Medical Image Segmentation

Lin Zhu(s232291), Cong Jin(s232254), Dong Yun(s232293), Yu Zhang(s230000)

DTU

Problem description

1. Design a generic segmentation architecture and apply to both **PH2** (skin lesion) and **DRIVE** (retinal blood vessel) datasets. Record and discuss the different performance measures.
2. Adding different **weak annotations** to the **PH2**(skin lesion) dataset, train a weakly supervised segmentation model, and evaluate the segmentation performance.

Abstract

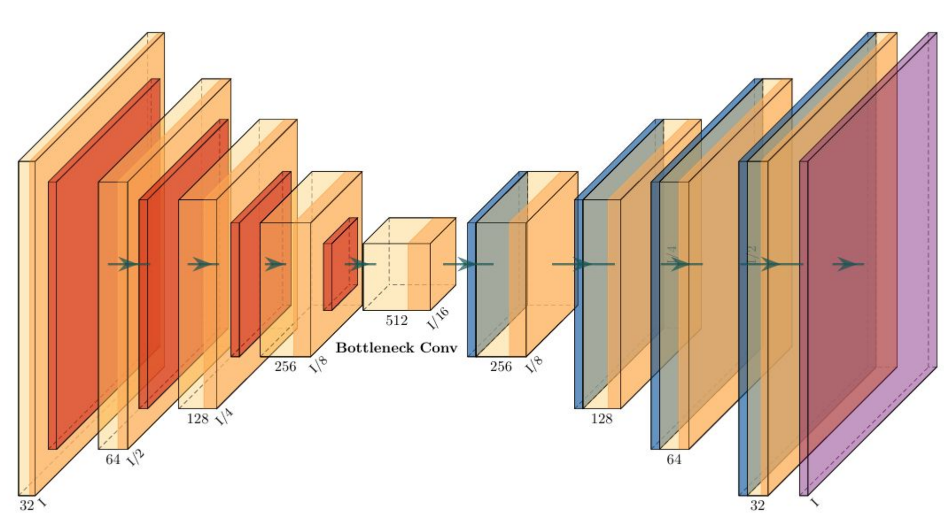
Result:
For DRIVE, after using data enhancement, the segmentation effect of blood vessels is much enhanced.
For PH2, the weakly supervised training effect using 8 click annotations can be close to the fully supervised training effect using mask. In addition, the effect of using box annotation alone for weak supervision is not better, which is consistent with what was mentioned in the paper.

(U-Net with focal loss)	DRIVE		PH2			
	with augmentation	without augmentation	with masks (50 epoches)	with 3 click annotations (50 epoches)	with 8 click annotations (50 epoches)	with box annotation (5 iteration 10 epoches each iteration)
performance metrics	validation / test	validation / test	validation / test	validation / test	validation / test	validation / test
Loss	**0.0179 / 0.0197**	0.0218 / 0.0220	**0.0320 / 0.2216**	0.0405 / 0.0705	**0.0556 / 0.0882**	0.8807 / 2.6025
Dice	**0.6988 / 0.6572**	0.6384 / 0.6023	**0.9295 / 0.8955**	0.8967 / 0.8608	**0.9078 / 0.8864**	0.8166 / 0.7413
IoU	**0.5370 / 0.4894**	0.4688 / 0.4309	**0.8684 / 0.8117**	0.8141 / 0.7585	**0.8323 / 0.7973**	0.6982 / 0.5990
Accuracy	**0.9575 / 0.9521**	0.9504 / 0.9464	**0.9615 / 0.8904**	0.9479 / 0.8561	**0.9518 / 0.8806**	0.9100 / 0.7606
Sensitivity	**0.5784 / 0.5105**	0.5139 / 0.4513	**0.9442 / 0.8358**	0.9376 / 0.7816	**0.9524 / 0.8268**	0.7327 / 0.6112
Specificity	**0.9928 / 0.9957**	0.9911 / 0.9953	**0.9682 / 0.9514**	0.9520 / 0.9658	**0.9514 / 0.9586**	0.9837 / 0.9803

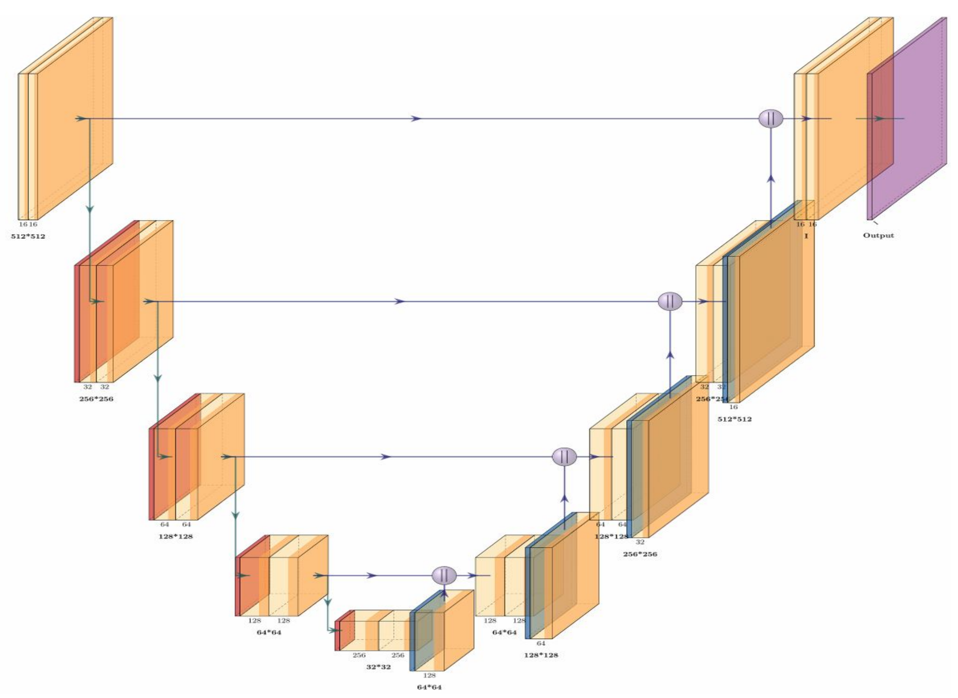
Dataset:
(1) DRIVE - train:val:test = 12:4:4 = 60%:20%:20% (without data augmentation)
- train:val:test = 48:4:4 (add augmentation to training set)

(2) PH2 - train:val:test = 140:30:30 = 75%:15%:15% (without data augmentation)

Architecture



EncDec



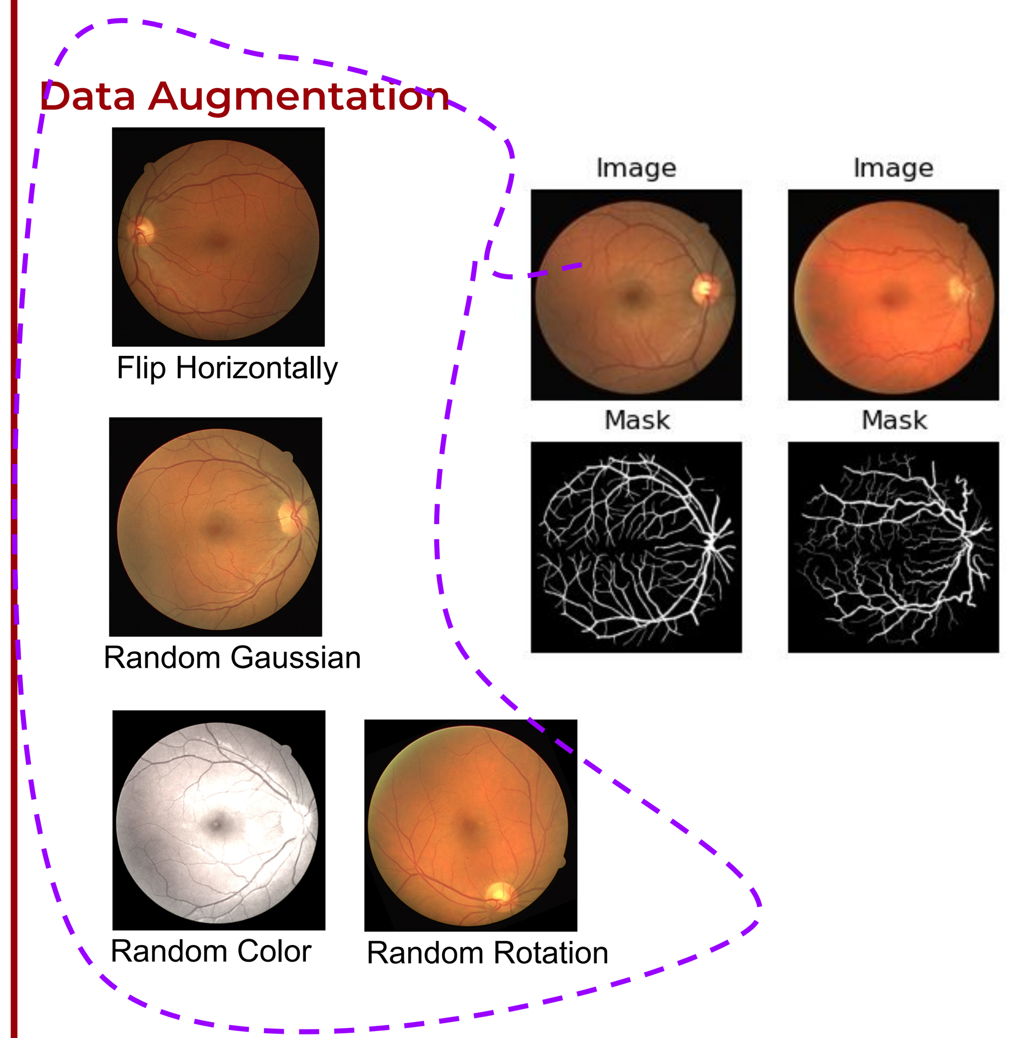
U-Net

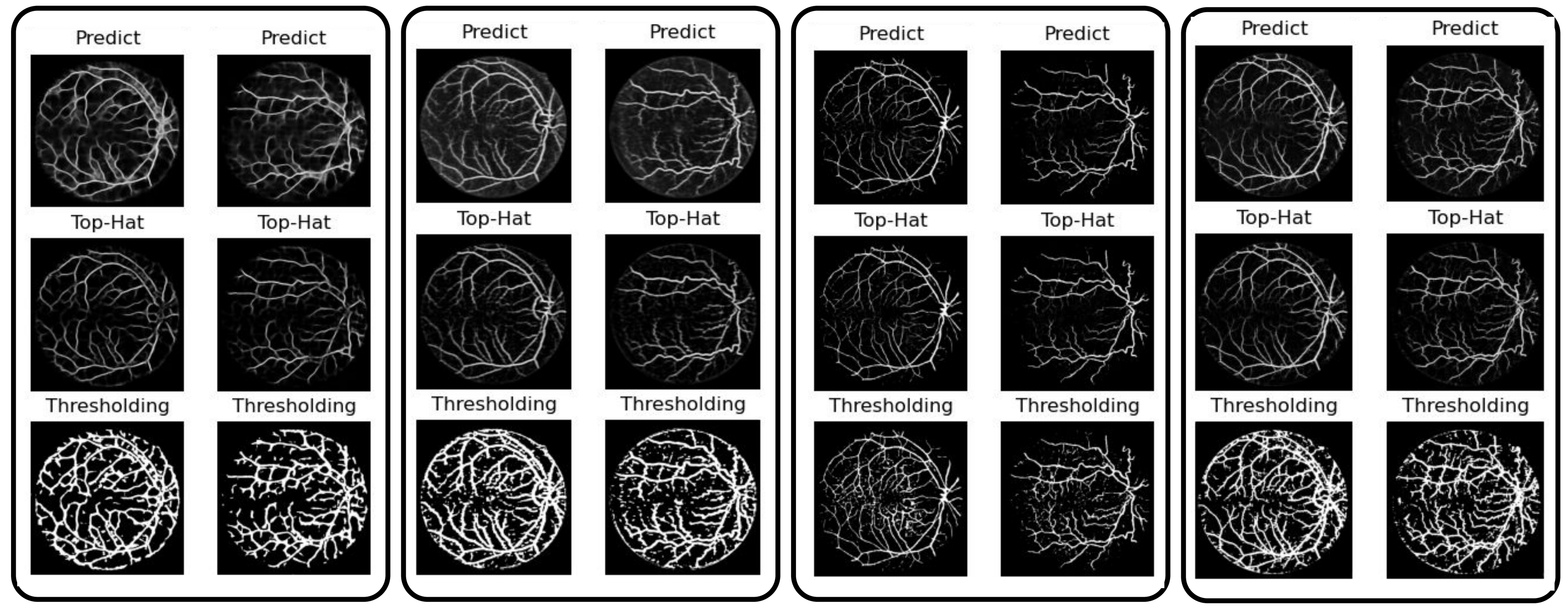
References

[1] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *arXiv preprint arXiv:1505.04597.*
[2] Guo, C., Szemenyei, M., Yi, Y., Wang, W., Chen, B., & Fan, C. (2020). SA-UNet: Spatial Attention U-Net for Retinal Vessel Segmentation. *arXiv preprint arXiv:2004.03696.*
[3]Papadopoulos, D. P., Uijlings, J. R., Keller, F., & Ferrari, V. (2017). *Extreme clicking for efficient object annotation.* In Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017.

Ablation study

For the sake of GPU load, we performed **data augmentation**, **different weak annotations** and **optimization loss functions** (cross entropy, focal_loss and weighted cross entropy) and the below results.





EncDec (550 epoches) **U-Net** (550 epoches) **U-Net** (augmented-550 epoches) **U-Net** (augmented-200 epoches)

Weak supervision & Annotation

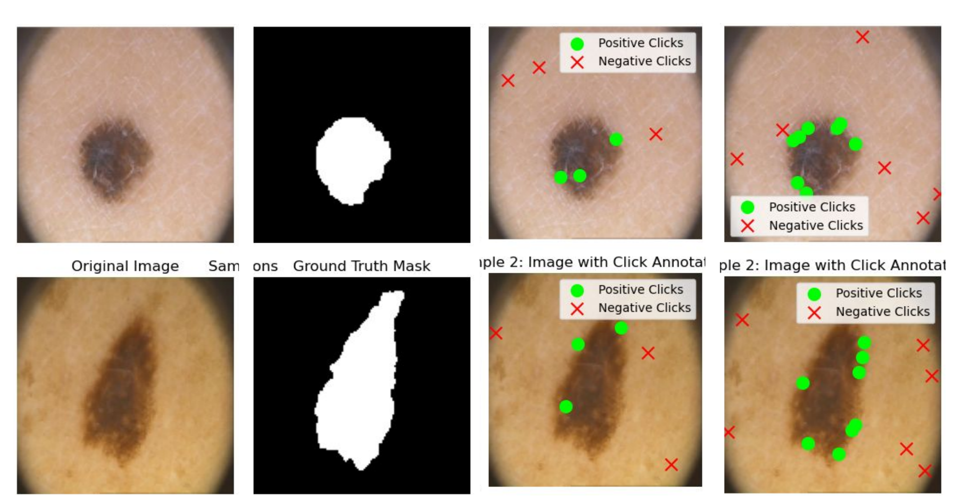
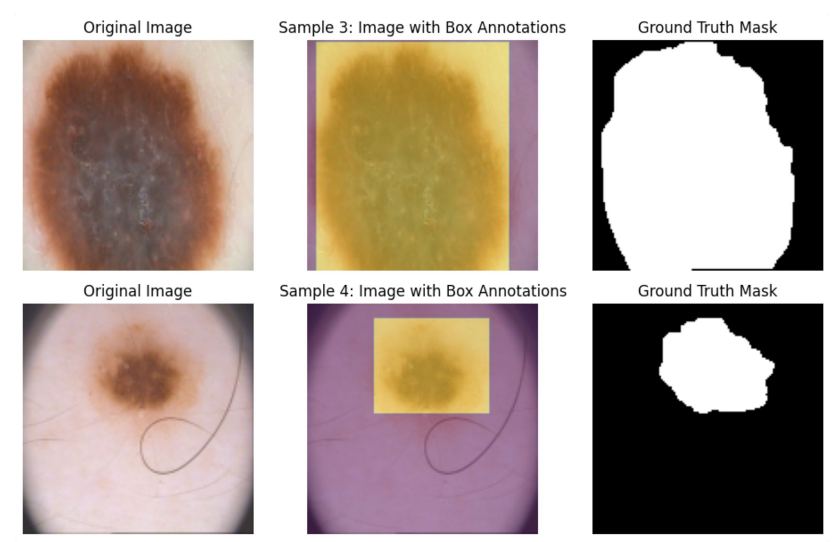
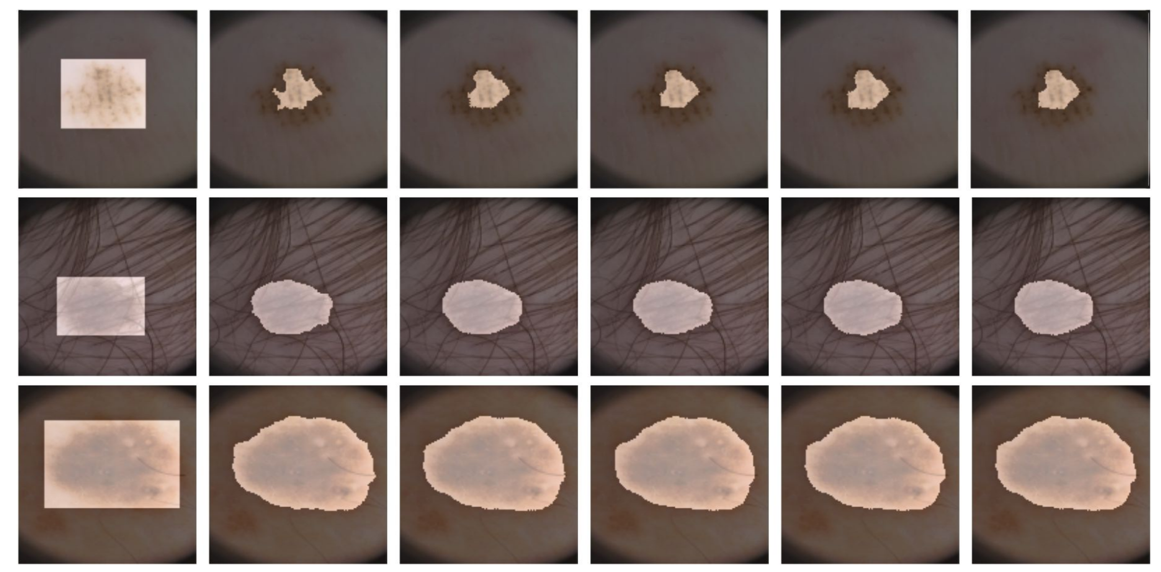


image mask 3 click annotation 8 click annotation





Start iteration 1 iteration 2 iteration 3 iteration 4 iteration 5

Loss functions

(U-Net)	retinal blood vessel		PH2 (3 click annotations)	
Loss functions	focal_loss	bce_loss	focal_loss	bce_loss
performance metrics	validation/test	validation/test	validation/test	validation/test
Loss	0.0179 / 0.0197	0.1149 / 0.1152	0.0405 / 0.0705	0.0886 / 0.3775
Dice	0.6988 / 0.6572	0.7536 / 0.7575	0.8967 / 0.8608	0.9004 / 0.8432
IoU	0.5370 / 0.4894	0.6047 / 0.6097	0.8141 / 0.7585	0.8192 / 0.7324
Accuracy	0.9575 / 0.9521	0.9601 / 0.9594	0.9479 / 0.8561	0.9465 / 0.8377
Sensitivity	0.5784 / 0.5105	0.7160 / 0.7050	0.9376 / 0.7816	0.9197 / 0.7490
Specificity	0.9928 / 0.9957	0.9828 / 0.9846	0.9520 / 0.9658	0.9559 / 0.9753