Cross-grained Contrastive Representation for Unsupervised Lesion Segmentation in Medical Images

Ziqi Yu †, Botao Zhao †, Yipin Zhang, Shengjie Zhang, Xiang Chen, Haibo Yang, Tingying Peng and Xiao-Yong Zhang*

Fudan University, Zhejiang Lab, Helmholtz AI

Background

- Segmentation for disease progression, pre- and post-operative treatment planning
- Labeling 3D MRI/CT volumes is time-consuming and requires specialized medical knowledge.
- Most of unsupervised segmentation are reconstruction-based methods.
- Instance-level contrastive loss does not take into account the pathological characteristics of lesions in medical images.

Method

1. Overall architecture (see Fig. 1)
2. Foreground-Background Determination
3. Lesion-Normal Tissue Discrimination
   a) Coarse-grained Discrimination: Cross entropy loss.
   b) Fine-grained Discrimination:

positive pair $\Omega^+ = \{\widetilde{z}_i(\widetilde{x}_i) | \forall C(\widetilde{x}_i) \in C(r_i^{lesion})\}$

negative ones $\Omega^- = \Omega \backslash \Omega^+$

$$\mathcal{L}_{cl}^{fg} = \sum_{i=1}^{N} \frac{-1}{|\Omega^+|} \sum_{x^+ \in \Omega^+} \log \frac{\exp(CL^+/\tau)}{\sum_{x_i \in \Omega'} \exp(CL/\tau)}$$

$$CL^+ = sim(\widetilde{z}_{x_i}, \widetilde{z}_{x^+}), CL = sim(\widetilde{z}_{x_i}, \widetilde{z}_x)$$

$$sim(x, y) = \frac{xy^{\mathrm{T}}}{\|x\| + \|y\|}$$

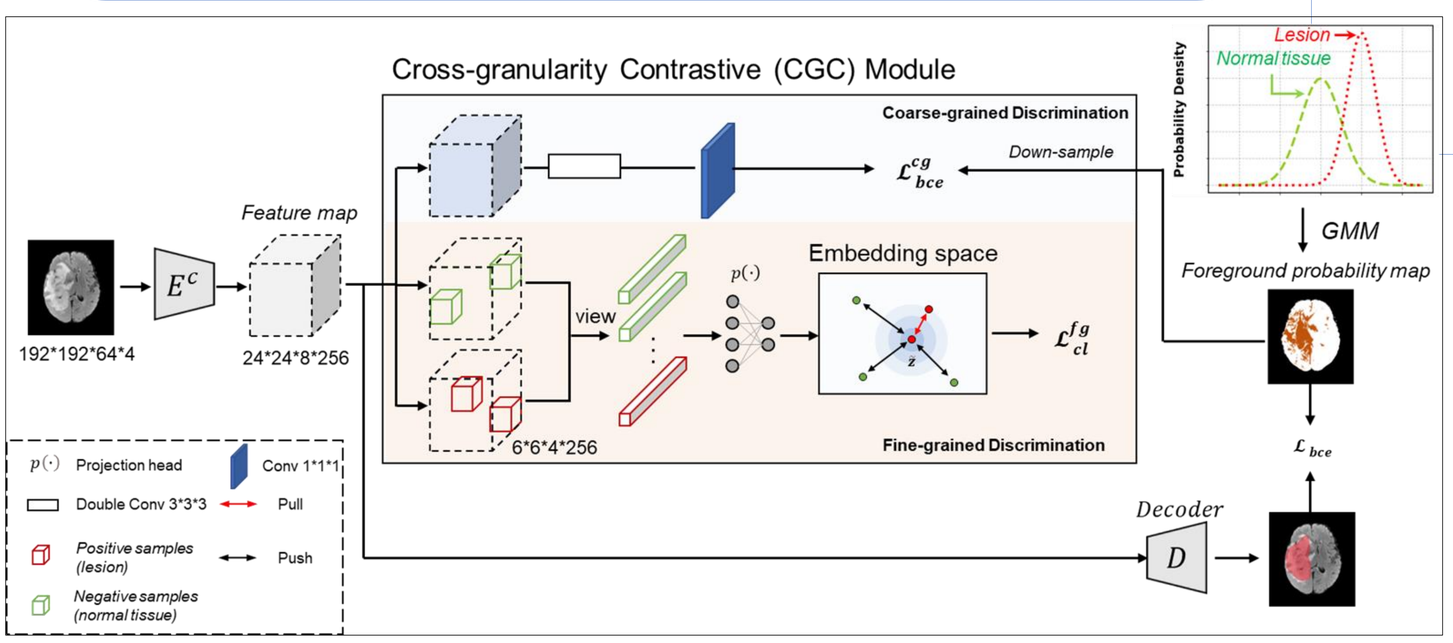

Figure 1. Schematic of our module.

Table 1. Comparison of brain tumor segmentation performance on BraTS and CQ500 dataset.

Methods	BraTS dataset		CQ500 dataset	
	Dice	Sensitivity	Dice	Sensitivity
Full supervision	0.9134± 0.1026	0.9136±0.1150	0.7962±0.1523	0.7840±0.2138
AE	0.3543±0.2462	0.4542±0.2401	0.3381±0.2369	0.5142±0.2616
Context VAE [3]	0.4261±0.1874	0.4371±0.2596	0.3969±0.2446	0.5348±0.2212
GMVAE [23]	0.4418±0.1726	0.5374±0.2127	0.4147±0.2184	0.5362±0.2449
f-AnoGAN [24]	0.4835±0.1675	0.5332±0.2446	0.4024±0.2172	0.5528±0.1882
Bayesian VAE [4]	0.5348±0.1618	0.5575±0.2375	0.4391±0.2474	0.5446±0.2076
AnoVAEGAN [25]	0.5184±0.1560	0.5737±0.2098	0.4467±0.2286	0.5649±0.1844
AMCons [15]	0.7362±0.1642	0.7684±0.2084	0.4741±0.2310	0.4588±0.2469
Mumford-Shah [26]	0.7156±0.1881	0.7063±0.2157	0.5206±0.1937	0.5087±0.2336
Ours (P from [26])	0.7743±0.1365	0.7576±0.1974	0.5569±0.1861	0.5348±0.1875
GMM w/ threshold	0.7585±0.2091	0.7965±0.1892	0.6490±0.1975	0.5535±0.1485
Ours w/o CGC	0.7929±0.1877	0.8013±0.1815	0.6625±0.1945	0.6427±0.2013
Ours (P from GMM)	**0.8405±0.1323**	**0.8178±0.1756**	**0.6993±0.1755**	**0.6768±0.1825**

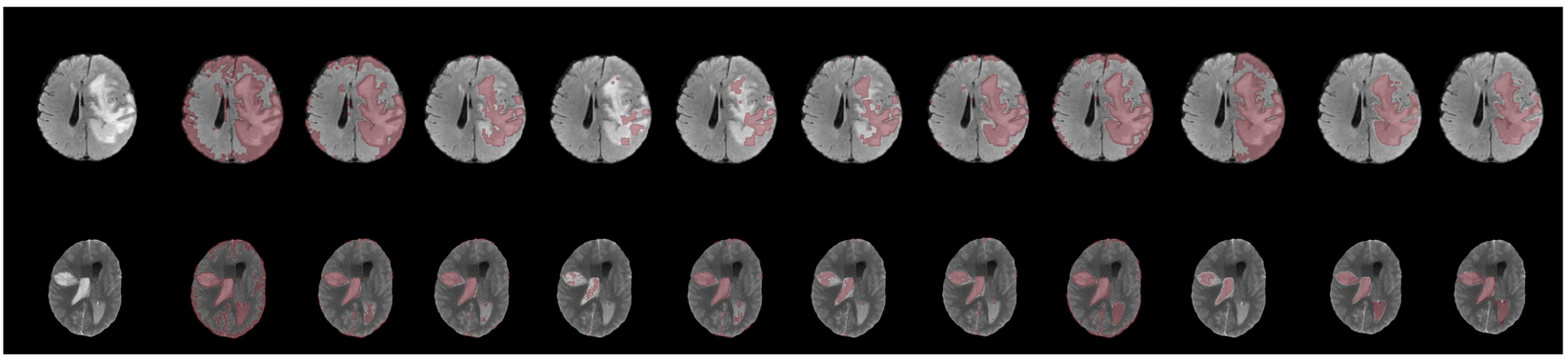

Figure 2. Visualization of exemplary segmentation results on BraTS and CQ500 dataset.

Table 2. Ablation study for each module and mini-path size on the BraTS dataset.

Methods	BraTS dataset	
	Dice	ASSD [mm]
Full supervision (backbone)	0.9134± 0.1028	1.1581±1.2894
Full supervision (nnUNet)	0.9161± 0.0873	1.0257±1.1316
P from GMM w/ threshold	0.7585±0.2091	2.1472±1.9426
Backbone + P	0.7929±0.1877	1.2330±1.1821
Backbone + P + Global granularity	0.8164±0.1631	1.3392±1.6271
Backbone + P + Local granularity	0.8248±0.1548	1.2513±1.2484
Ours ($\widetilde{z} : \mathbb{R}^{12\times12\times4}$)	0.8392±0.1389	1.2367±1.4205
Ours ($\widetilde{z} : \mathbb{R}^{6\times6\times4}$)	**0.8405±0.1323**	**1.1756±1.1439**
Ours ($\widetilde{z} : \mathbb{R}^{3\times3\times4}$)	0.8388±0.1410	1.1943±1.4263

Table 3. Ablation study of temperature on the BraTS dataset.

Temperature (τ)	0.07	0.1	0.2	0.3	0.7	0.9
Dice	0.8254	0.8375	**0.8405**	0.8392	0.8248	0.8211
ASSD [mm]	1.4744	1.1809	1.1756	**1.1550**	1.2665	1.1871

Results

Dataset:

- BraTS 2018 (MRI scans: T1, T1ce, T2, T2-F)
- CQ500 (CT scans): Preprocessed by Brain Extraction Net (BEN) [1]

Results:

1. Our proposed method achieves the Dice scores of 84.05% and 69.93% on BraTS and CQ500 datasets
2. Our method is not bound to GMM-initiated segmentation

Ablation Studies:

1. Effectiveness of Each Module (see Table 2)
2. Effectiveness of Temperature-calibrated Logits (see Table 3)

Conclusion

To summarize, we have presented a new unsupervised framework for medical image segmentation using a novel cross-granularity contrastive module. Our module contains coarse-grained and fine-grained discrimination paths, enabling the network to capture the distinctions between lesions and normal tissues at different levels of context. We evaluate our method on two large public datasets of CT/MRI scans and demonstrate that our approach improves a Gaussian mixture model-based segmentation by up to 9%, which surpasses all other unsupervised segmentation methods by a large margin. Additionally, our module can also be combined with other existing unsupervised segmentation methods to further enhance their performance.

[1] "BEN: a generalizable Brain Extraction Net for multimodal MRI data from rodents, nonhuman primates, and humans." eLife (2022).

Speaker: Ziqi Yu
E-mail: zqyu19@fudan.edu.cn