Author's version of Böhm et al. "DiskMask: Focusing Object Features for Accurate Instance Segmentation of Elongated or Overlapping Objects".2020 IEEE 17th International Symposium on Biomedical Imaging, DOI: 10.1109/ISBI45749.2020.9098435

Supplementary: Parameter, Augmentation, Pre- and Post-Processing

C. Elegans

Pre-Processing: We normalize data (division by 3200) and delete all loosely connected components in segmentation masks (area is smaller than 4px).

Default Augmentations: We don't use elastic deformations and scaling augmentation, but random cropping, flipping, rotation and image extrapolation with zero values. We don't use any intensity value augmentations.

Default Parameters: We set the reference points to the cell centroids, drawing disks with an adaptive radius of 9px around them. During the training, we drop the masks with a probability of 50%.

Note, inference time, reported for this dataset, was estimated using the *tile-predict strategy* (cf. [1]). It was implemented in MATLAB without any kind of parallelization.

OSC-ISBI

Note, for this experiment, we do *not* make use of synthetically generated data (as e.g. done in [2]:Tab. 2:"$\mathrm{ours}_{\mathrm{plus}}$"). We use 8 images for training and 9 for testing.

Pre-Processing: Since the "OSC-ISBI" dataset does not provide a mapping between the nuclei and their corresponding cells, we assigned them as follows (cf. [3]): We assume that the nucleus is located in the cell center. First, we separate merged nuclei using the Watershed transform, then we assign the nucleus segments to cells using the Hungarian method [4][1] based on their distances from the cells' centers of mass. Unmatched nucleus segments are assigned to the closest cell.

We train on the sub-sampled (2x) images, upscaling them to the original resolution in the inference time using bilinear (images) and nearest neighbour (object masks) interpolation.

Default Augmentations: We don't use scaling augmentation, but random cropping, flipping, rotation, image extrapolation with zero values and elastic deformation (cf. [1]) with 3×3 seed points for every of two dimensions, which are drawn from a Gaussian distribution with a standard deviation of 10px. To get a displacement vector for every position, we bilinearly interpolate between the seed points.

We randomly augment image intensities (strictly increasing intensity transformation, cf. [5]).

Default Parameters: We set the reference points to the nuclei centroids, drawing disks with an adaptive radius of 9px around them. During the training, we drop the masks with a probability of 50%.

[1] Used implementation: link.

LSC

Pre-Processing: We normalize the images (division by 255).

Default Augmentations: We don't use elastic deformations, but random cropping, flipping, rotation, image extrapolation with zero values and scaling in the range [1,1.5].

Default Parameters: We set the reference points to the leaf centroids, drawing disks with an adaptive radius of 9px around them. During the training, we drop the masks with a probability of 50%.

Post-Processing: We project the instance masks to a 2D plane (required by the competition), so that every instance has an unique ID. In case of an overlap, the overlapping region is randomly assigned to one of the masks which causes the overlap.

Supplementary: Dataset Splits

C. Elegans

Training-Split:

```
XXX_A01_w2_15...
XXX_A02_w2_B4...
XXX_A03_w2_34...
XXX_A04_w2_0A...
XXX_A05_w2_9F...
XXX_A06_w2_11...
XXX_A07_w2_26...
XXX_A08_w2_AC...
XXX_A09_w2_2F...
XXX_A10_w2_F2...
XXX_A11_w2_7B...
XXX_A12_w2_AD...
XXX_A13_w2_F3...
XXX_A14_w2_B8...
XXX_A15_w2_76...
XXX_A16_w2_43...
XXX_A17_w2_2D...
XXX_A18_w2_09...
XXX_A19_w2_7C...
XXX_A20_w2_E9...
XXX_A21_w2_E7...
XXX_A22_w2_27...
XXX_A23_w2_71...
XXX_A24_w2_FB...
XXX_B01_w2_33...
XXX_B02_w2_CF...
XXX_B03_w2_D8...
XXX_B04_w2_0D...
XXX_B05_w2_29...
XXX_B06_w2_2D...
XXX_B07_w2_F3...
```

```
XXX_B08_w2_E0...
XXX_B09_w2_44...
XXX_B10_w2_C2...
XXX_B11_w2_26...
XXX_B12_w2_8E...
XXX_B13_w2_70...
XXX_B14_w2_ED...
XXX_B15_w2_4E...
XXX_B16_w2_52...
XXX_B17_w2_ED...
XXX_B18_w2_9B...
XXX_B19_w2_B4...
XXX_B20_w2_62...
XXX_B21_w2_35...
XXX_B22_w2_7D...
XXX_B23_w2_FF...
XXX_B24_w2_3E...
XXX_C01_w2_CB...
XXX_C02_w2_B1...
```

Test-Split:

```
XXX_C03_w2_DB...
XXX_C04_w2_AB...
XXX_C05_w2_A2...
XXX_C06_w2_FF...
XXX_C07_w2_CA...
XXX_C08_w2_1A...
XXX_C09_w2_67...
XXX_C10_w2_D0...
XXX_C11_w2_C6...
XXX_C12_w2_A0...
XXX_C13_w2_19...
XXX_C14_w2_4A...
XXX_C15_w2_23...
XXX_C16_w2_82...
XXX_C17_w2_02...
XXX_C18_w2_89...
XXX_C19_w2_82...
XXX_C20_w2_5F...
XXX_C21_w2_10...
XXX_C22_w2_FB...
XXX_C23_w2_02...
XXX_C24_w2_29...
XXX_D01_w2_29...
XXX_D02_w2_F8...
XXX_D03_w2_6D...
```

Validation-Split:

```
XXX_D04_w2_90...
XXX_D05_w2_22...
XXX_D06_w2_1D...
XXX_D07_w2_1F...
XXX_D08_w2_7F...
XXX_D09_w2_89...
XXX_D10_w2_47...
XXX_D11_w2_E2...
XXX_D12_w2_EA...
XXX_D13_w2_ED...
XXX_D14_w2_13...
XXX_D15_w2_66...
XXX_D16_w2_B4...
XXX_D17_w2_20...
XXX_D18_w2_3E...
XXX_D19_w2_3E...
XXX_D20_w2_54...
XXX_D21_w2_87...
XXX_D22_w2_1E...
XXX_D23_w2_D3...
XXX_D24_w2_02...
XXX_E01_w2_0F...
XXX_E02_w2_AC...
XXX_E03_w2_31...
XXX_E04_w2_8A...
```

LSC

Training-Split:

```
plant001
plant005
plant006
plant007
plant008
plant011
plant012
plant013
plant015
plant017
plant018
plant020
plant021
plant024
plant026
plant027
plant029
plant031
plant032
plant033
plant035
plant037
plant038
plant039
plant040
plant043
plant044
plant045
plant046
plant048
```

plant049
plant050
plant051
plant053
plant054
plant055
plant057
plant059
plant060
plant061
plant062
plant064
plant065
plant067
plant068
plant070
plant071
plant072
plant073
plant077
plant078
plant079
plant080
plant083
plant084
plant085
plant086
plant089
plant090
plant091
plant092
plant096
plant098
plant099
plant100
plant102
plant104
plant105
plant106
plant108
plant109
plant110
plant113
plant115
plant116
plant118
plant119
plant121
plant123
plant124
plant126
plant128
plant129
plant130
plant132
plant134
plant135
plant137
plant138
plant141
plant142
plant143
plant144
plant146
plant147
plant148
plant149
plant152
plant153
plant154
plant156
plant161

Validation-Split:

plant002
plant010
plant016
plant022
plant030
plant036
plant042
plant047
plant052
plant058
plant063
plant069
plant076
plant082
plant088
plant094
plant101
plant107
plant114
plant120
plant127
plant133
plant139
plant145
plant151
plant159

Supplementary: Implementation Details

To accelerate the training process, we sample tiles from a distribution which is derived from the ground truth segmentation masks by filtering their projection with a Gaussian smoothing kernel (sigma=50px). Moreover, during training, we build the feature pools from the ground truth reference points and not from the predicted ones.

At inference time, we use a threshold of 0.5 to form the binary mask from the DecoNet's scores. In case, there is a disk-detection, but the corresponding binary segmentation map is empty, we exclude it from the evaluation.

To compute the objectness/confidence score (required for AP computation) for an object, we average the soft scores within the object's binary mask.

Supplementary: Corrections

- Typo in [6], Sec. 3.1.2: ”We set the nms-threshold to 0.5 and ~~double~~ *halve* the lengths of square anchors at all 5 resolution levels.”[2]

- Incomplete representation of Fig.2 in [6]: replace Fig.2 in [6] by Fig.1 [2]

1. REFERENCES

[1] 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*, 2015.

[2] Anton Böhm, Annekathrin Ücker, Tim Jäger, Olaf Ronneberger, and Thorsten Falk, “Isoo dl: Instance segmentation of overlapping biological objects using deep learning,” in *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th*, 2018.

[3] 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.

[4] Harold W Kuhn, “The hungarian method for the assignment problem,” *Naval research logistics quarterly*, 1955.

[5] Thorsten Falk, Dominic Mai, Robert Bensch, Özgün Çiçek, Ahmed Abdulkadir, Yassine Marrakchi, Anton Böhm, Jan Deubner, Zoe Jäckel, Katharina Seiwald, et al., “U-net: deep learning for cell counting, detection, and morphometry,” *Nature methods*, vol. 16, no. 1, pp. 67–70, 2019.

[2]no change in results

[6] 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.

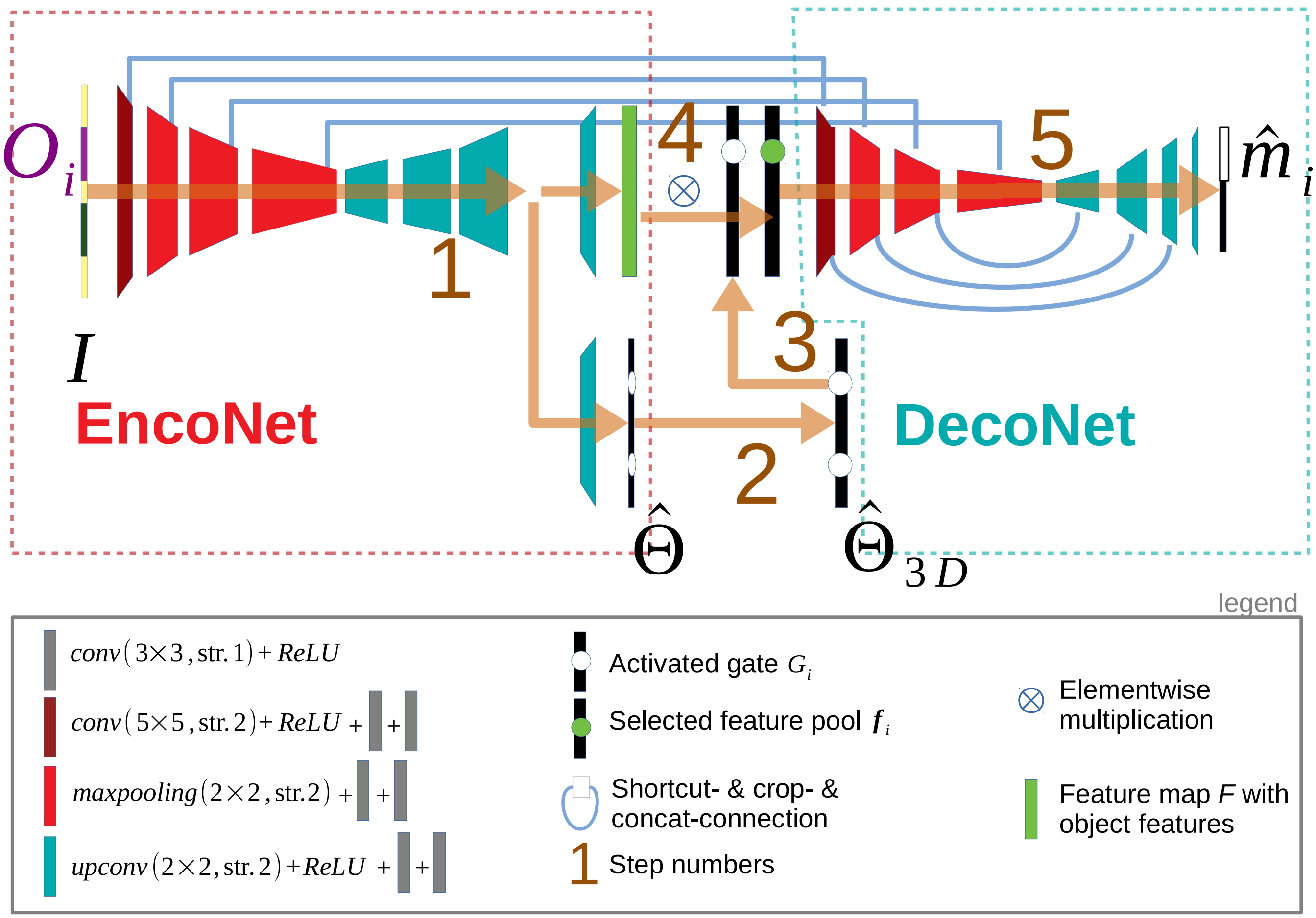

Fig. 1. Composite architecture with two encoder-decoder networks (EncoNet and DecoNet) at inference time (see Sec. 2.2).