#218

SEMI-SUPERVISED MEDICAL IMAGE SEGMENTATION VIA QUERY DISTRIBUTION CONSISTENCY

Rong Wu, Dehua Li, Cong Zhang
DesicionLinnc Dev Group

Introduction

Inspired by MaxQuery and KMaX-Deeplab, we introduce a novel **D**ual-**K**Max **UX**-Net (DKUXNet) for semi-supervised medical image segmentation. In our work, we leverage these strengths by adopting the general design of 3D UX-Net and kMax-decoder as our backbone meta-architecture.

Main contributions:
(1) Our model divides images into 3 categories: background, organ, and tumor and updates the distance of the cluster center. Its performance is similar to SOTA fully supervised models, only utilizing 20% training data.
(2) We utilizes the consistency loss of query distribution and segmentation outputs to enhance image consistency.

Methods

The multi-scale outputs from each stage in the encoder are connected to a ConvNet-based decoder via skip connections. Specifically, we extract the outputs for stage i (i = 1, 2, 3) in the encoder and further deliver the outputs into kMax decoder block for cluster center information learning.

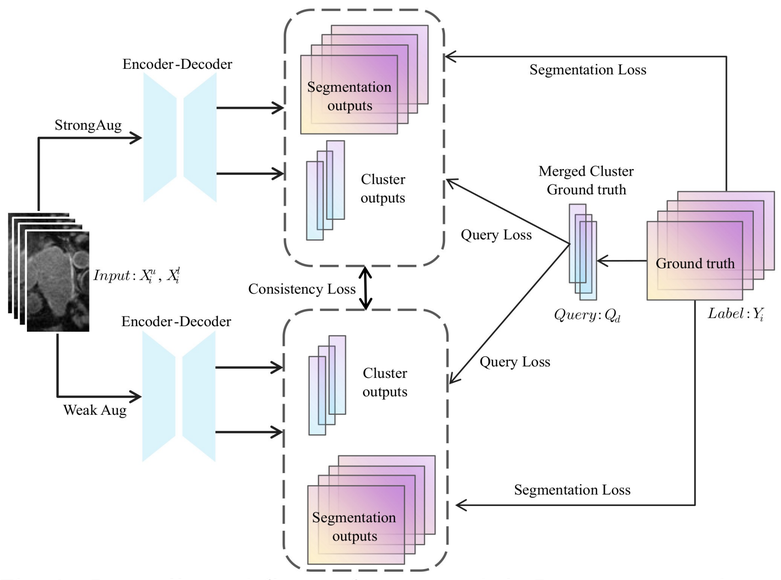

Fig 1. Overall workflow of our model. Our proposed dual KMax-based contrastive learning strategy

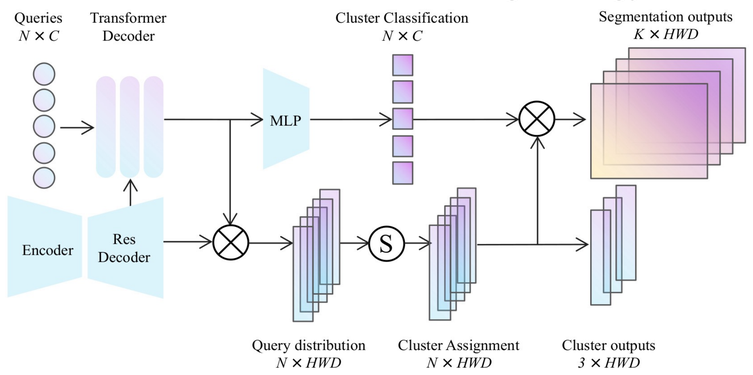

Fig 2. Decoder structure contains a pixel decoder, a transformer decoder and post-process modules.

Cluster Center Update:

$$\mathbf{C} = \mathbf{C} + \arg\max_N(\mathbf{Q}^c(\mathbf{K}^p)^T)\mathbf{V}^p$$

$$\mathbf{A} = \arg\max_N(\mathbf{C} \times \mathbf{P}^T),$$

$$\hat{\mathbf{C}} = \mathbf{A} \times \mathbf{P},$$

where $\mathbf{C} \in \mathbb{R}^{N \times C}$, $\mathbf{P} \in \mathbb{R}^{HWD \times C}$, and $\mathbf{A} \in \mathbb{R}^{HWD \times N}$ refers to cluster center, pixel features and cluster assignments.

Loss:

$$\mathcal{L}_{segc} = -\frac{1}{HWD} \log \frac{\exp(sim(X_i, X_j)/\tau)}{\sum_{k \neq i} \exp(sim(X_i, X_k)/\tau)},$$

$$\mathcal{L}_{qdc} = -\frac{1}{HWD} \log \frac{\exp(sim(Q_i, Q_j)/\tau)}{\sum_{k \neq i} \exp(sim(Q_i, Q_k)/\tau)},$$

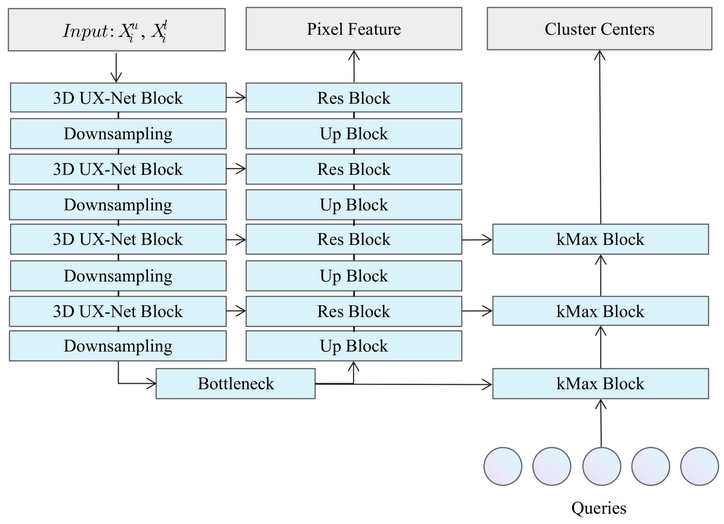

Fig 3. ConNeXt blocks with kMaX decoders

The ConvNeXt backbone includes transformer encoders to enhance the pixel features, and upsampling layers to generate higher-resolution features. We use four 3D UX-Net blocks and four Downsampling blocks as the depth-wise convolution encoder.

Experiments

To evaluate the proposed method, we apply our algorithm on Left Atrial (LA) dataset from the 2018 Atrial Segmentation Challenge. We use 80 scans for training and 20 scans for validation. All scans are centered at the heart region cropped accordingly, and then normalized to zero mean and unit variance.

In this work, we report the performance of all methods trained with 5%/10%/20% labeled images. The raw LA training data for each case are randomly cropped to 112*112*80 voxels. Results are evaluated on four metrics: Dice, Jaccard Index, 95% Hausdorff Distance (95HD), and Average Surface Distance (ASD). To ensure a fair comparison, we perform all experiments on the same machine and report the mean results from the final iteration.

Results

We compared DKUXNet with other baselines on LA. Our framework shows optimal performance in most metrics. Specifically, with only 5% labeled data, DKUXNet achieved 85.96% of the dice score. DKUXNet also achieved a 90.41% dice score with only 10% labeled data. When the labeled data volume increased to 20%, the results obtained by this model were comparable to those of V-Net trained in 100% labeled data, and compared to the 90.98% score of the upper bound model, Dice scored 91.70%.

Components			Metrics			
$\mathcal{L}_{seg}$	$\mathcal{L}_{qdc}$	$\mathcal{L}_{segc}$	Dice↑	Jaccard↑	95HD↓	ASD↓
✓			77.73	64.42	16.86	3.92
✓	✓		82.12	70.16	23.63	5.99
✓		✓	85.52	75.49	14.90	3.73
✓	✓	✓	90.41	82.69	7.32	1.71

Table 1. Ablation Studies

We conducted ablation experiments to verify the effectiveness of the key components in our proposed model. To investigate the individual impact of different tasks, we first only use labeled images for training and analyze how the dual-task consistency performs when only labeled images are used. As shown in Table 2, dual-contrastive loss substantially improves segmentation performance when labeled data is limited.

Our key idea is that cluster assignment should be considered for semi-supervised learning. However, the performance of the proposed method is not as outstanding at 5% labeled setting, we should further develop novel consistency loss for information transfer between unlabeled data and label data and test on more complicated segmentation datasets.

Method	Labeled Scans	Metrics			
		Dice↑	Jaccard↑	95HD↓	ASD↓
V-Net	80	90.98	83.61	8.58	2.10
ours		92.62	86.32	4.62	1.28
SASSNet [29]	4(5%)	78.07	65.03	29.17	8.63
DTC [30]		79.61	67.00	25.54	7.20
MC-Net [31]		80.14	67.88	24.08	7.18
URPC [11]		80.92	68.90	17.25	2.76
SS-Net [32]		80.75	68.54	19.81	4.98
MC-Net+ [33]		83.33	71.79	15.70	4.33
CAML [13]		**87.34**	**77.65**	**9.76**	**2.49**
ours		85.96	75.91	11.72	2.64
SASSNet [29]	8(10%)	85.71	75.35	14.74	4.00
DTC [30]		84.55	73.91	13.80	3.69
MC-Net [31]		86.87	78.49	11.17	2.18
URPC [11]		83.37	71.99	17.91	4.41
SS-Net [32]		86.56	76.61	12.76	3.02
MC-Net+ [33]		87.68	78.27	10.35	1.85
CAML [13]		89.62	81.28	8.76	2.02
ours		**90.41**	**82.69**	**7.32**	**1.71**
SASSNet [29]	16(20%)	88.11	79.08	12.31	3.27
DTC [30]		87.79	78.52	10.29	2.50
MC-Net [31]		90.43	82.69	6.52	1.66
URPC [11]		87.68	78.36	14.39	3.52
SS-Net [32]		88.19	79.21	8.12	2.20
MC-Net+ [33]		90.60	82.93	6.27	1.58
CAML [13]		90.78	83.19	6.11	1.68
ours		**91.70**	**84.82**	**5.81**	**1.62**

Table 2. Comparison with SOTA methods on LA dataset.

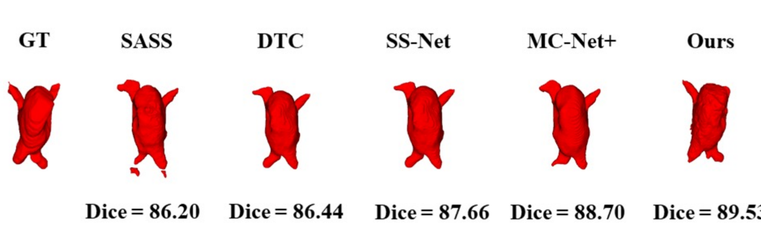

Fig 4. 3D Visualization of different ablation studies for LA segmentation. GT: ground truth.

ISBI 2024
ATHENS, GREECE

DecisionLinnc 决策链