CUTS: A Deep Learning and Topological Framework for Multigranular Unsupervised Medical Image Segmentation

Chen Liu[1], Matthew Amodio[1], Liangbo L. Shen[2], Feng Gao[1], Arman Avesta[1], Sanjay Aneja[1], Jay C. Wang[1], Lucian V. Del Priore[1], and Smita Krishnaswamy[1⋆]

[1] Yale University
[2] University of California, San Francisco

Abstract. Segmenting medical images is critical to facilitating both patient diagnoses and quantitative research. A major limiting factor is the lack of labeled data, as obtaining expert annotations for each new set of imaging data and task can be labor intensive and inconsistent among annotators. We present CUTS, an unsupervised deep learning framework for medical image segmentation. CUTS operates in two stages. For each image, it produces an embedding map via intra-image contrastive learning and local patch reconstruction. Then, these embeddings are partitioned at dynamic granularity levels that correspond to the data topology. CUTS yields a series of coarse-to-fine-grained segmentations that highlight features at various granularities. We applied CUTS to retinal fundus images and two types of brain MRI images to delineate structures and patterns at different scales. When evaluated against predefined anatomical masks, CUTS improved the dice coefficient and Hausdorff distance by at least 10% compared to existing unsupervised methods. Finally, CUTS showed performance on par with Segment Anything Models (SAM, MedSAM, SAM-Med2D) pre-trained on gigantic labeled datasets. Code is available at `https://github.com/KrishnaswamyLab/CUTS`.

Keywords: Unsupervised learning · Medical image segmentation.

1 Introduction

Medical image segmentation plays an increasingly crucial role in both research and clinical settings in a wide range of imaging modalities, including microscopy, X-ray, ultrasound, optical coherence tomography (OCT), computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), and others [13]. With high-quality image segmentation, clinicians can diagnose and monitor disease progression more easily to improve patient care. Traditional

⋆ C. Liu and M. Amodio are joint first authors. S. Aneja, J. C. Wang, L. V. Del Priore and S. Krishnaswamy are co-supervisory authors. Please direct correspondence to: `smita.krishnaswamy@yale.edu` or `lucian.delpriore@yale.edu`.

medical image segmentation methods rely on hand-crafted features [12, 29] or predefined atlases [18]. These methods are gradually being replaced by deep learning [15] as supervised neural networks demonstrate superior performance compared to feature-based methods and less overhead than atlas-based methods. Although supervised neural networks have been widely successful in image segmentation in recent years, there are several problems in applying them to medical images, particularly to make clinical inferences. First, these networks are dependent on expert annotations, so they require a large number of labels to adequately cover the variance of the data to produce reliable segmentations [11]. Second, supervised networks trained on one set of annotated images may not be able to generalize to similar images collected in very slightly different contexts, such as in different patient populations or on different devices [1]. Third, the desired segmentation granularity may vary across use cases even if the exact same image is concerned; for example, localizing a brain tumor would require finer segmentation compared to measuring the brain volume, yet this need is not easily accommodated by supervised approaches without updating the labels.

To address these issues, we propose an unsupervised framework that combines recent advances in representation learning with the frontiers of data geometry and topology. First, a convolutional patch encoder learns an embedding space from the image patches. Then, the learned embedding space serves as a feature-rich foundation for multiscale topological data coarse-graining. It not only circumvents the need for costly expert annotations and alleviates cross-domain generalization problem, but also produces multigranular segmentations which can potentially target multiple regions of interest without supervision.

2 Related Works

Before the introduction of contrastive learning, most unsupervised learning methods for medical image segmentation either learn from pseudo-labels generated by traditional methods [26] or perform registration onto an atlas [30]. Contrastive learning opens up many opportunities in this field. However, commonly used contrastive learning methods such as SimCLR [6] and SimSiam [7] focus on extracting image-level representations with an inter-image contrastive objective. These methods undermine intra-image features and are therefore unsuitable for tasks that require closer scrutiny within the same image, such as image segmentation. To adapt contrastive learning for image segmentation, [5] and [31] learn image and patch representations through global and local contrastive training. Both methods include a supervised fine-tuning stage, which still depends on labels. DFC [19] and STEGO [14] are two leading unsupervised segmentation methods that leverage contrastive learning concepts. Although STEGO can be trained without labels, it relies on pre-trained vision backbones for knowledge distillation, which is not a requirement in our method. DFC is by far the most similar to our approach, but with two key differences. First, DFC contrasts on pixels, while we operate on pixel-centered patches, which is semantically and

textually richer. Second, we perform multiscale coarse-graining that produces many segmentation maps at various granularities.

3 Methods

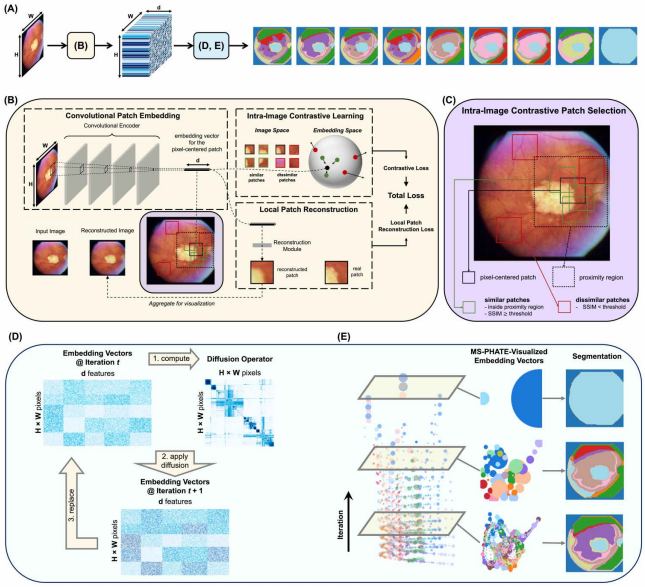

Fig. 1. The CUTS Framework. **(A)** Overview. **(B)** Pixel-centered patches are mapped into the embedding space, jointly optimized by two objectives. **(C)** Positive and negative patch pairs are selected based on proximity and structural similarity. **(D)** Diffusion condensation coarse grains embedding vectors at a series of granularities. **(E)** Segmentation for any granularity can be performed by mapping cluster assignments to the image space. Multiscale PHATE (MS-PHATE) [22] is used for visualization.

The CUTS framework contains two stages (Fig. 1(A)). In the first stage, it encodes each pixel along with the local neighborhood around it, denoted as a "pixel-centered patch", into a high-dimensional embedding space by jointly optimizing contrastive learning and autoencoding objectives (Fig. 1(B)). Unlike most contrastive learning methods that learn from augmented versions of full images, CUTS learns from regions within the same image (Fig. 1(C)). This emphasizes learning of local, intra-image features instead of invariance over known image transformations or noise models. This is especially critical for medical images, since they are globally homogeneous (i.e., images from different participants capture the same body part) yet locally heterogeneous (i.e., nuances in structures or textures within small areas of the image are essential). In the second stage, these embedding vectors are coarse-grained to many levels of granularity

by diffusion condensation [3, 21]. Metastable granularities can be automatically identified from the condensation homology as granularities with zero topological activity [21]. Segmentation is performed by assigning labels to pixels that correspond to clusters arising from a particular metastable granularity (Fig. 1(D-E)).

Learning an embedding space for pixel-centered patches CUTS uses a convolutional neural network as a patch encoder to map pixel-centered patches from the image space to a latent embedding space. It has convolution, batch norm, activation but no pooling – to ensure identical spatial dimension between the image and feature map. Two objectives are jointly optimized.

Intra-image contrastive loss For any anchor patch $\mathcal{P}_{ij} \in \mathbb{R}^{p\times p\times c}$ centered at coordinates (i, j), we sample positive patches $\{\mathcal{P}_{ij}^{+}\}$ and negative patches $\{\mathcal{P}_{ij}^{-}\}$. Let f denote the convolutional encoder. Anchor embedding $z_{ij} = f(\mathcal{P}_{ij})$, positive embeddings $\Omega^{+} := \{z_{ij}^{+}\} = \{f(\mathcal{P}_{ij}^{+})\}$, and negative embeddings $\Omega^{-} := \{z_{ij}^{-}\} = \{f(\mathcal{P}_{ij}^{-})\}$. After projecting the patches to the latent embedding space, we can perform contrastive learning on their respective embedding vectors z_{ij}^{+} and z_{ij}^{-}. We mine these positive and negative patches using a combination of a proximity heuristic and an image similarity metric. Only patches nearby (within $\pm$ one patch size) and structurally similar (SSIM [16] > 0.5) to the anchor patch are considered positive patches. The contrastive loss is defined as:

$$l_{contrast} = -\log \frac{\text{pos}}{\text{neg}}, \quad \text{pos} = \sum_{z_{ij}^{+}\in\Omega^{+}} e^{sim(z_{ij},z_{ij}^{+})/\tau}, \quad \text{neg} = \sum_{z_{ij}^{-}\in\Omega^{-}} e^{sim(z_{ij},z_{ij}^{-})/\tau}$$

Local patch reconstruction loss In addition to the contrastive loss, we ensure that our embedding of each pixel-centered patch retains information about the patch around it through a reconstruction loss. For an embedding $z_{ij} \in \mathbb{R}^{d}$, the patch reconstruction loss is $l_{recon} = ||\mathcal{P}_{ij} - f_{recon}(z_{ij})||_2^2$, where $f_{recon}(\cdot) : \mathbb{R}^{d} \to \mathbb{R}^{p\times p\times c}$ is a patch reconstruction module. In implementation, $f_{recon}(\cdot)$ is a two-layered fully-connected network with ReLU activation.

Final objective function The final objective function balances the two losses with a weighting coefficient $\lambda \in [0, 1]$. $loss = \lambda \cdot l_{contrast} + (1 - \lambda) \cdot l_{recon}$.

Hyperparameters Hyperparameters (patch size, number of patches, contrastive loss coefficient) are empirically selected. See Supplementary Materials.

Coarse-graining for multiscale segmentation For each image patch $\mathcal{P}_{ij}$ centered at coordinates (i, j), the patch encoder encodes it to $z_{ij} \in \mathbb{R}^{d}$. We can assign them to n different clusters $\{c_1, c_2, ..., c_n\}$ using a clustering algorithm $cls(\cdot) : \mathbb{R}^{d} \to \mathbb{R}$. Then, we can create a label map $L \in \mathbb{R}^{H\times W}$ where $L_{ij} = cls(z_{ij})$. The label map L will be the end product of CUTS segmentation. Notably, with diffusion condensation, $cls(\cdot)$ changes throughout the process, and therefore we can generate a rich set of labels. Diffusion condensation [3, 21] is

a dynamic process that sweeps through various levels of granularities to identify natural groupings of data. It iteratively condenses data points towards their neighbors through a diffusion process, at a rate defined by the diffusion probability between the points. Unlike most clustering methods, diffusion condensation constructs a full hierarchy of coarse-to-fine granularities where the number of clusters at each granularity is not arbitrarily set but rather inferred from the underlying structure of the data.

We can identify the segments that occur consistently over the series of segmentations, called persistent structures. This can be achieved by rank-ordering different segments based on their persistence levels, which is quantified by the number of consecutive diffusion iterations in which the segment stays intact and refrains from being merged into another segment.

For binary segmentation, we need to convert the multi-class label maps to binary segmentation masks. Following standard practices [14, 24], we use the ground truth segmentation mask to provide a hint on how to select the foreground for each image. Specifically, we iterate over each foreground pixel in the ground truth mask and find the most frequently associated cluster of the corresponding embedding vector. Then we set all pixels whose embeddings match that cluster label as the foreground. This process effectively finds the most probable cluster label if a pixel is randomly selected from the foreground region of the ground truth and thus is objective and unbiased.

4 Experiments

We prepared three medical image datasets to evaluate our proposed framework. The datasets are chosen to demonstrate the breadth of applications, as they cover variation in color channels (e.g., RGB versus intensity-only), imaging sequences (e.g., T1 versus T2 FLAIR), and organs of interest (e.g., eye versus brain).
Retinal fundus images We used retinal color fundus images of eyes with Geographic Atrophy (GA) in the age-related eye disease study group [10, 28]. GA regions were segmented by two graders and reviewed by a retinal specialist, resulting in 56 retinal images with accurate segmentations.
Brain MRI images (ventricles) We used MRIs of patients from the Alzheimer's Disease Neuroimaging Initiative study [9]. A radiologist manually segmented the brain ventricles on 100 T1-weighted brain MRIs for our study.
Brain MRI images (tumor) We used MRIs of patients with glioma that were scanned by several healthcare facilities. Tumor regions of 200 fluid-attenuated inversion recovery (FLAIR) brain MRIs are segmented by trained medical students and finalized by a board-certified attending neuroradiologist.

Qualitative results on multigranular segmentation As shown in Fig. 2, our multiscale segmentation method provides delineation of image structures at various granularities. The diffusion condensation process starts when all pixels are isolated from each other (pure noise, not shown in the figure). After a few iterations, fine-grained structures begin to emerge, as the most similar pixels are

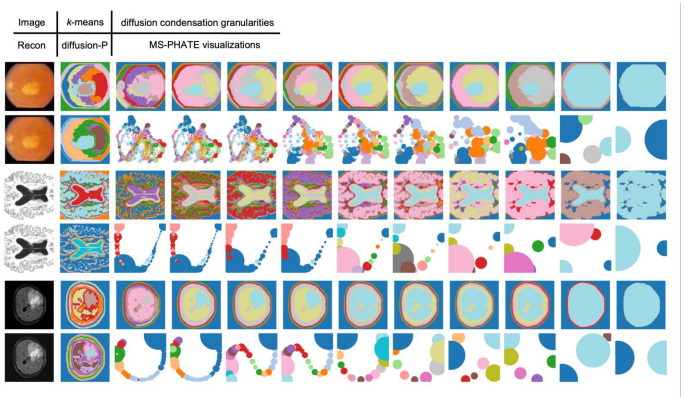

Fig. 2. Multigranular segmentation (odd rows) captures distinctive patterns at various scales. Multiscale PHATE (even rows) is used to visualize the diffusion condensation process. The results of CUTS + spectral k-means clustering ("k-means") and CUTS + diffusion condensation persistent structures ("diffusion-P") are also shown for reference.

clustered together (leftmost columns starting from the third column). On these finest scales, even the smallest structures are delineated, such as the retinal vessels in the retinal images (first row). Moving toward the coarser scales, anatomical structures arise as tiny patterns collectively form larger groups. Signature structures include the optic disc and geographic atrophy in the retinal images (first row), white and gray matter in the brain ventricles images (third row), and tumor region in the brain tumor images (fifth row). Detection of these anatomical structures can facilitate automatic measurements of their sizes, shapes, and locations for clinical interventions. On the coarser side of the spectrum, most structures are iteratively merged through diffusion condensation, leaving only the most distinctive objects in the image. The final resolution (rightmost column) identifies the two remaining clusters which correspond to the foreground and background, respectively.

Qualitatively, we show that CUTS is able to automatically detect meaningful structures and patterns at multiple granularities within medical images of various modalities. It enables users to determine their desired level of detail without the necessity of manually annotating data for the model's training.

Qualitative and quantitative results on binary segmentation We compared the performance of CUTS on the three datasets with several alternative methods. We first compared it with three traditional unsupervised methods: Otsu's watershed [29], Felzenszwalb [12], and SLIC [2]. We then compared with DFC [19] and STEGO [14], two recent unsupervised models based on deep learning. For each experiment, we re-trained DFC, STEGO, and CUTS on the images only. Next, we compared against the Segment Anything Model (SAM) [20, 24] which was pre-trained on 11 million images and 1.1 billion masks, as well as its

medical-image variants (MedSAM [23]/SAM-Med2D [8]) pre-trained on 1.6/4.6 million images and 1.6/19.7 million masks, respectively. For SAM variants, we provided a center point of the ground truth label as a prompt for segmentation of each image [24]. Lastly, for reference, we benchmarked a random labeler as the performance lower bound and two fully supervised methods, UNet [27] and nnUNet [17], as the upper bound. For coarse-graining of the pixel embeddings, we also implemented a spectral k-means clustering [32] alternative, which segments at only one granularity level. For a fair comparison, we applied the same binarization approach described in the Methods section to *all unsupervised methods*.

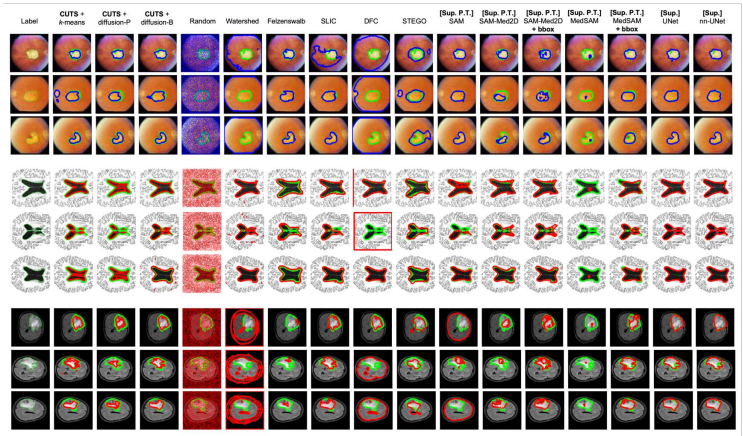

Fig. 3. Qualitative segmentation comparison. Green curves outline the ground truth labels while blue or red curves outline the predictions. "diffusion-B": the best diffusion condensation granularity. "**Sup.**": supervised "**P.T.**": pre-training. "**+bbox**": using bounding box as input; included for completeness but would be unfair for comparison.

Geographic atrophy segmentation in retinal fundus images Our first experiment aims to segment regions of geographic atrophy (GA) in retinal fundus images. GA is an advanced stage of age-related macular degeneration (AMD) characterized by progressive macula degeneration. CUTS accurately selects the region of atrophy. Qualitatively, CUTS is better at delineating the boundaries of atrophy compared to all other unsupervised methods (Fig. 3). The quantitative results (Table 1) also confirmed this observation. CUTS created better segmentations than other unsupervised methods, as indicated by a higher dice score and a lower Hausdorff distance.

Ventricle segmentation in brain MRI images In our next experiment, we tried to segment the brain ventricles in MRI images of patients at various stages of Alzheimer's disease. This task is considered clinically important because the volume of the brain ventricles can predict the progression of dementia [4, 25].

Table 1. Quantitative comparisons from 3 random seeds. Among unsupervised methods, the best is **bolded** and runner-up is underscored. §Using bounding box instead of point as input; included for completeness but would be unfair for comparison. ‡Diffusion condensation will not run since #features = 1 for each pixel in single-channel images.

	Deep learning?	Topological?	Retinal Atrophy DSC ↑	Retinal Atrophy HD ↓	Brain Ventricles DSC ↑	Brain Ventricles HD ↓	Brain Tumor DSC ↑	Brain Tumor HD ↓
Unsupervised, without learning								
Watershed (IEEE TPAMI'91 [29])	✗	✗	0.192±0.000	56.32±0.00	0.781±0.000	30.25±0.00	0.073±0.000	95.42±0.00
Felzenszwalb (IJCV'04 [12])	✗	✗	0.592±0.000	27.60±0.00	0.759±0.000	44.80±0.00	0.316±0.000	**21.41**±0.00
SLIC (IEEE TPAMI'12 [2])	✗	✗	0.567±0.000	28.76±0.00	0.475±0.000	37.96±0.00	0.242±0.000	47.51±0.00
Unsupervised, with learning								
DFC (IEEE TIP'20 [19])	✓	✗	0.300±0.020	46.47±1.42	0.631±0.024	34.28±0.57	0.197±0.004	52.51±0.09
STEGO (ICLR'22 [14])	✓	✗	0.649±0.025	34.12±4.06	0.725±0.050	12.59±4.43	0.176±0.104	57.16±14.09
(Ours) CUTS + Spectral k-means	✓	✗	0.675±0.014	26.82±0.88	0.774±0.008	8.31±0.23	0.432±0.010	33.94±0.65
(Ours) CUTS + Diffusion (pers.)	✓	✓	0.604±0.003	21.69±0.44	0.495±0.002	13.36±0.60	0.390±0.004	33.66±0.24
(Ours) CUTS + Diffusion (best)	✓	✓	**0.741**±0.007	**17.76**±0.13	**0.810**±0.006	**7.17**±0.18	**0.486**±0.007	25.16±1.12
Ablation: image pixels instead of latent embeddings								
Image pixels + Spectral k-means	✗	✗	0.560±0.000	37.97±0.00	0.386±0.000	26.11±0.00	0.240±0.000	51.69±0.00
Image pixels + Diffusion (pers.)	✗	✓	0.405±0.000	61.67±0.00	‡	‡	‡	‡
Image pixels + Diffusion (best)	✗	✓	0.538±0.000	45.16±0.00	‡	‡	‡	‡
Lower bound: random label								
Random	✗	✗	0.132±0.000	78.45±0.07	0.149±0.000	61.40±0.02	0.057±0.000	95.53±0.02
Upper bound: supervised								
SAM (ICCV'23 [20], MedIA'23 [24])	✓	✗	0.924±0.000	9.18±0.01	0.644±0.003	30.24±0.19	0.405±0.000	36.14±0.14
SAM-Med2D (ArXiv [8])	✓	✗	0.548±0.001	14.69±0.00	0.736±0.000	17.38±0.02	0.591±0.001	12.93±0.01
SAM-Med2D+bbox§	✓	✗	0.882±0.000	5.31±0.00	0.849±0.000	9.78±0.00	0.686±0.000	8.74±0.00
MedSAM (Nat. Commun.'24 [23])	✓	✗	0.079±0.000	32.29±0.02	0.053±0.000	64.00±0.04	0.088±0.001	33.54±0.02
MedSAM+bbox§	✓	✗	0.889±0.000	5.21±0.00	0.829±0.000	10.60±0.00	0.702±0.000	7.61±0.00
UNet (MICCAI'15 [27])	✓	✗	0.965±0.014	3.78±1.08	0.989±0.001	1.05±0.10	0.867±0.016	8.84±1.10
nnUNet (Nat. Methods'21 [17])	✓	✗	0.937±0.014	6.00±1.35	0.984±0.005	2.10±0.42	0.834±0.024	8.64±1.60

Qualitatively, CUTS delineated the brain ventricles in a wide variety of settings (Fig. 3). Due to the general trend that ventricles appear consistently darker than the rest of the image, most methods are able to achieve good overall performance on several cases. However, our method usually delineates the boundaries better than competing methods, especially for images showing noncontiguous ventricles. The quantitative results (Table 1) also indicate the superior performance of CUTS over other unsupervised methods.

Tumor segmentation in brain MRI images Our final experiment investigated a different segmentation target in brain MRI images – brain tumors, or more specifically, glioma. Accurate segmentation of tumor areas is crucial for the diagnosis and treatment of brain tumors. This process can help radiologists provide vital details about the size, position, and form of tumors, which is important to determine the most appropriate course of clinical care. Qualitatively, our method demonstrated superior segmentation compared to other unsupervised methods, as shown in Fig. 3. As a general observation, competing methods struggle to identify tumors, although they manage to segment the ventricles in a similar imaging modality. This disparity in performance was anticipated, given the pronounced complexity associated with tumor segmentation compared to ventricles, due to considerably more subtle contrast and morphological distinctions. Nevertheless, CUTS overcomes the inherent challenges and successfully segments tumor regions. Quantitatively, CUTS led the other unsupervised methods by a larger margin compared to the less demanding task of ventricle segmentation.

More impressively, CUTS achieved better results than every SAM variant on most datasets under fair comparison without relying on billions of annotations.

Ablation study We confirmed that applying diffusion condensation or spectral k-means on the raw image pixels is suboptimal compared to CUTS (Table 1).

5 Conclusion

CUTS is a deep learning and topological framework that identifies important medical image structures with self-supervision. Despite the emergence of foundation models, such as variants of SAM, CUTS remains relevant and insightful. It is lightweight and does not require extensive annotation and pre-training in large compute warehouses. Additionally, it is clear that foundation models necessitate domain-specific fine-tuning for tasks not covered by the initial supervised pre-training, which highlights the relevance of approaches like CUTS that investigate objectives and techniques to inject the correct inductive biases. Therefore, CUTS offers a practical alternative in the evolving landscape of medical imaging.

Acknowledgements This work was supported in part by NSF DMS 2327211, NSF Career 2047856, NIH 1R01GM130847-01A1, and NIH 1R01GM135929-01.

Disclosure of Interests The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Abràmoff, M.D., Lavin, P.T., Birch, M., Shah, N., Folk, J.C.: Pivotal trial of an autonomous ai-based diagnostic system for detection of diabetic retinopathy in primary care offices. NPJ digital medicine **1**(1), 1–8 (2018)
2. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: Slic superpixels compared to state-of-the-art superpixel methods. IEEE transactions on pattern analysis and machine intelligence **34**(11), 2274–2282 (2012)
3. Brugnone, N., Gonopolskiy, A., Moyle, M.W., Kuchroo, M., van Dijk, D., Moon, K.R., Colon-Ramos, D., Wolf, G., Hirn, M.J., Krishnaswamy, S.: Coarse graining of data via inhomogeneous diffusion condensation. In: 2019 IEEE International Conference on Big Data (Big Data). pp. 2624–2633. IEEE (2019)
4. Carmichael, O.T., Kuller, L.H., Lopez, O.L., Thompson, P.M., Dutton, R.A., Lu, A., Lee, S.E., Lee, J.Y., Aizenstein, H.J., Meltzer, C.C., et al.: Ventricular volume and dementia progression in the cardiovascular health study. Neurobiology of aging **28**(3), 389–397 (2007)
5. Chaitanya, K., Erdil, E., Karani, N., Konukoglu, E.: Contrastive learning of global and local features for medical image segmentation with limited annotations (2020)
6. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations (2020)
7. Chen, X., He, K.: Exploring simple siamese representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15750–15758 (2021)
8. Cheng, J., Ye, J., Deng, Z., Chen, J., Li, T., Wang, H., Su, Y., Huang, Z., Chen, J., Jiang, L., et al.: Sam-med2d. arXiv preprint arXiv:2308.16184 (2023)
9. Crawford, K.L., Neu, S.C., Toga, A.W.: The image and data archive at the laboratory of neuro imaging. Neuroimage **124**, 1080–1083 (2016)
10. Davis, M.D., Gangnon, R.E., Lee, L.Y., Hubbard, L.D., Klein, B., Klein, R., Ferris, F.L., Bressler, S.B., Milton, R.C., et al.: The age-related eye disease study severity scale for age-related macular degeneration: Areds report no. 17. Archives of ophthalmology (Chicago, Ill.: 1960) **123**(11), 1484–1498 (2005)
11. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., Dean, J.: A guide to deep learning in healthcare. Nature medicine **25**(1), 24–29 (2019)
12. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient graph-based image segmentation. International journal of computer vision **59**(2), 167–181 (2004)
13. Fu, Y., Lei, Y., Wang, T., Curran, W.J., Liu, T., Yang, X.: A review of deep learning based methods for medical image multi-organ segmentation. Physica Medica **85**, 107–122 (2021)
14. Hamilton, M., Zhang, Z., Hariharan, B., Snavely, N., Freeman, W.T.: Unsupervised semantic segmentation by distilling feature correspondences. In: International Conference on Learning Representations (2022)
15. Haque, I.R.I., Neubert, J.: Deep learning approaches to biomedical image segmentation. Informatics in Medicine Unlocked **18**, 100297 (2020)
16. Hore, A., Ziou, D.: Image quality metrics: Psnr vs. ssim. In: 2010 20th international conference on pattern recognition. pp. 2366–2369. IEEE (2010)
17. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)

18. Isgum, I., Staring, M., Rutten, A., Prokop, M., Viergever, M.A., Van Ginneken, B.: Multi-atlas-based segmentation with local decision fusion—application to cardiac and aortic segmentation in ct scans. IEEE transactions on medical imaging **28**(7), 1000–1010 (2009)
19. Kim, W., Kanezaki, A., Tanaka, M.: Unsupervised learning of image segmentation based on differentiable feature clustering. IEEE Transactions on Image Processing **29**, 8055–8068 (2020)
20. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4015–4026 (2023)
21. Kuchroo, M., DiStasio, M., Song, E., Calapkulu, E., Zhang, L., Ige, M., Sheth, A.H., Majdoubi, A., Menon, M., Tong, A., et al.: Single-cell analysis reveals inflammatory interactions driving macular degeneration. Nature Communications **14**(1), 2589 (2023)
22. Kuchroo, M., Huang, J., Wong, P., Grenier, J.C., Shung, D., Tong, A., Lucas, C., Klein, J., Burkhardt, D.B., Gigante, S., et al.: Multiscale phate identifies multimodal signatures of covid-19. Nature biotechnology **40**(5), 681–691 (2022)
23. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**(1), 654 (2024)
24. Mazurowski, M.A., Dong, H., Gu, H., Yang, J., Konz, N., Zhang, Y.: Segment anything model for medical image analysis: an experimental study. Medical Image Analysis p. 102918 (2023)
25. Ott, B.R., Cohen, R.A., Gongvatana, A., Okonkwo, O.C., Johanson, C.E., Stopa, E.G., Donahue, J.E., Silverberg, G.D., Initiative, A.D.N., et al.: Brain ventricular volume and cerebrospinal fluid biomarkers of alzheimer's disease. Journal of Alzheimer's disease **20**(2), 647–657 (2010)
26. Ouyang, C., Biffi, C., Chen, C., Kart, T., Qiu, H., Rueckert, D.: Self-supervision with superpixels: Training few-shot medical image segmentation without annotation. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIX 16. pp. 762–780. Springer (2020)
27. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
28. Shen, L.L., Sun, M., Ahluwalia, A., Young, B.K., Park, M.M., Toth, C.A., Lad, E.M., Del Priore, L.V.: Relationship of topographic distribution of geographic atrophy to visual acuity in nonexudative age-related macular degeneration. Ophthalmology Retina **5**(8), 761–774 (2021)
29. Vincent, L., Soille, P.: Watersheds in digital spaces: an efficient algorithm based on immersion simulations. IEEE Transactions on Pattern Analysis & Machine Intelligence **13**(06), 583–598 (1991)
30. Wang, S., Cao, S., Wei, D., Wang, R., Ma, K., Wang, L., Meng, D., Zheng, Y.: Lt-net: Label transfer by learning reversible voxel-wise correspondence for one-shot medical image segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9162–9171 (2020)
31. Yan, K., Cai, J., Jin, D., Miao, S., Harrison, A.P., Guo, D., Tang, Y., Xiao, J., Lu, J., Lu, L.: Self-supervised learning of pixel-wise anatomical embeddings in radiological images (2020)
32. Zha, H., He, X., Ding, C., Gu, M., Simon, H.: Spectral relaxation for k-means clustering. Advances in neural information processing systems **14** (2001)