Seminar -1/12

Presenter： Wentao Lei

Paper List

- (CVPR2021)

  Every Annotation Counts: Multi-label Deep Supervision for Medical Image Segmentation

- (CVPR2021)

  clDice - a Novel Topology-Preserving Loss Function for Tubular Structure Segmentation

Every Annotation Counts: Multi-label Deep Supervision for Medical Image Segmentation

Simon Reiß[1,2] Constantin Seibold[1] Alexander Freytag[2] Erik Rodner[2,3] Rainer Stiefelhagen[1]
[1]Karlsruhe Institute of Technology [2]Carl Zeiss AG [3]University of Applied Sciences Berlin
{simon.reiss,constantin.seibold,rainer.stiefelhagen}@kit.edu

Background

任务：半监督学习辅助影像分割

传统的分割网络是用encoder-decoder的方式，在decoder 的最后一层输出 prediction heatmap，然后用这个输出的heatmap 和 mask 标签构建损失函数。但是作者认为这种方式是有缺陷的：用上采样的方式类似与无中生有，会引入很多额外的信息。而且这种方式生成的方式对于生成准确的分类和空间信息的作用存疑。具体来说，一个10 x 10 的特征图，要上采样回原图的 100 x 100，意味着每个像素要扩大一百倍，那么在这种情况下要保证分割的精确性，难度是比较大的。

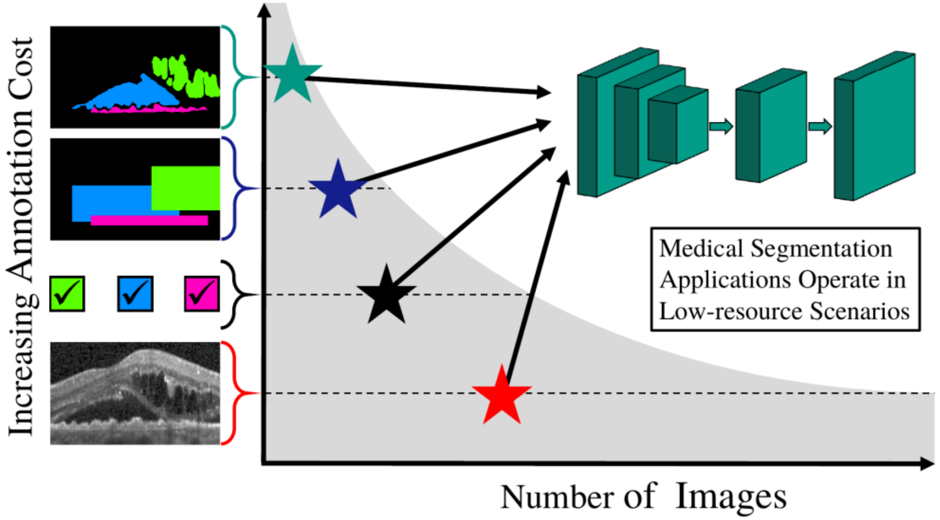


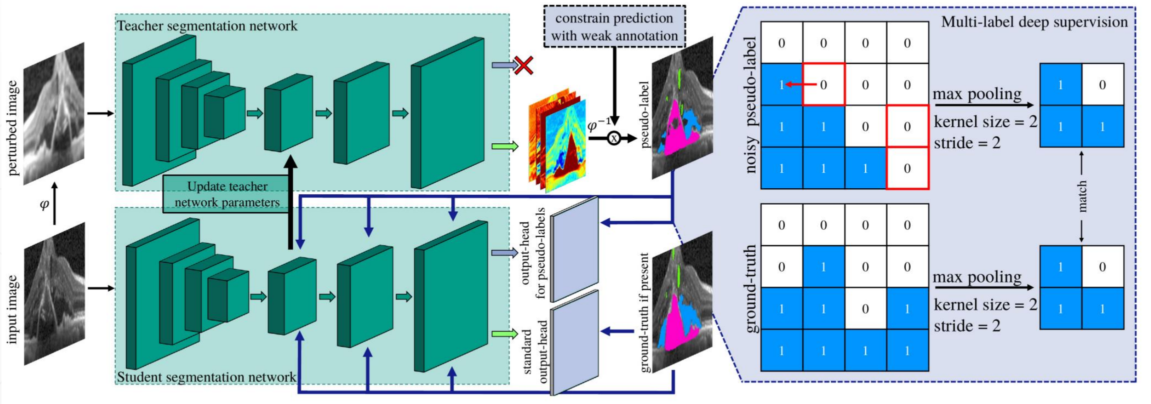
Figure 1. Annotations for segmentation are costly, especially when experts need to provide them. We show how our semi-weakly semantic segmentation method can use different annotation types and how the recognition performance benefits from them.

Method

核心思想：
把 mask 标签通过降采样方式到 decoder 的每一层进行监督学习。

作者认为与其将特征图逐步上采样到进行损失函数计算，不如将 mask 标签进行下采样到每一个特征图的大小，然后对 decoder 的每一个特征层做限制。

这种把 mask 标签进行下采样用于限制 feature map 的方案归结起来有两个好处：（1）从全量信息中做减法，总比无中生有的信息更加准确些。（2）在下采样过程中能够平滑一些mask 标签中的噪声，即使mask不是特别准确，只要轮廓相似，可能在 featuremap 尺寸比较小的时候也能得到比较好的监督。这为 boundingbox 标签的使用提供了便利条件。



Method

多标签信息（bounding box，类别信息等）的利用。本文中可以利用多种标签，除了 pixel-wise 的 mask，还可以利用 boundingbox-level 生成的 mask 和 image-level 的标签。boundingbox-level 生成 mask 的方式见下图：背景为 0，boundingbox 框住目标的位置为前景，里面的数值与类别的编号一致，所以与 pixel-level mask唯一不同的就是 boundingbox-level 生成的mask 边缘边缘是boundingbox 的边缘。

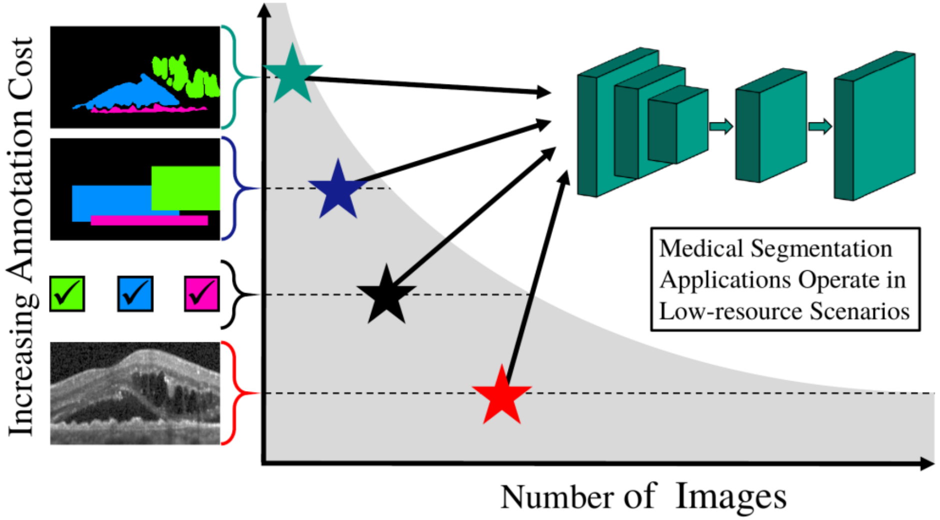
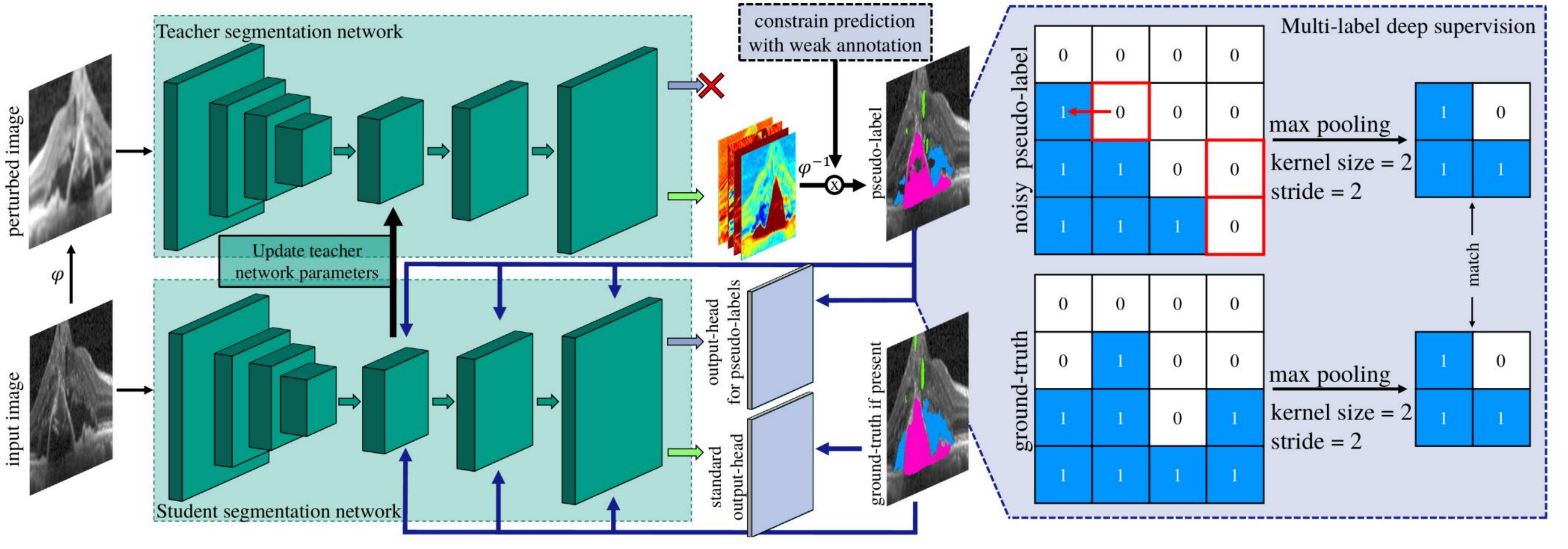


Figure 1. Annotations for segmentation are costly, especially when experts need to provide them. We show how our semi-weakly semantic segmentation method can use different annotation types and how the recognition performance benefits from them.

Method

自监督学习框架上，本文借鉴了 Mean-Teacher 的框架 [1]，利用了一组迭代的 student-teacher 的方式进行模型更新。

$$\theta_t^{\text{teacher}} = \alpha \cdot \theta_{t-1}^{\text{teacher}} + (1-\alpha) \cdot \theta_t^{\text{student}} \quad . \qquad (10)$$



Method

如图3所示，我们通过在feature map上应用output-heads，将这种包含语义、下采样的多标签ground-truth进行集成，并执行

$$\mathcal{L}_{BCE}(f,m) = -\frac{1}{\Omega_2}\sum_{i,j,c=1}^{H,W,C} BCE(\kappa(f)^{c,i,j}, m^{c,i,j})$$

$$\mathcal{L}(f_1,\ldots,f_h,m_1^*,\ldots,m_h^*) = \frac{1}{h}\sum_{k=1}^{h}\mathcal{L}_{BCE}(f_k,m_k^*). \quad (9)$$

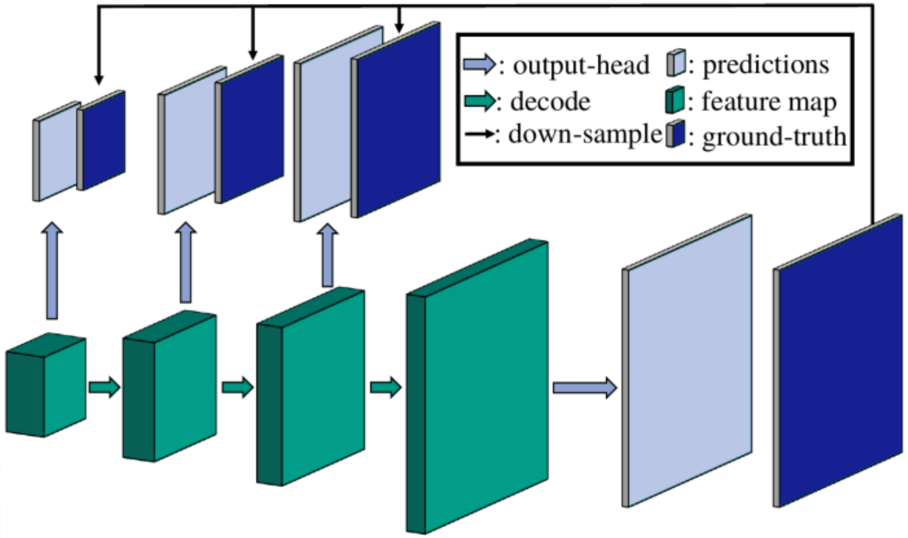


Figure 3. Our proposed method of integrating deep supervision into the decoder of segmentation networks by down-sampling the pixel-mask and enforcing a multi-label classification loss.

Experiment

本文的实验建立在公开可用的RETOUCH数据集上，用于视网膜流体分割。 它是包含不同视网膜疾病的光学相干断层扫描(OCT) 的集合。 这些数据通过不同供应商的成像工具获得，这些数据集有三种特征:Spectralis、Cirrus和Topcon。

一般来说，不同厂家的b型扫描在外观上是不同的，本文将在实验中单独考虑b型扫描。 此外，作为研究稀缺数据场景的工作的一部分，主要实验是在这三家公司中最小的Spectralis(每卷49个scan)上进行的，而对其余供应商(每卷128个scan)的性能进行评估，以强调方法的通用性。 该数据集有pixel-wise标签的三种类型的视网膜液:视网膜内液，视网膜下液和色素上皮脱离。 在实验中，从mask中获得边界框和图像级标签。

Experiment

Method	$\mathcal{U}$	6	12	24	Full Access
		Cirrus			
Baseline [50]		12.31 ± 5.41	19.43 ± 8.00	30.10 ± 9.34	48.92 ± 11.94
Multi-label Deep Supervision (Ours)		$\mathbf{15.99 \pm 6.87}$	$\mathbf{25.12 \pm 8.58}$	33.53 ± 9.44	50.47 ± 10.84
Perone and Cohen-Adad[10] [48]	✓	12.36 ± 6.12	24.99 ± 6.49	33.79 ± 10.15	49.75 ± 12.87
Mean-Taught Deep Supervision[10] (Ours)	✓	9.18 ± 8.53	23.33 ± 7.37	$\mathbf{35.82 \pm 11.40}$	$\mathbf{51.24 \pm 10.94}$
		Topcon			
Baseline [50]		14.79 ± 9.34	21.19 ± 11.57	27.61 ± 10.31	42.22 ± 10.42
Multi-label Deep Supervision (Ours)		$\mathbf{18.20 \pm 10.48}$	20.92 ± 13.02	33.71 ± 11.92	$\mathbf{45.85 \pm 10.32}$
Perone and Cohen-Adad[10] [48]	✓	15.26 ± 12.74	21.88 ± 12.48	27.67 ± 13.81	41.43 ± 8.18
Mean-Taught Deep Supervision[10] (Ours)	✓	14.39 ± 11.19	$\mathbf{23.92 \pm 15.25}$	$\mathbf{33.87 \pm 8.25}$	42.70 ± 10.97

Experiment

larger φ	inference	α	MSE	validation (mIoU)
✓	teacher	0.5	–	61.24 ± 3.69
✓	teacher	0.5	✓	$\mathbf{61.36} \pm \mathbf{4.73}$
✓	teacher	0.1	✓	60.15 ± 4.14
✓	student	0.1	✓	58.54 ± 3.62
–	student	0.1	✓	58.26 ± 4.27
–	student	0.0	–	57.80 ± 4.68

Table 1. Ablation for *Mean-Taught Deep Supervision* using 24 pixel-masks and the remaining image-level labels. Last line shows *Self-Taught Deep Supervision* performance for comparison.

Experiment

Method	$\mathcal{G}$	$\mathcal{U}$	3	6	12	24	Full Access
					Mask Supervision		
Baseline [50]			14.80 ± 6.50	26.98 ± 7.83	35.39 ± 6.36	48.63 ± 5.17	62.09 ± 4.77
Multi-Label Deep Supervision (Ours)			**17.98 ± 8.20**	**32.92 ± 7.35**	**42.96 ± 6.71**	**52.68 ± 6.82**	**65.82 ± 4.64**
IIC Baseline[8] [27]		✓	**22.45 ± 9.36**	32.02 ± 7.23	41.48 ± 7.26	53.08 ± 6.13	65.16 ± 3.80
Deeply Supervised IIC[8]		✓	20.78 ± 8.83	31.39 ± 10.26	39.18 ± 6.94	50.10 ± 7.92	65.18 ± 3.85
Perone and Cohen-Adad[10] [48]		✓	16.17 ± 10.74	33.10 ± 10.24	45.80 ± 7.51	54.75 ± 5.96	65.49 ± 4.14
Self-Taught Deep Supervision (Ours)		✓	10.37 ± 8.29	28.62 ± 12.96	43.57 ± 9.97	56.11 ± 6.30	66.24 ± 4.67
Mean-Taught Deep Supervision[10] (Ours)		✓	16.31 ± 15.48	**35.17 ± 11.35**	**53.52 ± 8.72**	**58.84 ± 6.57**	**66.31 ± 4.66**
MIL Baseline	✓		15.44 ± 11.10	25.46 ± 8.57	41.34 ± 9.66	49.07 ± 8.20	61.50 ± 5.64
Deeply Supervised MIL	✓		20.02 ± 9.17	31.50 ± 8.88	44.29 ± 5.03	51.13 ± 3.93	62.04 ± 3.92
Self-Taught Deep Supervision (Ours)	✓		20.47 ± 8.62	36.40 ± 8.91	49.39 ± 9.95	59.29 ± 7.52	66.34 ± 3.81
Mean-Taught Deep Supervision[10] (Ours)	✓		**21.91 ± 13.49**	**42.14 ± 14.25**	**54.70 ± 9.26**	**60.45 ± 5.71**	**66.39 ± 4.29**
					Bounding Box Supervision		
Baseline [50]			12.49 ± 4.28	18.32 ± 4.94	25.62 ± 3.08	29.55 ± 2.77	38.45 ± 4.44
Multi-Label Deep Supervision (Ours)			**14.59 ± 5.81**	**19.62 ± 6.21**	**27.89 ± 3.44**	**32.02 ± 4.78**	**38.66 ± 3.36**
IIC Baseline[8] [27]		✓	**15.40 ± 7.07**	18.15 ± 7.49	26.05 ± 6.00	30.07 ± 4.32	38.45 ± 4.65
Deeply Supervised IIC[8]		✓	12.77 ± 7.15	17.76 ± 6.26	**28.99 ± 4.60**	30.64 ± 3.05	38.81 ± 4.48
Perone and Cohen-Adad[10] [48]		✓	11.17 ± 7.41	**19.02 ± 8.46**	27.44 ± 5.81	31.72 ± 3.87	39.38 ± 3.56
Self-Taught Deep Supervision (Ours)		✓	5.14 ± 3.84	9.62 ± 7.35	24.47 ± 6.12	32.71 ± 3.56	**39.39 ± 3.63**
Mean-Taught Deep Supervision[10] (Ours)		✓	8.21 ± 3.96	14.28 ± 7.48	24.79 ± 5.79	**34.14 ± 3.10**	39.04 ± 4.15
MIL Baseline	✓		15.82 ± 6.55	16.95 ± 6.19	22.56 ± 4.56	26.48 ± 5.51	37.15 ± 4.06
Deeply Supervised MIL	✓		**17.14 ± 8.06**	20.18 ± 4.61	24.15 ± 4.95	29.12 ± 4.75	37.94 ± 3.35
Self-Taught Deep Supervision (Ours)	✓		16.04 ± 8.52	**22.15 ± 6.29**	28.63 ± 4.04	32.37 ± 3.75	**38.97 ± 3.59**
Mean-Taught Deep Supervision[10] (Ours)	✓		15.81 ± 8.59	21.97 ± 8.17	**29.83 ± 5.30**	**34.81 ± 3.62**	38.66 ± 4.73

clDice - a Novel Topology-Preserving Loss Function for Tubular Structure Segmentation

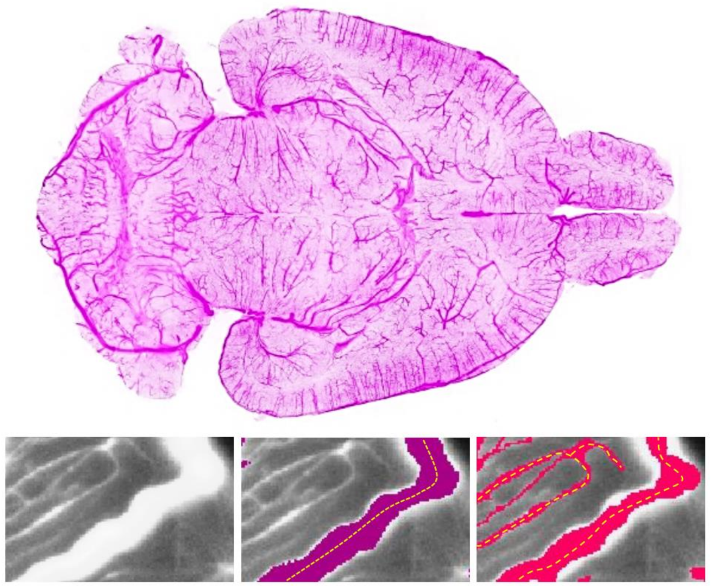
Suprosanna Shit *[1] Johannes C. Paetzold *[1] Anjany Sekuboyina[1] Ivan Ezhov[1]
Alexander Unger[1] Andrey Zhylka[2] Josien P. W. Pluim[2] Ulrich Bauer[1] Bjoern H. Menze[1]
[1]Technical University of Munich [2] Eindhoven University of Technology

Background

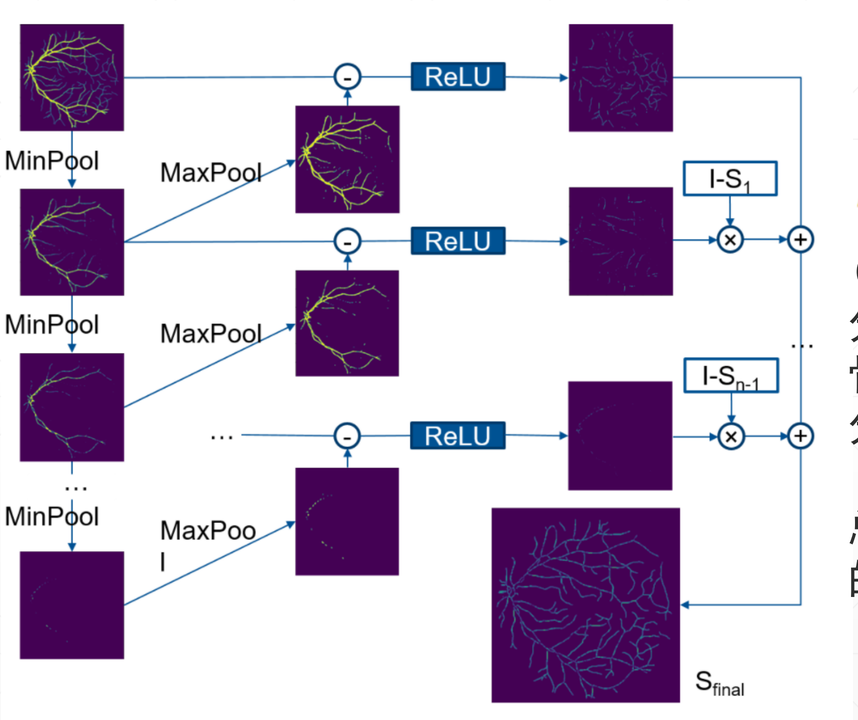
■ 管状、网状结构(如血管、神经元或道路)的精确分割与许多研究领域有关。

■ 对于这种结构，拓扑是其最重要的特征; 尤其是保持连通性:在血管网络的情况下，失去连接的血管会完全改变血流的动态。

■ 为了保证分割准确度的基础上，保持拓扑特性，本文引入了一种新的相似度度量方法，称为centerlineDice(简称clDice)，它是根据分割蒙版与其(形态学)骨架的交集计算的。本文证明了clDice在二维和三维分割中能保证拓扑等价。



Method



$$S \leftarrow S + (1 - S) \circ ReLU(I - I')$$

(1-S)是上图骨架中不存在的部分，和下图取交集就是得到上图骨架不存在，但是下图存在的部分

总体思路就是用下图逐步对上图的骨架做一个完善

Method

clDice Metric

在本文中，展示了如何将clDice用作管状结构的基准分割性能指标。 度量clDice是通过使用scikit-image库中的骨架，使用“硬”骨架来计算的。 其他潜在的更复杂的骨架技术也可以集成到clDice度量中。

$$\mathrm{clDice}(V_P, V_L) = 2 \times \frac{\mathrm{Tprec}(S_P, V_L) \times \mathrm{Tsens}(S_L, V_P)}{\mathrm{Tprec}(S_P, V_L) + \mathrm{Tsens}(S_L, V_P)} \tag{2}$$

$$\mathrm{Tprec}(S_P, V_L) = \frac{|S_P \cap V_L|}{|S_P|}; \quad \mathrm{Tsens}(S_L, V_P) = \frac{|S_L \cap V_P|}{|S_L|}$$

Method

clDice as a Loss function

为了用clDice训练神经网络，本文设计了一个损失函数。 为了稳定的原因，并确保良好的体积分割，本文将clDice与常规的Dice或二元交叉熵损失函数相结合。 此外，我们需要引入一个软骨架来使骨架化完全可微。

$$\mathrm{clDice}(V_P, V_L) = 2 \times \frac{\mathrm{Tprec}(S_P, V_L) \times \mathrm{Tsens}(S_L, V_P)}{\mathrm{Tprec}(S_P, V_L) + \mathrm{Tsens}(S_L, V_P)} \tag{2}$$

Method

为了使用clDice作为损失函数，本文引入了一种可微的软骨架化，其中迭代的最小和最大池化被用作形态侵蚀和扩张。

$$\mathrm{clDice}(V_P, V_L) = 2 \times \frac{\mathrm{Tprec}(S_P, V_L) \times \mathrm{Tsens}(S_L, V_P)}{\mathrm{Tprec}(S_P, V_L) + \mathrm{Tsens}(S_L, V_P)} \tag{2}$$

$$s = 1 - \frac{2|X \cap Y|}{|X| + |Y|}$$

由于我们的目标是在实现精确分割的同时保持拓扑，而不是学习骨架，我们将我们提出的soft-clDice与soft-Dice以以下方式结合:

$$\mathcal{L}_c = (1-\alpha)(1 - \mathit{softDice}) + \alpha(1 - \mathit{softclDice}) \tag{3}$$

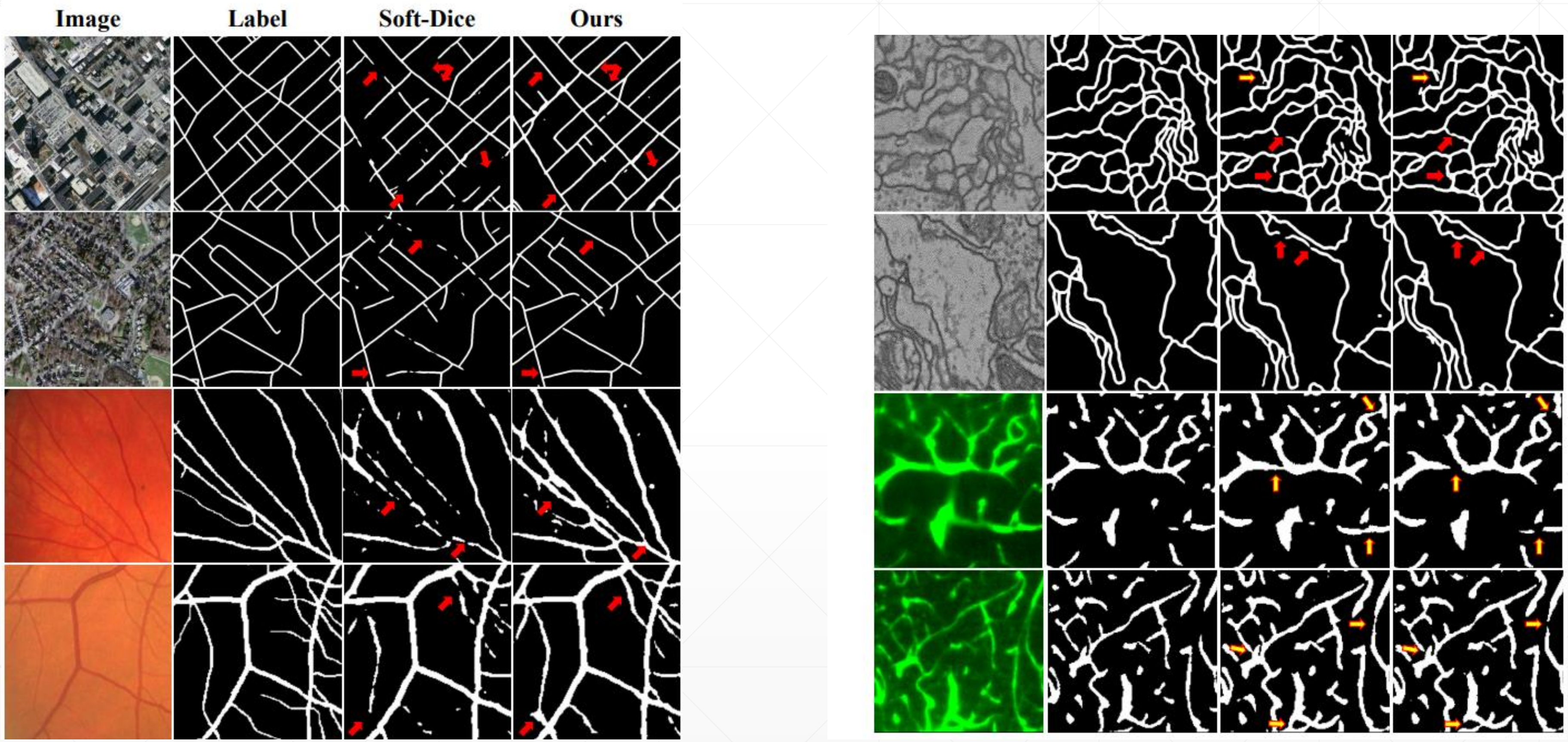
Experiment

Dataset	Network	Loss	Dice	Accuracy	*clDice*	β_0 Error	β_1 Error	SMD [4]	χ_{error}	Opt-J F1 [7]
Roads	FCN	*soft-dice*	64.84	95.16	70.79	1.474	1.408	0.1216	2.634	0.766
		$\mathcal{L}_c, \alpha = 0.1$	66.52	95.70	74.80	0.987	1.227	0.1002	2.625	0.768
		$\mathcal{L}_c, \alpha = 0.2$	**67.42**	**95.80**	76.25	**0.920**	1.280	**0.0954**	2.526	0.770
		$\mathcal{L}_c, \alpha = 0.3$	65.90	95.35	74.86	0.974	1.197	0.1003	2.448	0.775
		$\mathcal{L}_c, \alpha = 0.4$	67.18	95.46	**76.92**	0.934	**1.092**	0.0991	**2.183**	**0.803**
		$\mathcal{L}_c, \alpha = 0.5$	65.77	95.09	75.22	0.947	1.184	0.0991	2.361	0.782
	U-NET	*soft-dice*	76.23	96.75	86.83	0.491	1.256	0.0589	1.120	0.881
		$\mathcal{L}_c, \alpha = 0.1$	**76.66**	**96.77**	87.35	0.359	**0.938**	0.0457	0.980	0.878
		$\mathcal{L}_c, \alpha = 0.2$	76.25	96.76	87.29	**0.312**	1.031	**0.0415**	0.865	0.900
		$\mathcal{L}_c, \alpha = 0.3$	74.85	96.57	86.10	0.322	1.062	0.0504	0.827	0.913
		$\mathcal{L}_c, \alpha = 0.4$	75.38	96.60	86.16	0.344	1.016	0.0483	**0.755**	**0.916**
		$\mathcal{L}_c, \alpha = 0.5$	76.45	96.64	**88.17**	0.375	0.953	0.0527	1.080	0.894
	Mosinska et al.	[29, 17]	-	97.54	-	-	2.781	-	-	-
	Hu et al.	[17]	-	97.28	-	-	1.275	-	-	-
CREMI	U-NET	*soft-dice*	91.54	97.11	95.86	0.259	0.657	0.0461	1.087	0.904
		$\mathcal{L}_c, \alpha = 0.1$	91.76	**97.21**	96.05	0.222	0.556	**0.0395**	1.000	0.900
		$\mathcal{L}_c, \alpha = 0.2$	91.66	97.15	96.01	0.231	0.630	0.0419	0.991	0.902
		$\mathcal{L}_c, \alpha = 0.3$	**91.78**	97.18	**96.21**	**0.204**	**0.537**	0.0437	**0.919**	**0.913**
		$\mathcal{L}_c, \alpha = 0.4$	91.56	97.12	96.09	0.250	0.630	0.0444	0.995	0.902
		$\mathcal{L}_c, \alpha = 0.5$	91.66	97.16	96.16	0.231	0.620	0.0455	0.991	0.907
	Mosinska et al.	[29, 17]	82.30	94.67	-	-	1.973	-	-	-
	Hu et al.	[17]	-	94.56	-	-	1.113	-	-	-
DRIVE retina	FCN	*soft-Dice*	78.23	96.27	78.02	2.187	1.860	0.0429	3.275	0.773
		$\mathcal{L}_c, \alpha = 0.1$	78.36	96.25	79.02	2.100	1.610	0.0393	3.203	0.777
		$\mathcal{L}_c, \alpha = 0.2$	**78.75**	96.29	80.22	1.892	1.382	0.0383	2.895	0.793
		$\mathcal{L}_c, \alpha = 0.3$	78.29	96.20	80.28	1.888	**1.332**	**0.0318**	2.918	**0.798**
		$\mathcal{L}_c, \alpha = 0.4$	78.00	96.11	80.43	2.036	1.602	0.0423	3.141	0.764
		$\mathcal{L}_c, \alpha = 0.5$	77.76	96.04	**80.95**	**1.836**	1.408	0.0394	**2.848**	0.794
	U-Net	*soft-Dice*	74.25	95.63	75.71	1.745	1.455	0.0649	2.997	0.760
		$\mathcal{L}_c, \alpha = 0.5$	**75.21**	**95.82**	**76.86**	**1.538**	**1.389**	**0.0586**	**2.737**	**0.767**
	Mosinska et al.	[29, 17]	-	95.43	-	-	2.784	-	-	-
	Hu et al.	[17]	-	95.21	-	-	1.076	-	-	-

Experiment

Dataset	Network	Loss	Dice	Accuracy	*clDice*	β_0 Error	β_1 Error	SMD [4]	χ_{error}	Opt-J F1 [7]
Vessap data	FCN, 1 ch	*soft-dice*	85.21	**96.03**	90.88	3.385	4.458	0.00459	5.850	0.862
		$\mathcal{L}_c, \alpha = 0.5$	**85.44**	95.91	**91.32**	**2.292**	**3.677**	**0.00417**	**5.620**	**0.864**
	FCN, 2 ch	*soft-dice*	85.31	95.82	90.10	2.833	4.771	0.00629	6.080	0.849
		$\mathcal{L}_c, \alpha = 0.1$	85.96	95.99	91.02	2.896	**4.156**	0.00447	5.980	0.860
		$\mathcal{L}_c, \alpha = 0.2$	**86.45**	**96.11**	91.22	2.656	4.385	0.00466	5.530	0.869
		$\mathcal{L}_c, \alpha = 0.3$	85.72	95.93	91.20	2.719	4.469	**0.00423**	5.470	0.866
		$\mathcal{L}_c, \alpha = 0.4$	85.65	95.95	**91.65**	2.719	4.469	**0.00423**	5.670	0.869
		$\mathcal{L}_c, \alpha = 0.5$	85.28	95.76	91.22	**2.615**	4.615	0.00433	**5.320**	**0.870**
	U-Net, 1 ch	*soft-dice*	87.46	96.35	91.18	3.094	5.042	0.00549	5.300	0.863
		$\mathcal{L}_c, \alpha = 0.5$	**87.82**	**96.52**	**93.03**	**2.656**	**4.615**	**0.00533**	**4.910**	**0.872**
	U-Net, 2 ch	*soft-dice*	87.98	96.56	90.16	2.344	4.323	0.00507	5.550	0.855
		$\mathcal{L}_c, \alpha = 0.1$	88.13	96.59	91.12	2.302	4.490	0.00465	5.180	**0.872**
		$\mathcal{L}_c, \alpha = 0.2$	87.96	96.74	92.52	2.208	**3.979**	0.00342	**4.830**	0.861
		$\mathcal{L}_c, \alpha = 0.3$	87.70	96.71	92.56	**2.115**	4.521	**0.00309**	5.260	0.858
		$\mathcal{L}_c, \alpha = 0.4$	**88.57**	**96.87**	**93.25**	2.281	4.302	0.00327	5.370	0.868
		$\mathcal{L}_c, \alpha = 0.5$	88.14	96.74	92.75	2.135	4.125	0.00328	5.390	0.864

Experiment



Qualitative results

Thanks for listening!