nature methods

ARTICLES

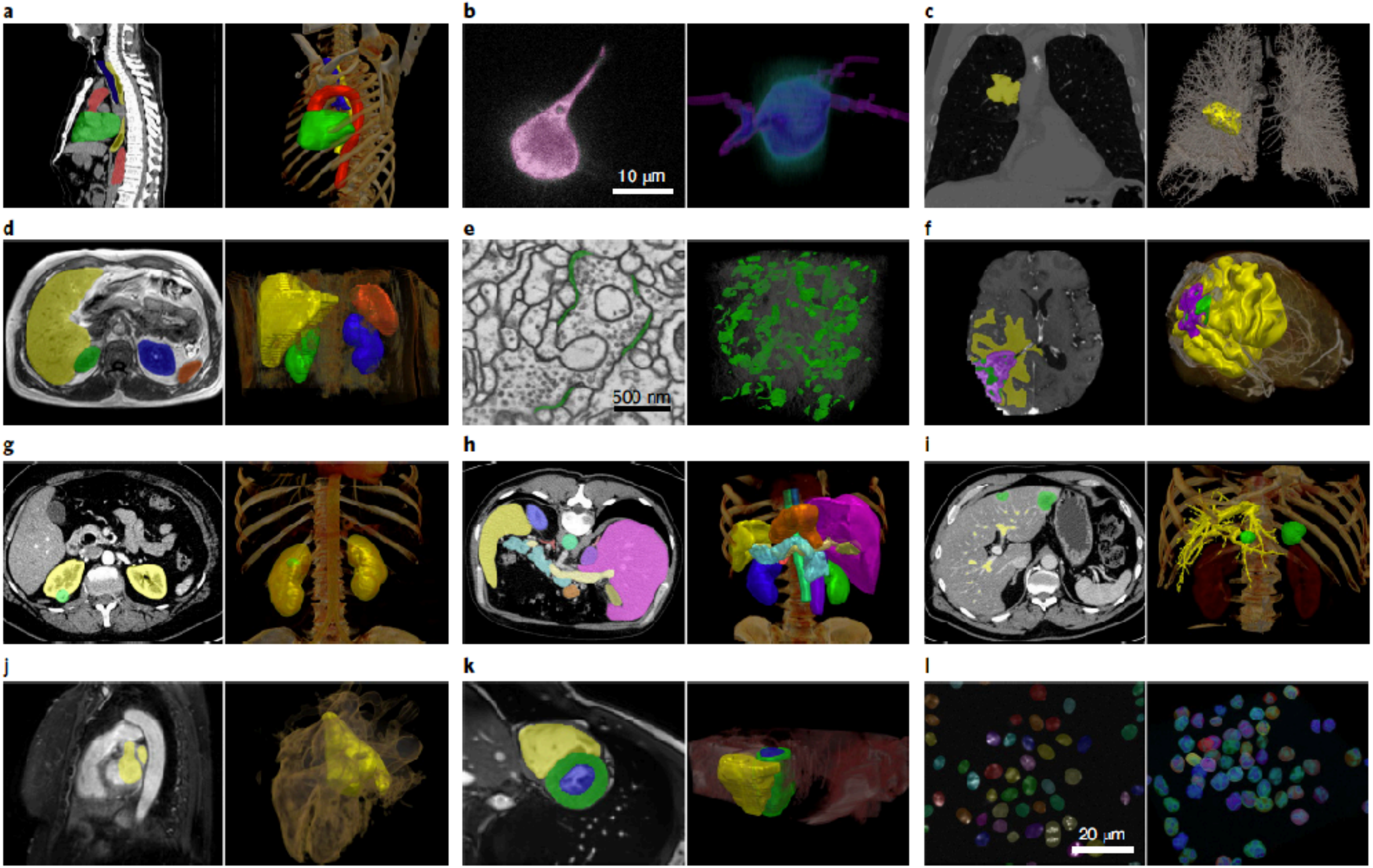
https://doi.org/10.1038/s41592-020-01008-z

Check for updates

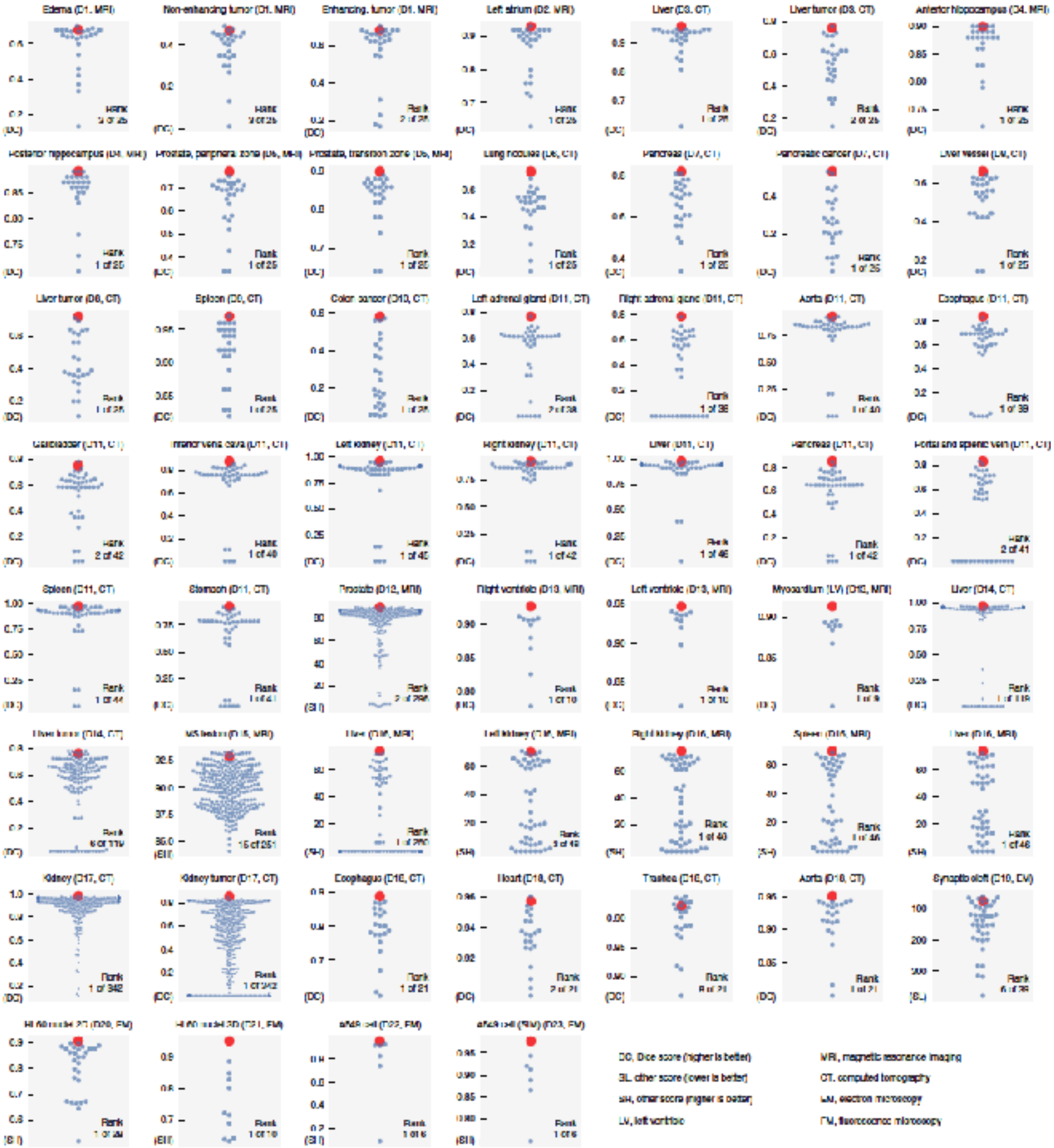
nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation

Fabian Isensee[1,2,6], Paul F. Jaeger[1,6], Simon A. A. Kohl[1,3], Jens Petersen[1,4] and Klaus H. Maier-Hein[1,5]

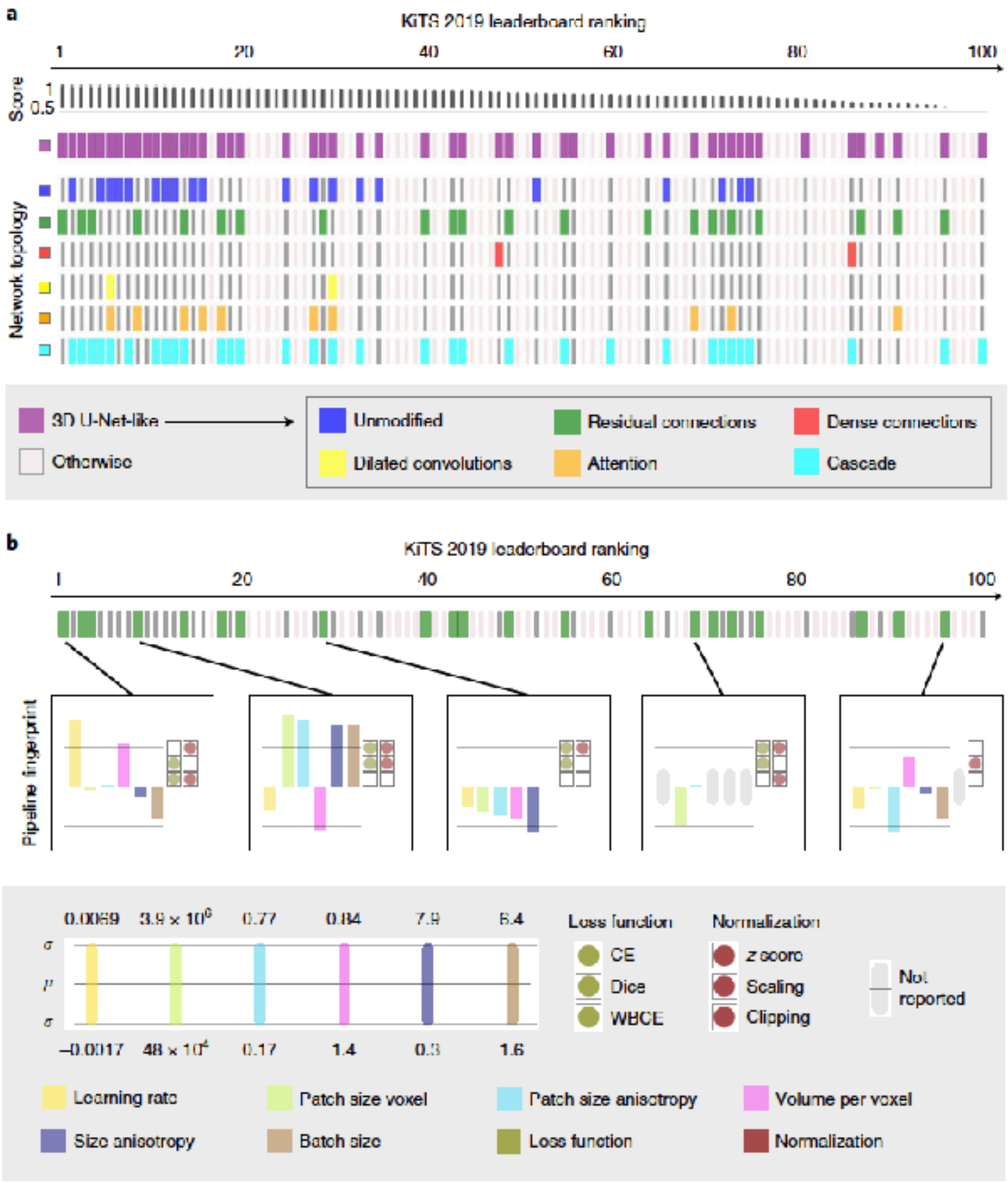
Without manual intervention, nnU-Net surpasses most existing approaches, including highly specialized solutions on 23 (53 in total) public datasets used in international biomedical segmentation competitions.

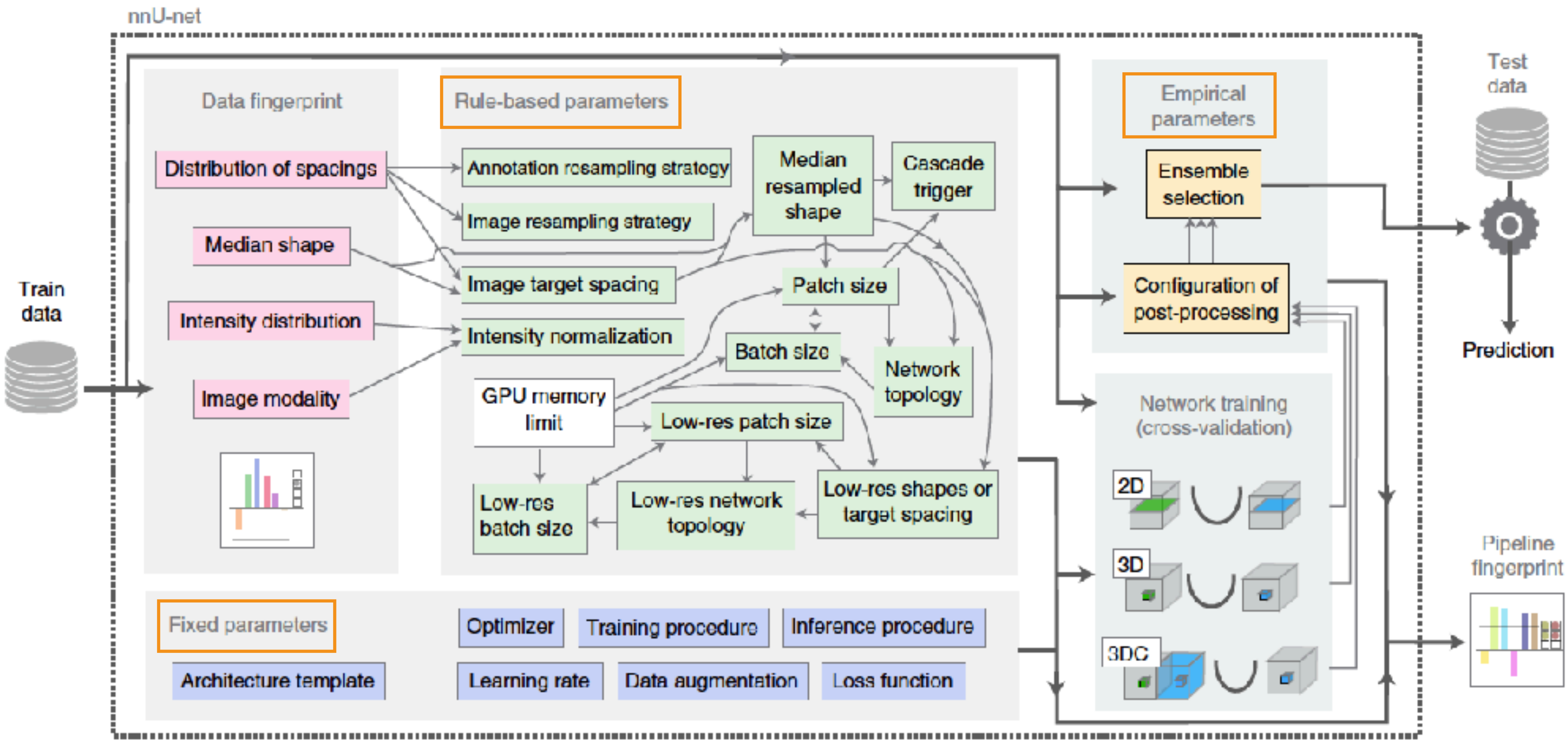

a
b
c
d
e
f
g
h
i
j
k
l
10 μm
500 nm
20 μm

ranking on 53 public challenges



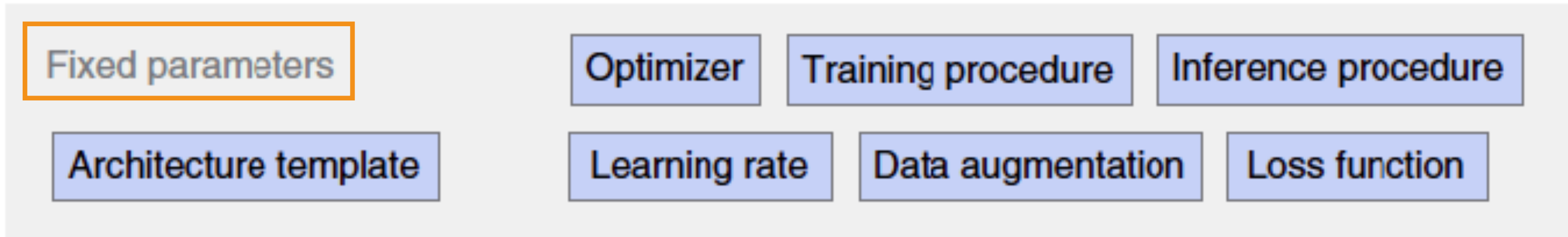
Motivation


a
KiTS 2019 leaderboard ranking
1
20
40
60
80
100
Score
1
0.5
Network topology
3D U-Net-like
Otherwise
Unmodified
Residual connections
Dense connections
Dilated convolutions
Attention
Cascade
b
KiTS 2019 leaderboard ranking
Pipeline fingerprint
0.0069
0.77
0.84
7.9
6.4
−0.0017
0.17
1.4
0.3
1.6
Loss function
CE
Dice
WBCE
Normalization
z score
Scaling
Clipping
Not reported
Learning rate
Patch size voxel
Patch size anisotropy
Volume per voxel
Size anisotropy
Batch size
Loss function
Normalization



train 10~20 models

Fixed Parameters



- **Architecture Template**

1. 2D/3D/cascaded 3D U-net without variantion.
2. two blocks per resolution
3. each block: conv(3*3) - instance norm, leaky ReLU.
4. Downsampling is done with strided convolutions, upsampling is done with convolutions transposed.

- **Training procedure**

1. Device: one GPU with more than 11GB memory.
2. 1000 epochs, 250 itertions;

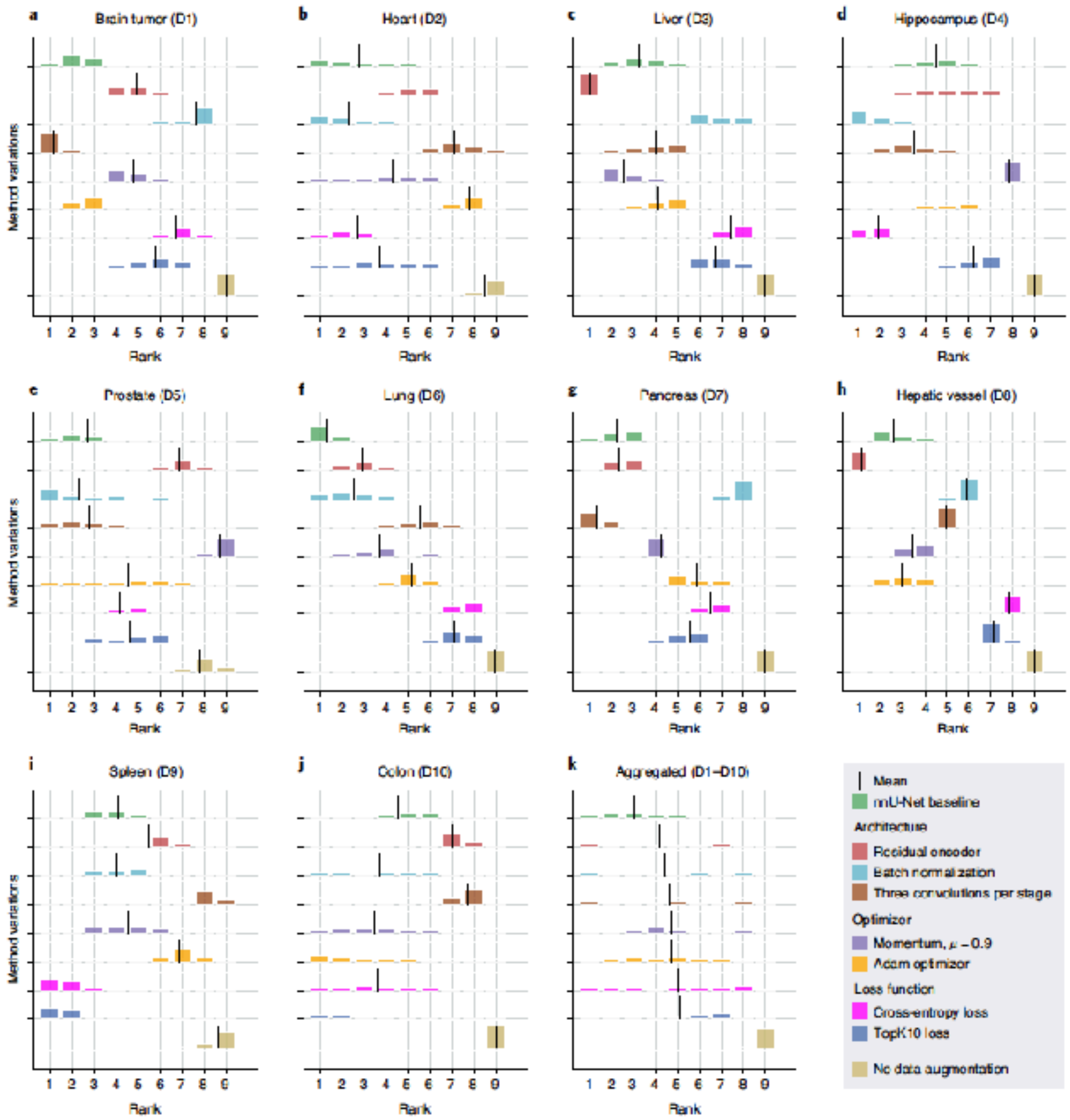
- **Loss**

Combining the Dice loss with a cross-entropy loss

- **Learning rate**

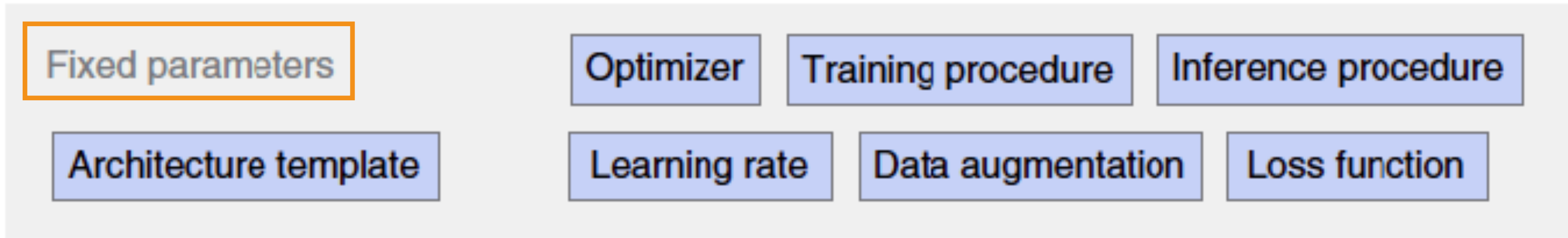
Initial learning rate 0.01 with nesterov momentum 0.99

Fixed Parameters



- One thousand **virtual validation** sets were generated via bootstrapping (drawn with replacement).

Fixed Parameters



- **Data augmentation**

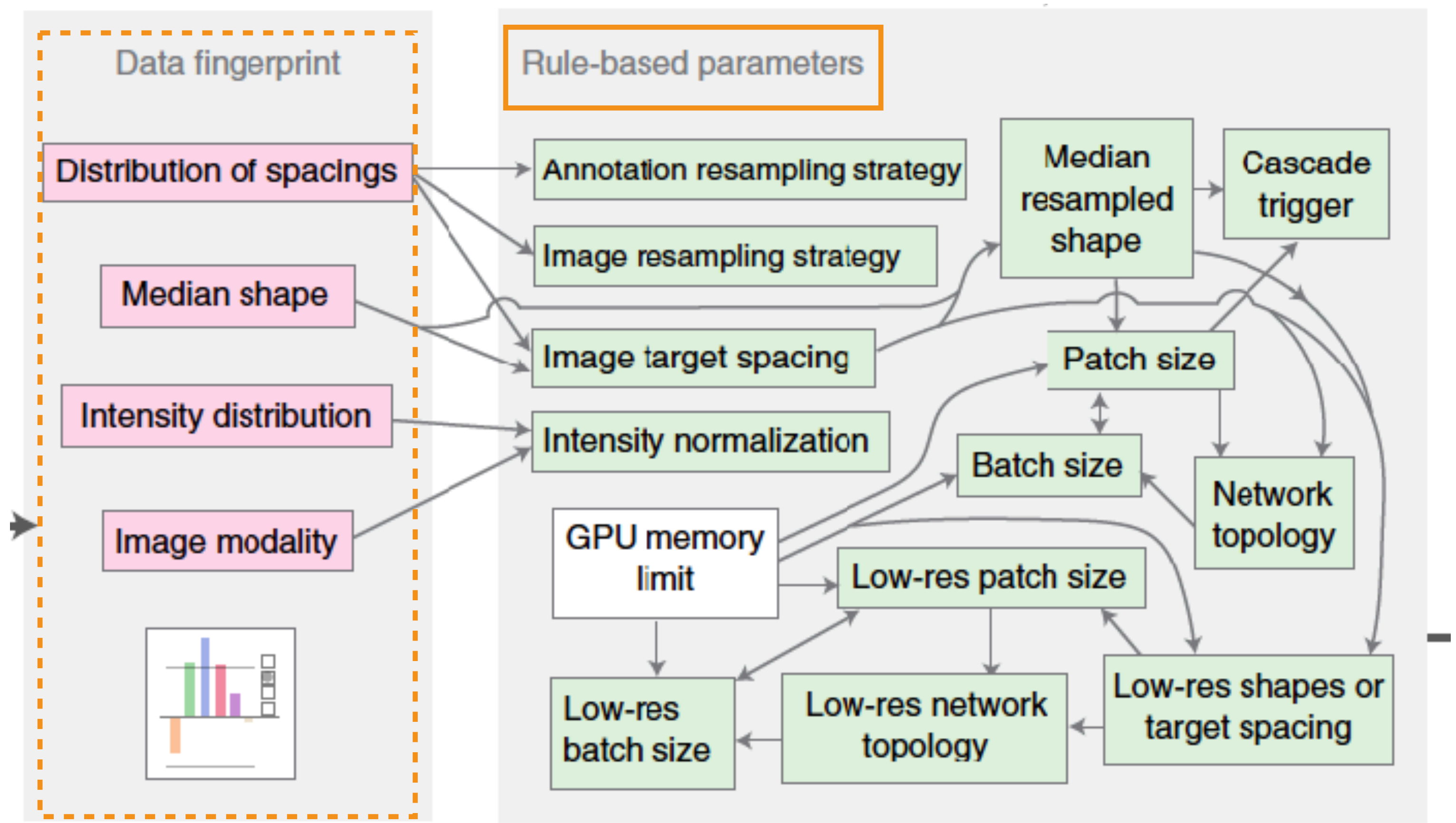
1. Oversampling foreground regions
2. Rotation and scaling
3. Gaussian noise
4. Gaussian blur
5. Brightness
6. Contrast
7. Simulation of low resolution
8. Gamma augmentation
9. Mirroring

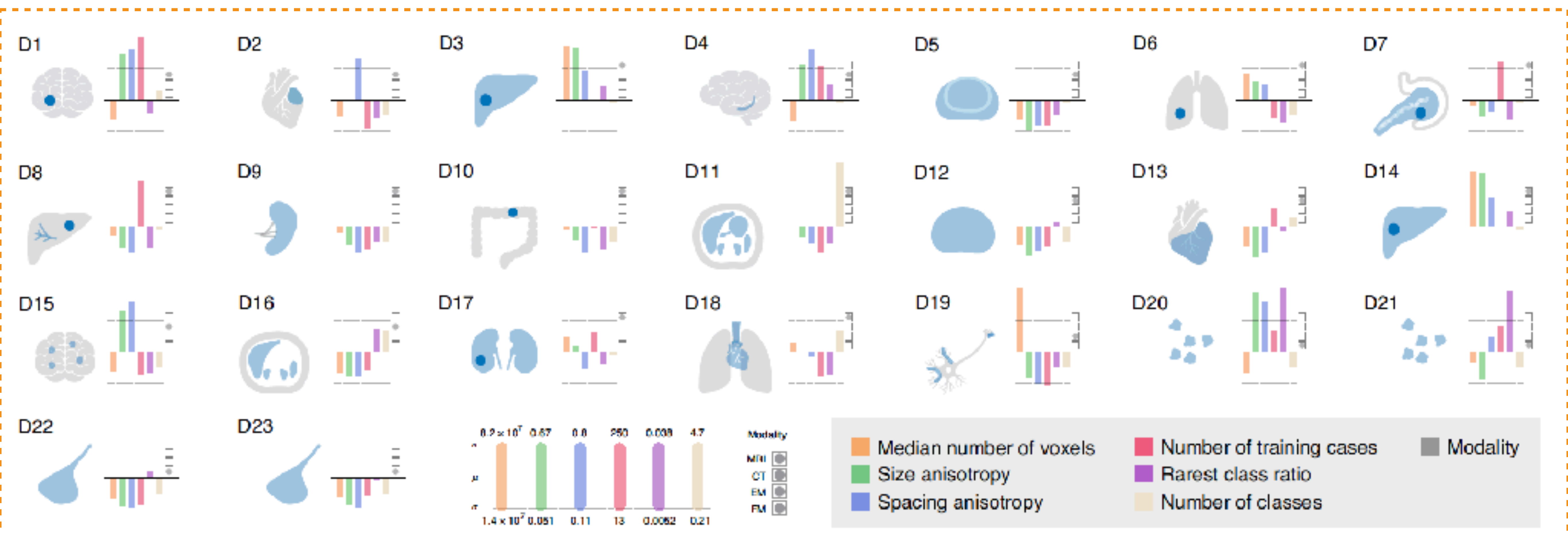
Essential in competitions

- **Inference procedure**

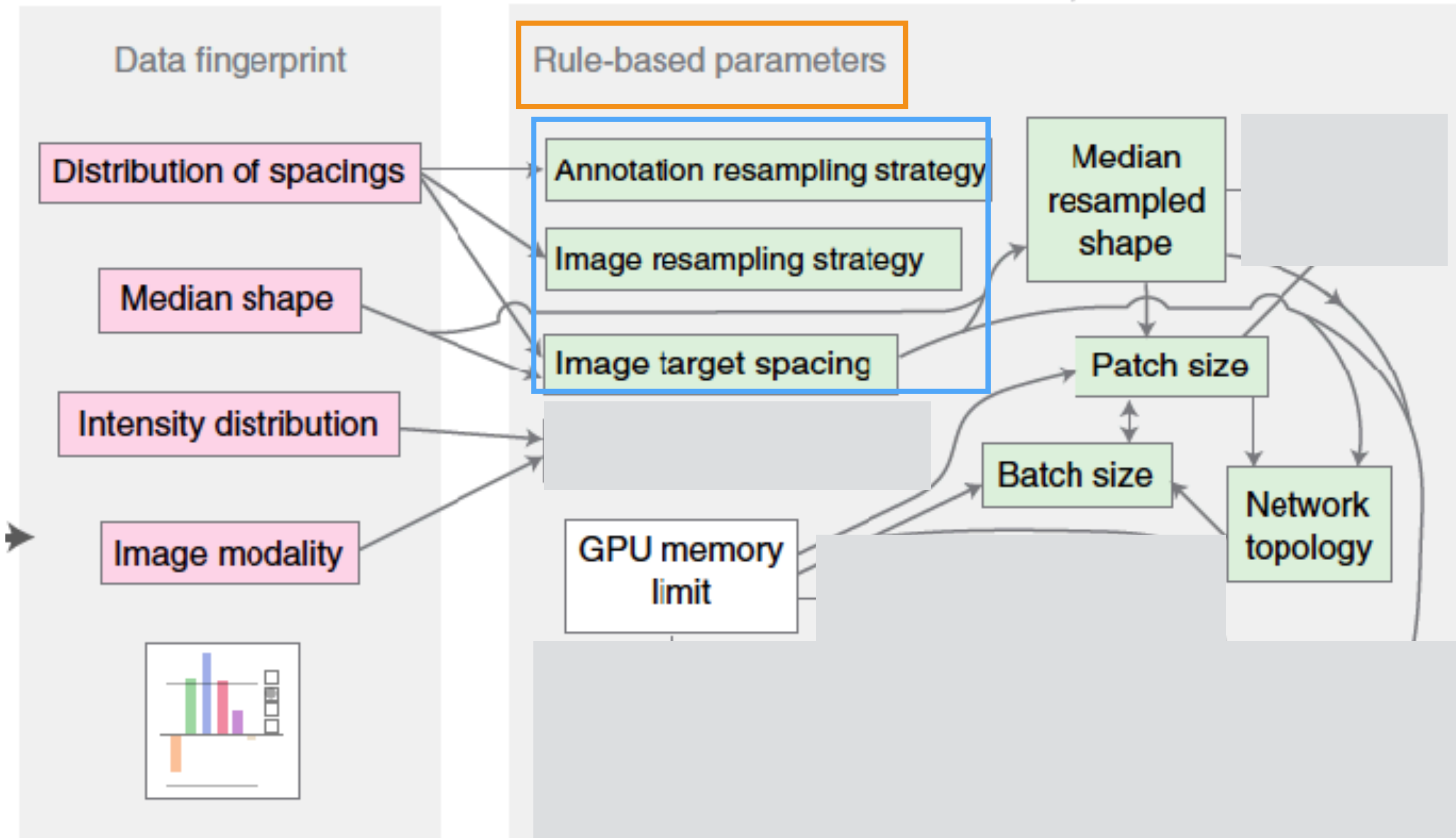
1. 5-fold cross validation;
2. Inference is done patch based with the same patch size as used during training;
3. Overlapping predictions with a distance of patch_size/ 2, aggravate with Gaussian weight.

Rule-based Parameters





Rule-based Parameters



- **Target spacing and resampling**

Resampling with third order spline (data) and linear interpolation to the median spacing of training cases, except for the out of plane axis in anisotropic data.

- **Network size**

1. Additional loss functions are applied to all but the two lowest resolutions of the decoder to inject gradient deep into the network;
2. For anisotropic data, no pooling and 3D conv in low resolution.
3. downsampling until the feature maps are relatively small (minimum is 4*4(*4));

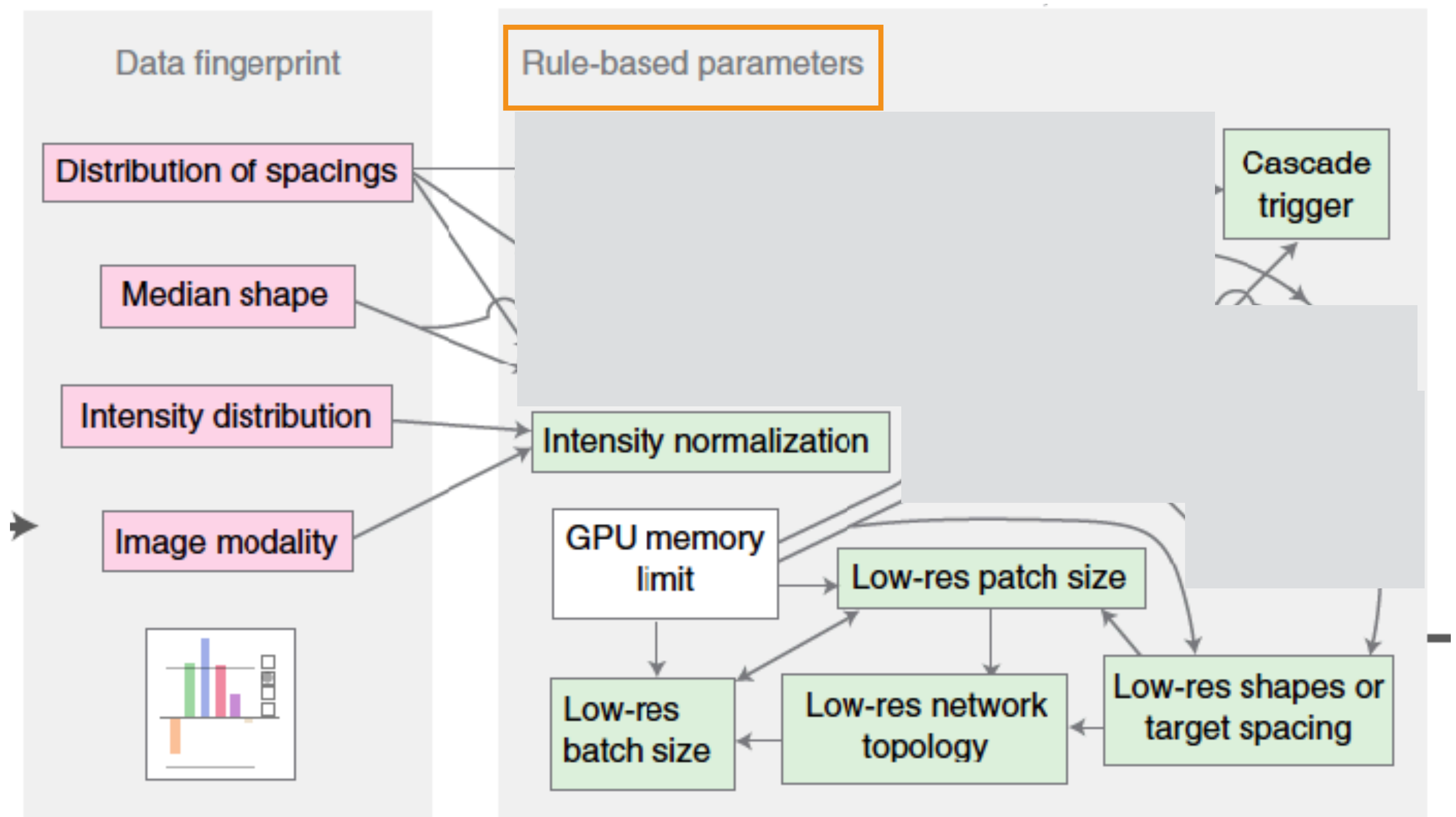
- **Patch size**

1. As large as possible while still allowing a batch size of 2;
2. Aspect ratio follows media image shape after resampling.

- **Batch size**

1. Minimum of 2;
2. As large as possible after patch configuration.

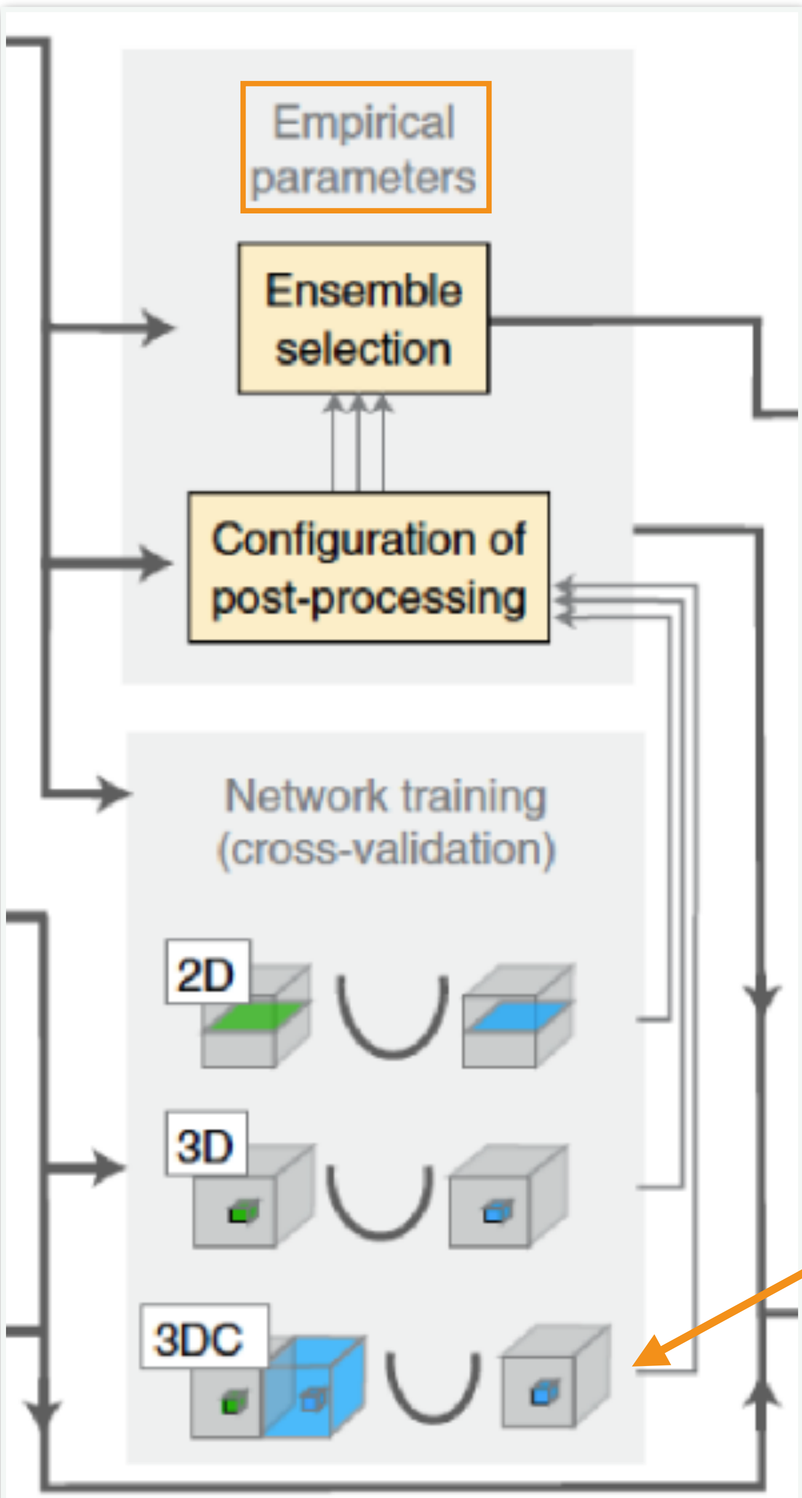
Rule-based Parameters



- **Intensity normalization:**

 1. Z-score per image (mean substraction and division by standard deviation)
We deviate from this default only for CT images, where a global normalization scheme is determined.

Empirical Parameters

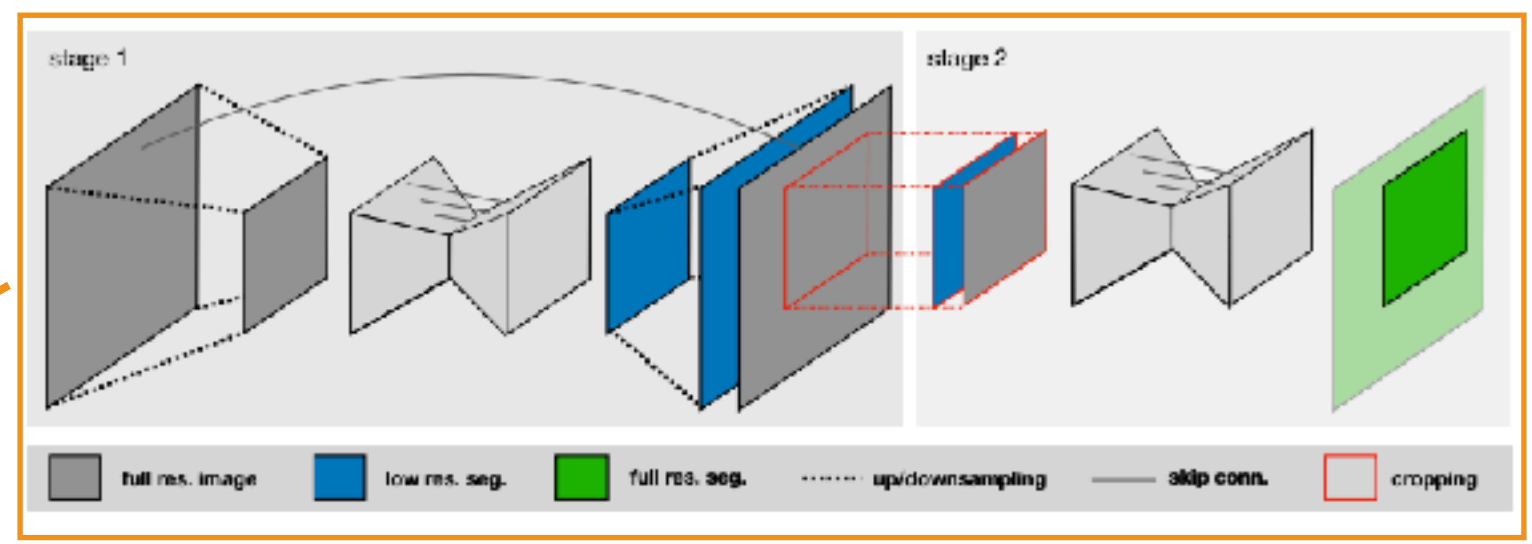


Model selection:

From 2D/3D/cascaded 3D Unet, select the best performing method (or ensemble of methods) after cross-validation. This includes training 20 models (5/5/2*5 for 2D/3D/cascaded 3D Unet).

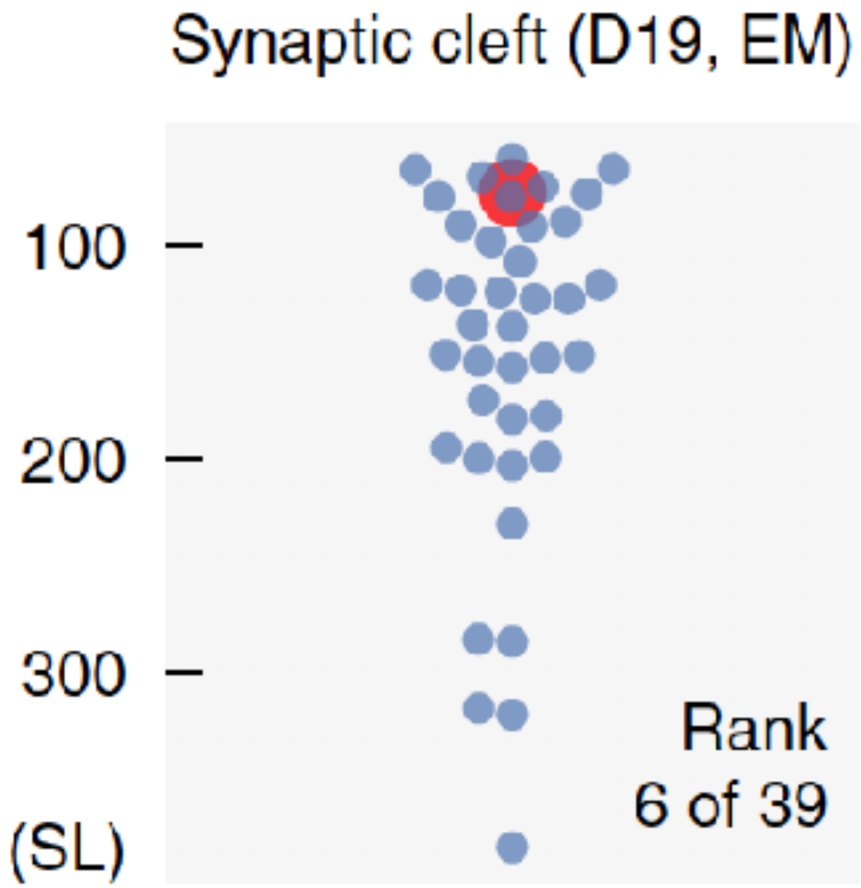
Postprocessing:

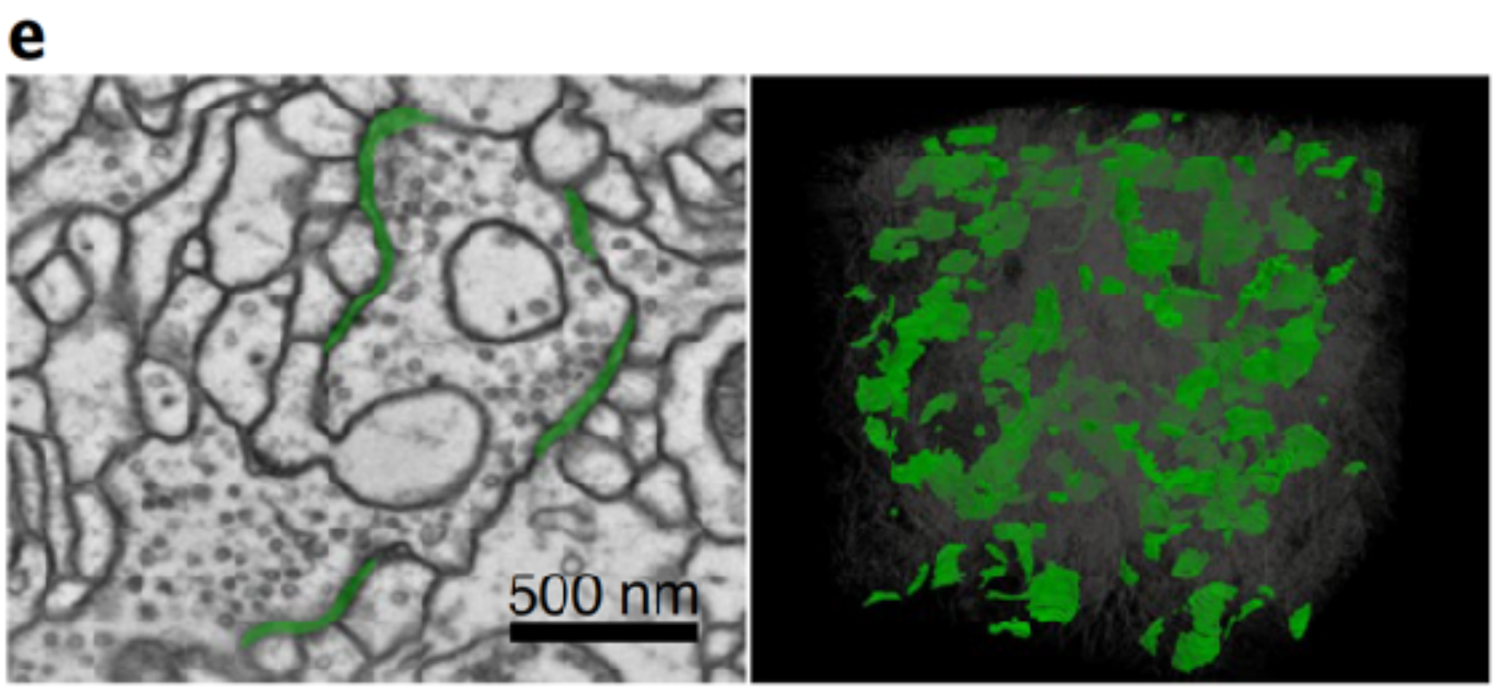
Whether to remove all but the largest component.



cascaded 3D Unet for large images (not necessary)

Failure Case





manual adaptation of the loss function, as well as EM-specific preprocessing, may be necessary to surpass state-of-the-art performance