Organ segmentation using U-Net like models

Hidy Gábor

Superviosr: dr. András Lukács

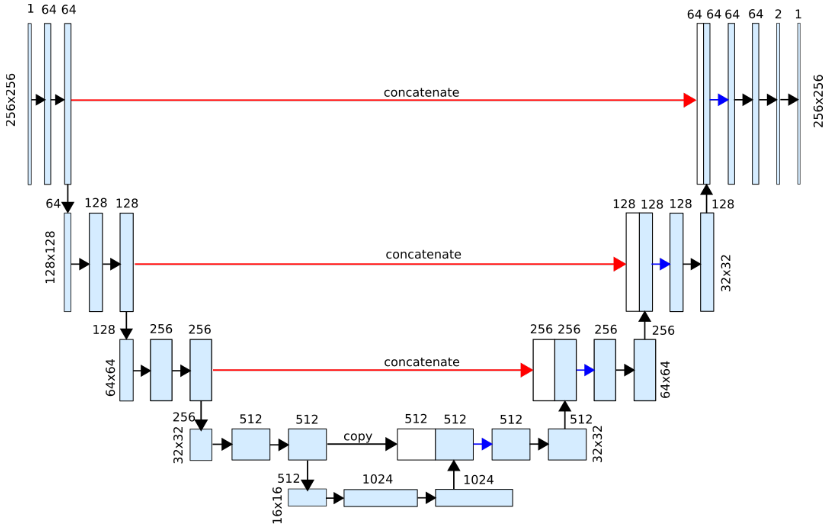
Semantic segmentation

- input: $x \in [0, 1]^{D \times H \times W}$ image
- output: $\hat{y} \in [0, 1]^{C \times H \times W}$ prediction
- target: $y \in \{0, 1\}^{C \times H \times W}$ mask

Metrics

- accuracy: $\dfrac{|y \cap \hat{y}|}{HW}$
- Dice similarity coefficient: $\dfrac{2|y \cap \hat{y}|}{|y| + |\hat{y}|}$
- IoU score: $\dfrac{|y \cap \hat{y}|}{|y \cup \hat{y}|}$
- average modified (95$^{\text{th}}$ percentile) Hausdorff distance
- area under ROC, PR curves

U-Net


1 64 64
256x256
concatenate
64 64 64 64 2 1
256x256
64 128 128
128x128
concatenate
128 128 128 128
32x32
128 256 256
64x64
concatenate
256 256 256 256
64x64
256 512 512
32x32
copy
512 512 512 512
32x32
512 1024 1024
16x16

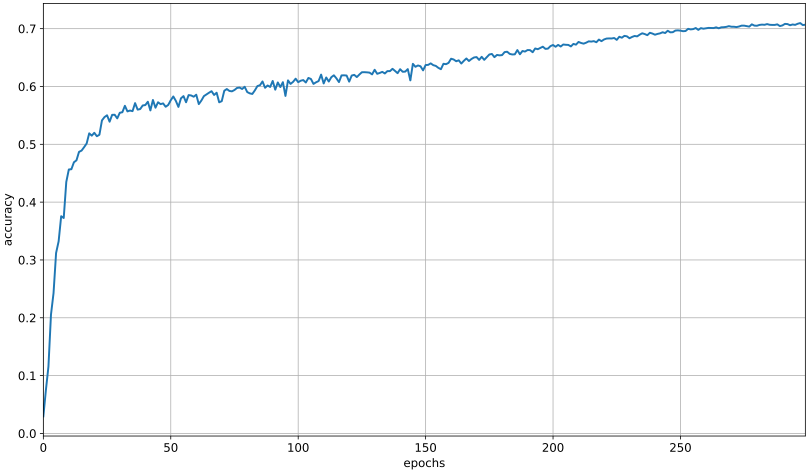
Synapse

- multi-organ CT dataset
- segmentation mask for eight organs
- 2D slices of 3D CT images
- train: 12 samples, 1800 slices
- validation: 4 samples, 300 slices

Experiments with weight initialisation

weight init	acc	AUC	AP	DSC	IoU	HD95
He (fan out)	0.990	0.969	0.785	0.735	0.619	0.003
He (fan in)	0.990	0.979	0.776	0.731	0.612	0.003
Glorot	0.990	0.976	0.762	0.718	0.601	0.003
orthogonal	0.991	0.976	0.789	0.737	0.614	0.003
pretrained encoder	0.990	0.983	0.817	0.750	0.626	0.004

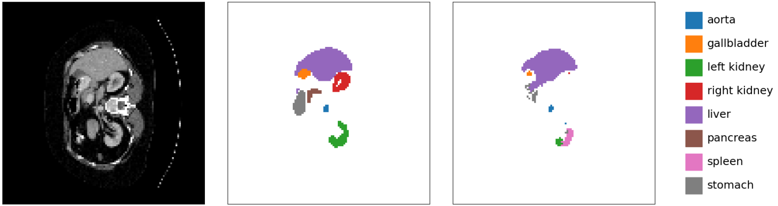
Training on Imagenet



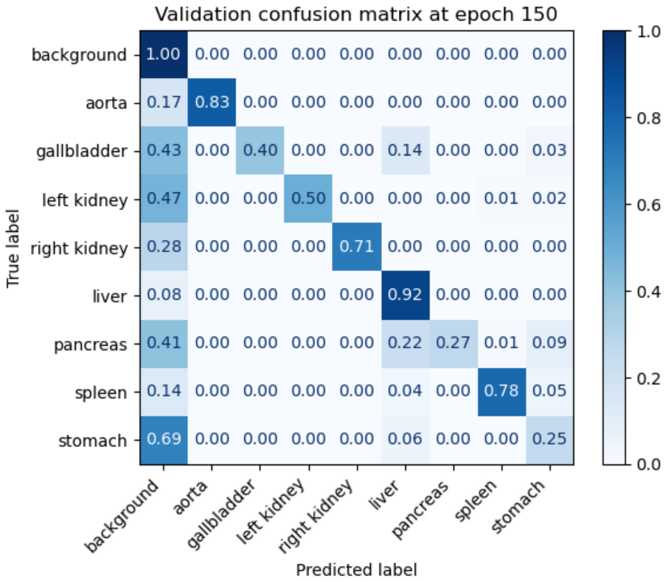
Experiments with downsampling methods

downsampling	acc	AUC	AP	DSC	IoU	HD95
maxpool	0.989	0.981	0.796	0.773	0.659	0.004
conv (channel change)	0.989	0.945	0.710	0.709	0.595	0.005
conv (no channel change)	0.988	0.945	0.710	0.709	0.599	0.004

Example prediction



original image | ground truth | prediction


Validation confusion matrix at epoch 150
True label
Predicted label
background
aorta
gallbladder
left kidney
right kidney
liver
pancreas
spleen
stomach
1.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
0.17 0.83 0.00 0.00 0.00 0.00 0.00 0.00 0.00
0.43 0.00 0.40 0.00 0.00 0.14 0.00 0.00 0.03
0.47 0.00 0.00 0.50 0.00 0.00 0.00 0.01 0.02
0.28 0.00 0.00 0.00 0.71 0.00 0.00 0.00 0.00
0.08 0.00 0.00 0.00 0.00 0.92 0.00 0.00 0.00
0.41 0.00 0.00 0.00 0.00 0.22 0.27 0.01 0.09
0.14 0.00 0.00 0.00 0.00 0.04 0.00 0.78 0.05
0.69 0.00 0.00 0.00 0.00 0.06 0.00 0.00 0.25
1.0
0.8
0.6
0.4
0.2
0.0

References

[1] O. Ronneberger, P. Fischer, T. Brox. U-Net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention*, November 2015, pp. 234–241.

[2] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[3] Hu Cao et al. Swin-Unet: Unet-like pure transformer for medical image segmentation. arXiv preprint, May 2021. arXiv:2105.05537v1 [eess.IV]

[4] O. Russakovsky et al. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision* **115**, 2015, pp. 211–252.

[5] O. Oktay et al. Attention U-Net: Learning where to look for the pancreas. arXiv preprint, April 2018. arXiv:1804.03999v3 [cs.CV]