-Supplementary Material-
DoNet: Deep De-overlapping Network for Cytology Instance Segmentation

Hao Jiang[1] Rushan Zhang[1] Yanning Zhou[2] Yumeng Wang[1] Hao Chen[1]
[1]The Hong Kong University of Science and Technology [2]Tencent AI Lab
{hjiangaz,jhc}@cse.ust.hk, {rzhangbq,ywanglu}@connect.ust.hk, amandayzhou@tencent.com

A. Data Pre-processing

Compared to standard annotation in COCO format, we need annotations of the instance masks $\mathcal{E}_k = \{e_{k,i}\}_{i=1}^{N_k}$ together with its sub-region masks, $\mathcal{O}_k = \{o_{k,i}\}_{i=1}^{N_k}$, and $\mathcal{M}_k = \{m_{k,i}\}_{i=1}^{N_k}$. We don not need extra annotating for $\mathcal{O}_k = \{o_{k,i}\}_{i=1}^{N_k}$, $\mathcal{M}_k = \{m_{k,i}\}_{i=1}^{N_k}$, which can be obtained by logical operations from $\mathcal{E}_k = \{e_{k,i}\}_{i=1}^{N_k}$, as follows,

$$\begin{aligned} \mathcal{O}_k &= \{o_{k,i}\}_{i=1}^{N_k}, o_{k,i} = e_{k,i} \cap e_{k,j}, j = 1, ...i-1, i+1, ..., N_k \\ \mathcal{M}_k &= \{m_{k,i}\}_{i=1}^{N_k}, m_{k,i} = o_{k,i} \oplus e_{k,i} \end{aligned} \tag{1}$$

B. Data Augmentation

We propose to extend the CPS dataset with instance-level data augmentation, which can be divided into two steps: instance generation and image synthesis. In the first step, we build an isolated instance set and a clustered instance set. Then, we generate a series of synthesized images based on the two instance banks controlling the instance popularity and distribution (Algorithm 1).

C. Structure of Units

Figure 1 shows details of Concatenation Unit (CU), Fusion Unit (FU), Mergence Unit (MU), and Mask Head. Firstly, CU is utilized to concate two extracted feature maps $f_{k,i}^c$ and $f_{k,i}^{roi}$, which contains rich semantic information and morphological information, respectively. CU is composed of 3 convolution layers with the kernel size of 3×3 + ReLU. Secondly, FU is a key component designed to recombine predictions of intersection, complement, and integral instances, which is also composed of 3 convolution layers with the kernel size of 3×3 + ReLU. Then, MU in CRM is designed to merge intersection $\hat{o}_{k,i}$ masks and complement $\hat{m}_{k,i}$ masks, where they pass through a sigmoid function for normalization, and then through a pixel-level Exclusive-OR operation, obtaining the merged mask $\hat{e}_{k,i}^e$. Finally, we illustrate the structure of the mask head adopted in H_i, H_o, and H_m, which consists of 4 convolution layers with the kernel size of 3×3, a deconvolution layer with the kernel size of 2×2, and a convolution layer for final prediction.

D. Comparison with State-of-the-arts

We provide more qualitative comparisons among our DoNet and other methods on CPS (Figure 2) and ISBI2014 (Figure 3) datasets. In addition, detailed results of the CPS dataset for every fold in terms of 6 metrics: TPp, FNo, mAP, Dice, F1, and AJI are shown in Table 1-4. We also list the results on both categories of instances (cytoplasm and nuclei) and their average value. Among these metrics, our DoNet not only achieves the best average performance, but also achieves the best performance on each fold. After the adoption of data augmentation, DoNet continues to gain performance improvements. Note that our model performs better on cytoplasm than on nucleus. This is because the main focus of our proposed Decompose-and-Recombined strategy is on solving the overlapping problem in cytoplasm regions.

Algorithm 1: An synthetic pipeline for data augmentation

Input: Dataset $\mathcal{D} = \{(\mathcal{X}_k, \mathcal{Y}_k)\}_{k=1}^{K}$ with annotations of instance categories $\mathcal{C}_k = \{c_{k,i}\}_{i=1}^{N_k}$, and instance masks $\mathcal{E}_k = \{e_{k,i}\}_{i=1}^{N_k}$, total instance number $\mathcal{N} = \{n_i\}_{i=1}^{N_n}$, lowest ratio $\mathcal{L} = \{l_i\}_{i=1}^{N_L}$, high ratio $\mathcal{H} = \{h_j\}_{j=1}^{N_H}$, and transparency ratio r_t.

Output: $\mathcal{D}_{\text{a}} = \{(\mathcal{X}_p, \mathcal{Y}_p)\}_{p=1}^{P}$.

Initialize two instance sets, $S^I = \emptyset$ and $S^C = \emptyset$;
for $k = 1, ..., K$ **do**
  for $i = 1, ..., N_k$ **do**
    Crop instance mask $m_{k,i}$ from $\mathcal{X}_k$ using $e_{k,i}$;
    $S^I \leftarrow S^I \cup \{(m_{k,i}, e_{k,i})\}$ for isolated instances;
    $S^C \leftarrow S^C \cup \{(m_{k,i}, e_{k,i})\}$ for cluster instances;
  end
end
for $g = 1, ..., N_n$ **do**
  for $i = 1, ..., N_L$ **do**
    for $j = i, ..., N_H$ **do**
      $\{x_q\}_{q=1}^{g} \leftarrow Sample(S^I, S^C)$;
      Geometric transformation: $x_q^t = \mathcal{T}\{x_q\}, \mathcal{T} \in \{\text{rotation, scaling, affine}\}$;
      Adjust the overlap ratio between (l_i, h_j);
      Adjust the transparency in the overlapping regions by r_t;
    end
  end
end

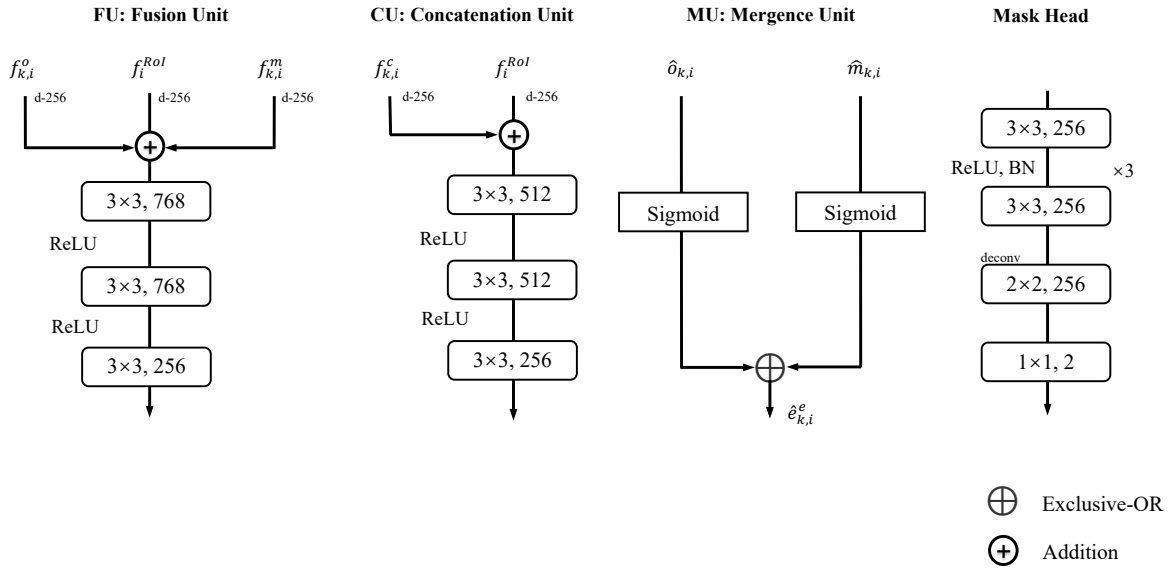
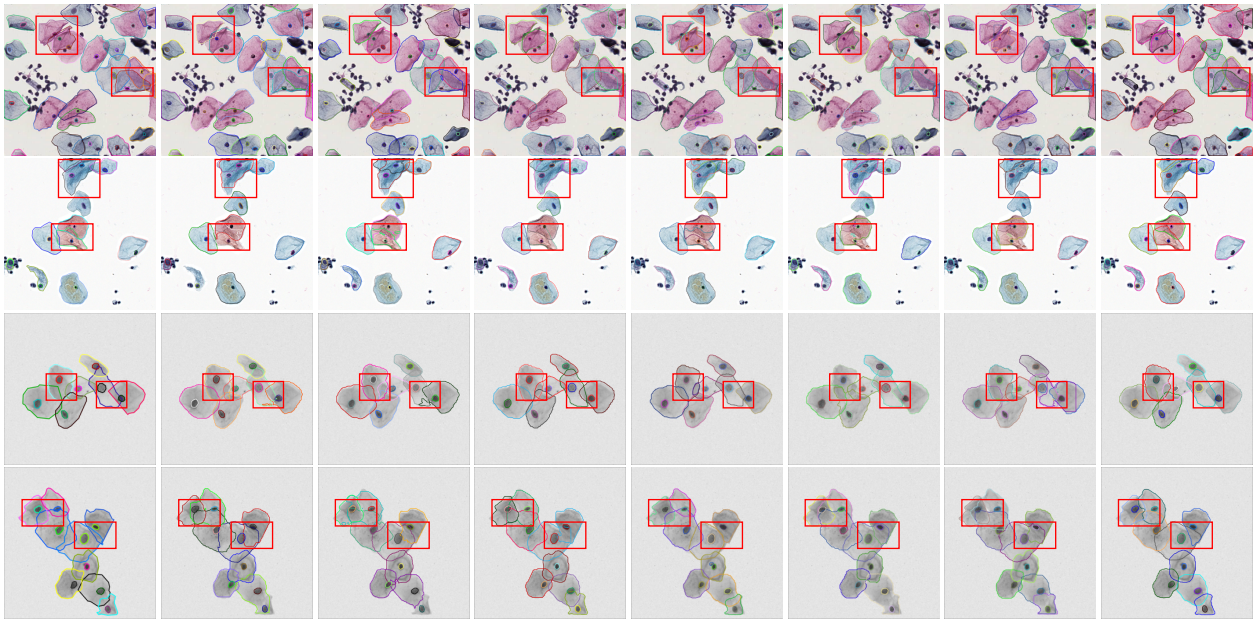
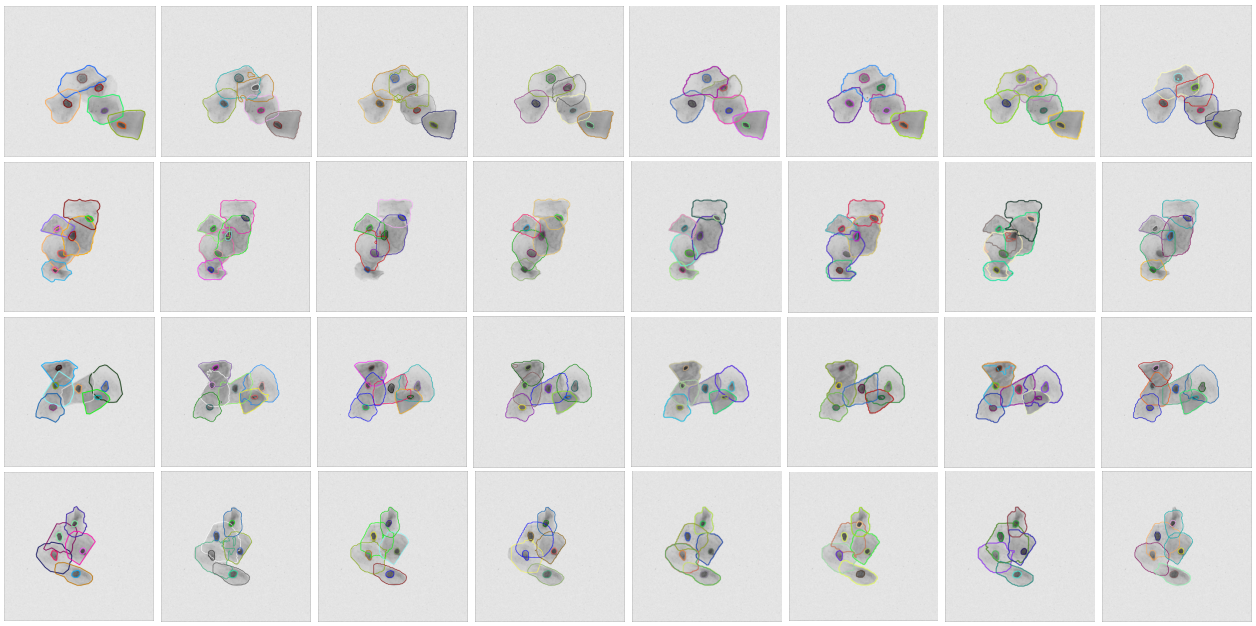


Figure 1. Structural designs CU, FU, MU, and Mask Head in DoNet.



(a) GT (b) MRCNN (c) ORCNN (d) SSP (e) CMRCNN (f) HTC (g) MSRCNN (h) DoNet (our)

Figure 2. Qualitative results of our DoNet and other SOTA methods on CPS dataset. (a) Ground Truth; (b) Mask R-CNN [4]; (c) Occlusion R-CNN [3]; (d) Xiao et al. [6]; (e) Cascade R-CNN [1]; (f) Hybrid Task Cascade [2]; (g) Mask Scoring R-CNN [5]; (h) Our proposed DoNet.



(a) GT (b) MRCNN (c) ORCNN (d) SSP (e) CMRCNN (f) HTC (g) MSRCNN (h) DoNet (our)

Figure 3. Qualitative results of our DoNet and other SOTA methods on ISBI2014 dataset. (a) Ground Truth; (b) Mask R-CNN [4]; (c) Occlusion R-CNN [3]; (d) Xiao et al. [6]; (e) Cascade R-CNN [1]; (f) Hybrid Task Cascade [2]; (g) Mask Scoring R-CNN [5]; (h) Our proposed DoNet.

Table 1. Quantitative segmentation results of DoNet and other state-of-the-art methods on CPS dataset (mAP↑).

Methods	Cytoplasm				Nuclei				Average
	Fold1	Fold2	Fold3	Average	Fold1	Fold2	Fold3	Average	
Mask R-CNN [4]	63.48	57.35	56.83	59.22 ± 3.70	39.97	**35.57**	36.28	37.27 ± 2.36	48.24 ± 3.03
Cascade R-CNN [1]	62.10	55.62	55.72	57.81 ± 3.72	41.08	35.60	37.13	**37.93 ± 2.83**	47.87 ± 3.27
Mask Scoring R-CNN [5]	64.85	58.19	58.50	60.51 ± 3.76	38.50	33.55	36.73	36.26 ± 2.51	48.38 ± 3.13
HTC [2]	62.10	55.05	54.76	57.30 ± 4.16	41.08	35.23	**37.36**	37.89 ± 2.96	47.60 ± 3.56
Occlusion R-CNN [3]	63.66	58.04	56.87	59.52 ± 3.63	38.80	34.70	36.79	36.76 ± 2.05	48.14 ± 2.84
Xiao et al. [6]	63.84	57.61	57.57	59.67 ± 3.61	39.61	35.44	37.15	37.40 ± 2.10	48.53 ± 2.85
DoNet	66.20	59.95	58.23	61.46 ± 4.20	**41.19**	34.44	36.56	37.40 ± 3.45	49.43 ± 3.83
DoNet w/ Aug.	**67.28**	**59.99**	**59.75**	**62.34 ± 4.28**	40.05	34.77	36.06	36.96 ± 2.75	**49.65 ± 3.52**

Table 2. Quantitative segmentation results of DoNet and other state-of-the-art methods on CPS dataset (Dice↑).

Methods	Cytoplasm				Nuclei				Average
	Fold1	Fold2	Fold3	Average	Fold1	Fold2	Fold3	Average	
Mask R-CNN [4]	92.65	91.31	**92.24**	92.07 ± 0.68	86.11	86.36	86.74	86.40 ± 0.32	89.23 ± 0.50
Cascade R-CNN [1]	92.98	91.72	91.92	92.21 ± 0.67	86.36	86.05	86.43	86.28 0.20 ±	89.24 ± 0.44
Mask Scoring R-CNN [5]	92.77	**92.16**	91.93	92.29 ± 0.43	86.49	86.47	86.56	86.50 ± 0.05	89.39 ± 0.24
HTC [2]	92.98	91.59	91.72	92.10 ± 0.77	86.36	85.90	85.92	86.06 ± 0.26	89.08 ± 0.51
Occlusion R-CNN [3]	92.41	91.37	92.01	91.93 ± 0.53	86.20	86.27	86.20	86.22 ± 0.04	89.08 ± 0.28
Xiao et al. [6]	92.36	91.75	92.06	92.06 ± 0.31	86.53	86.37	86.71	86.53 ± 0.17	89.29 ± 0.24
DoNet	92.74	91.93	91.94	92.20 ± 0.46	**86.87**	**86.91**	**86.83**	**86.87 ± 0.04**	**89.54 ± 0.25**
DoNet w/ Aug.	**93.08**	91.97	92.08	**92.38 ± 0.61**	86.50	86.77	86.60	86.62 ± 0.14	89.50 ± 0.38

Table 3. Quantitative segmentation results of DoNet and other state-of-the-art methods on CPS dataset (F1↑).

Methods	Cytoplasm				Nuclei				Average
	Fold1	Fold2	Fold3	Average	Fold1	Fold2	Fold3	Average	
Mask R-CNN [4]	84.06	87.40	85.18	85.55 ± 1.70	**88.95**	82.30	83.57	84.94 ± 3.53	85.24 ± 2.62
Cascade R-CNN [1]	82.58	84.77	84.34	83.90 ± 1.16	85.15	81.01	82.15	82.77 ± 2.14	83.33 ± 1.65
Mask Scoring R-CNN [5]	82.61	85.15	83.33	83.70 ± 1.31	84.99	80.47	81.32	82.26 ± 2.40	82.98 ± 1.86
HTC [2]	82.58	82.21	80.71	81.83 ± 0.99	85.15	76.95	80.21	80.77 ± 4.13	81.30 ± 2.56
Occlusion R-CNN [3]	84.54	88.18	85.17	85.97 ± 1.94	87.93	82.73	85.59	85.42 ± 2.61	85.69 ± 2.28
Xiao et al. [6]	84.31	88.18	84.83	85.77 ± 2.10	88.68	**82.88**	83.88	85.15 ± 3.10	85.46 ± 2.60
DoNet	85.47	87.62	84.87	85.99 ± 1.44	88.55	82.21	84.37	85.04 ± 3.22	85.51 ± 2.33
DoNet w/ Aug.	**87.42**	**88.63**	**86.37**	**87.47 ± 1.13**	88.11	82.34	**84.93**	**85.12 ± 2.89**	**86.30 ± 2.01**

Table 4. Quantitative segmentation results of DoNet and other state-of-the-art methods on CPS dataset (AJI↑).

Methods	Cytoplasm				Nuclei				Average
	Fold1	Fold2	Fold3	Average	Fold1	Fold2	Fold3	Average	
Mask R-CNN [4]	74.32	78.84	75.23	76.13 ± 2.39	65.60	**59.89**	61.71	62.40 ± 2.92	69.27 ± 2.65
Cascade R-CNN [1]	74.76	78.13	74.91	75.93 ± 1.90	67.74	58.08	59.55	61.79 ± 5.20	68.86 ± 3.55
Mask Scoring R-CNN [5]	73.84	78.88	73.84	75.52 ± 2.91	61.56	57.66	58.92	59.38 ± 1.99	67.45 ± 2.45
HTC [2]	74.76	77.73	73.19	75.22 ± 2.31	60.72	53.97	57.77	57.49 ± 3.38	66.35 ± 2.84
Occlusion R-CNN [3]	75.60	79.78	75.05	76.81 ± 2.58	64.26	59.69	62.70	62.22 ± 2.32	69.51 ± 2.45
Xiao et al. [6]	74.55	79.81	74.80	76.39 ± 2.97	65.41	59.95	61.69	62.35 ± 2.79	69.37 ± 2.88
DoNet	76.29	79.84	75.64	77.26 ± 2.26	**66.54**	59.77	62.42	**62.91 ± 3.41**	70.08 ± 2.84
DoNet w/ Aug.	**78.09**	**80.84**	**76.77**	**78.56** ± 2.08	64.78	59.69	**63.17**	62.55 ± 2.60	**70.56 ± 2.34**

References

[1] Zhaowei Cai and Nuno Vasconcelos. Cascade r-cnn: Delving into high quality object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6154–6162, 2018. 3, 4

[2] Kai Chen, Jiangmiao Pang, Jiaqi Wang, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jianping Shi, Wanli Ouyang, et al. Hybrid task cascade for instance segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4974–4983, 2019. 3, 4

[3] Patrick Follmann, Rebecca König, Philipp Härtinger, Michael Klostermann, and Tobias Böttger. Learning to see the invisible: End-to-end trainable amodal instance segmentation. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1328–1336. IEEE, 2019. 3, 4

[4] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017. 3, 4

[5] Zhaojin Huang, Lichao Huang, Yongchao Gong, Chang Huang, and Xinggang Wang. Mask scoring r-cnn. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6409–6418, 2019. 3, 4

[6] Yuting Xiao, Yanyu Xu, Ziming Zhong, Weixin Luo, Jiawei Li, and Shenghua Gao. Amodal segmentation based on visible region segmentation and shape prior. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 2995–3003, 2021. 3, 4