universitaet Heidelberg. Downloaded on February 07,2022 at 11:07:02 UTC from IEEE Xplore. Restrictions apply.
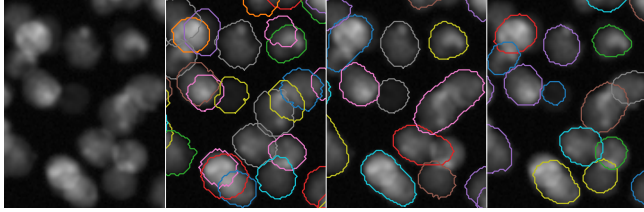
Fig. 4. Example image from the **DSB-OV** dataset. From left to right: original image, ground truth segmentation, pre-trained *StarDist* segmentation, *MultiStar* segmentation.

on images of overlapping cell nuclei, which we believe could be further improved by using a more elaborate training procedure.

5. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using human and animal subject data made available in open access by [12], [13] and [14].

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] Anton Böhm, Annekathrin Ücker, Tim Jäger, Olaf Ronneberger, and Thorsten Falk, "Isoo dl: Instance segmentation of overlapping biological objects using deep learning," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 1225–1229.

[2] Anton Böhm, Maxim Tatarchenko, and Thorsten Falk, "Isoo v2 dl-semantic instance segmentation of touching and overlapping objects," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 2019, pp. 343–347.

[3] Anton Böhm, Nikolaus Mayer, and Thomas Brox, "Diskmask: Focusing object features for accurate instance segmentation of elongated or overlapping objects," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2020, pp. 230–234.

[4] Yanning Zhou, Hao Chen, Jiaqi Xu, Qi Dou, and Pheng-Ann Heng, "Irnet: Instance relation network for overlapping cervical cell segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 640–648.

[5] Csaba Molnar, Zoltan Kato, and Ian H Jermyn, "A new model for the segmentation of multiple, overlapping, near-circular objects," in *2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*. IEEE, 2015, pp. 1–5.

[6] Uwe Schmidt, Martin Weigert, Coleman Broaddus, and Gene Myers, "Cell detection with star-convex polygons," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 265–273.

[7] Martin Weigert, Uwe Schmidt, Robert Haase, Ko Sugawara, and Gene Myers, "Star-convex polyhedra for 3d object detection and segmentation in microscopy," in *The IEEE Winter Conference on Applications of Computer Vision*, 2020, pp. 3666–3673.

[8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[9] Sergey Ioffe and Christian Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[10] Alex Kendall, Yarin Gal, and Roberto Cipolla, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7482–7491.

[11] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[12] Zhi Lu, Gustavo Carneiro, Andrew P Bradley, Daniela Ushizima, Masoud S Nosrati, Andrea GC Bianchi, Claudia M Carneiro, and Ghassan Hamarneh, "Evaluation of three algorithms for the segmentation of overlapping cervical cells," *IEEE journal of biomedical and health informatics*, vol. 21, no. 2, pp. 441–450, 2016.

[13] Zhi Lu, Gustavo Carneiro, and Andrew P Bradley, "An improved joint optimization of multiple level set functions for the segmentation of overlapping cervical cells," *IEEE Transactions on Image Processing*, vol. 24, no. 4, pp. 1261–1272, 2015.

[14] Juan C Caicedo, Allen Goodman, Kyle W Karhohs, Beth A Cimini, Jeanelle Ackerman, Marzieh Haghighi, CherKeng Heng, Tim Becker, Minh Doan, Claire McQuin, et al., "Nucleus segmentation across imaging experiments: the 2018 data science bowl," *Nature methods*, vol. 16, no. 12, pp. 1247–1253, 2019.