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**[3] **Arman Avesta**[4]
Sanjay Aneja[4§] **Jay C. Wang**[5,6§] **Lucian V. Del Priore**[5§] **Smita Krishnaswamy**[1,3§]

[1]Yale University Department of Computer Science
[2]University of California, San Francisco, Department of Ophthalmology
[3]Yale University Department of Genetics [4]Yale University Department of Therapeutic Radiology
[5]Yale University Department of Ophthalmology [6]Northern California Retina Vitreous Associates

* These authors are joint first authors. § Senior authors.
Please direct correspondence to: smita.krishnaswamy@yale.edu or lucian.delpriore@yale.edu.

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.
The code is available at https://github.com/KrishnaswamyLab/CUTS.

1 Introduction

Medical image segmentation plays an increasingly crucial role in both research and clinical settings in a wide array 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 [1]. With high-quality image segmentation, clinicians can more easily diagnose and monitor the progression of diseases to improve patient care. Traditional medical image segmentation methods rely on hand-crafted features [2–7] or predefined atlases [8–10]. These methods are gradually being replaced by deep learning [11–14] as supervised neural networks demonstrate superior performance than 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 issues in applying them to medical images, particularly in order to make clinical inferences. First, these networks depend on expert annotations, so they require a large number of labels to adequately cover the data variance to produce reliable segmentations [12]. Second, supervised networks trained on one set of annotated images can fail to generalize to similar

27th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2024).

images collected in very slightly different contexts, such as in different patient populations or on different devices [15]. 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 a 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 to automatically segment medical images using an entirely unsupervised framework that combines recent advances in representation learning with advances in data geometry and topology. An unsupervised approach circumvents the need for costly expert annotations and alleviates the cross-domain generalization problem. More importantly, we also design our approach to produce multigranular segmentations, which can potentially target multiple regions of interest without supervision.

Our framework, which we denote **C**ontrastive and **U**nsupervised **T**raining for multigranular medical image **S**egmentation (**CUTS**), was named as an homage to the renowned painter Henri Matisse, who famously used a "cut-up" method he called "drawing with scissors" to assemble an image based on patches of material from different sources. Our technique is in essence the reverse of this process, as we start with the initial image and use unsupervised machine learning to segment the initial figure into a collection of relatively homogeneous patches using data coarse graining in a learned latent space. Although it may seem trivial to identify the different pieces of paper cut up by scissors, segmentation of medical images is more challenging as the boundaries between biological structures, such as between healthy and pathological tissues, are not always sharp and clean.

CUTS is designed as an unsupervised segmentation pipeline. The images are processed in units of pixel-centered patches, which consists of a fixed-size crop of image centered on an image pixel. A convolutional encoder is then trained on these pixel-centered patches with both intra-image contrastive learning and local patch reconstruction as optimization objectives. We note that contrastive patches should come from the domain of the medical image itself to create a meaningful pixel embedding. Thus, we find suitable contrastive patches within each image itself using an image similarity metric. Subsequently, the learned embedding space serves as a stronger feature-rich foundation for a multiscale, topology-based data coarse graining method called diffusion condensation that produces multigranular segmentations.

Our main contributions include:

- CUTS, a novel unsupervised framework with a two-stage approach: it first produces an image-specific pixel-centered patch embedding via a convolutional encoder, and subsequently uses diffusion condensation to coarse-grain these patches into clusters at various levels of granularity to perform multiscale segmentation.
- (Specific to the first stage) A novel optimization objective that combines **intra-image contrastive learning** with local patch reconstruction to help the convolutional encoder learn an expressive embedding space.
- (Specific to the second stage) A multiscale cluster assignment approach that utilizes diffusion condensation, which provides clinicians with labels at **multiple levels of segmentation granularities**, potentially highlighting clinically relevant regions at various scales.

The remainder of this paper is organized as follows. First, we discuss related work in the field. With this provided context, we then introduce our framework detailing the neural network architecture, optimization objective, and multiscale segmentation. Finally, we apply our framework on a series of medical image datasets consisting of retinal fundus images and brain MRI images We evaluate the performance of CUTS through qualitative and quantitative metrics and compare to several baselines including other unsupervised approaches and supervised approaches.

2 Related Works

Traditional methods for medical image segmentation Traditional image segmentation methods generally fall into two categories. The first category relies on hand-crafted image features, such as line/edge detection [2], graph cuts [3, 7], active contours [4], watershed [5, 6], level-set [16], and feature clustering [17]. The methods in this category are simple to execute, but they usually struggle

with images with more complicated colors and textures. The second category utilizes a precomputed and annotated atlas to propagate prior knowledge, by warping a predefined set of labels onto new images through image registration [8–10]. These methods require building and annotating an atlas for each image dataset – a time-consuming process that may sometimes be impractical.

Supervised learning for medical image segmentation Supervised deep learning has outperformed traditional methods in segmenting medical images in the past decade [13]. In supervised deep learning, a neural network learns to perform a designated task through a data-driven parameter optimization process [11]. In medical image segmentation, the most well-known method is U-Net [18], followed by a proliferation of variants with skip connections, attention mechanisms, etc. [19–24]. They are all supervised learning methods and thereby require an abundance of expert annotations.

Towards unsupervised learning With a growing emphasis on avoiding reliance on human expert annotations, researchers have been exploring unsupervised learning approaches for medical image segmentation. Many works focused on training with fewer data [25–29]. SSL-ALPNet [30] proposed to directly learn from pseudo-labels generated from Felzenswalb segmentation [7], thus categorizing it as an unsupervised learning method despite a supervised learning approach. DCGN [31] used a constrained Gaussian mixture model to cluster pixel representations in histopathology images. It assumes that different tissue types correspond to different colors, which is not necessarily true in many other medical image modalities.

Atlas-based unsupervised learning is another promising direction. Compared to their traditional counterparts [8–10], the versions empowered by deep learning [32, 33] have improved results. When the domain gap is small, they can be highly effective; otherwise, these methods could fail similarly. Given their requirement for spatial registration, they are more suitable for clearly defined structures that show little variation among individuals and thus are less applicable to image domains with greater variability.

Contrastive learning Contrastive learning [34] was proposed as a generic self-supervised method to address the issue of limited annotations. Conceptually, it allows neural networks to learn meaningful representations in the embedding space by encouraging similar image pairs to be embedded closer to each other and vice versa. After a meaningful embedding space is trained, additional layers can be attached and fine-tuned for downstream tasks. In particular, commonly used contrastive learning methods such as SimCLR [34], SwaV [35], MoCo [36], BYOL [37], BarlowTwins [38] and SimSiam [39] focus on extracting image-level representations with an inter-image contrastive objective. These image-level contrastive learning methods yield no information about intra-image features and are therefore unsuitable for tasks that require closer scrutiny within the same image, such as image segmentation. In an attempt to adapt contrastive learning to tackle the image segmentation task, [40] proposed learning image and patch representations through global and local contrastive training. In [41], the authors used a similar approach, although they coined different terminologies. Both methods include a supervised fine-tuning stage after contrastive pre-training, which still depends on labels.

Unsupervised image segmentation with contrastive learning Two leading unsupervised image segmentation methods, DFC [42] and STEGO [43], both utilize contrastive learning concepts. STEGO learns feature relationships between an image and itself, its k most similar images, and dissimilar images. 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, yet with two key differences. First, DFC contrasts on pixels, while we operate on pixel-centered patches. Pixel-centered patches contain significantly richer semantic and textural information than pixels. Second, we achieve segmentation through a topological multiscale coarse-graining method that produces many segmentation maps at various granularities rather than a single segmentation map.

Segment Anything Model (SAM) and medical variants Segment Anything Model (SAM) [44] recently introduced a general-purpose segmentation tool pre-trained on a gigantic dataset of natural images. As previous researchers have shown [45], SAM offers an alternative solution to label-free medical image segmentation through an interface called "zero-shot transfer", where a single point is

provided as a prompt which is deciphered by a prompt encoder and sent to a mask model to produce a segmentation mask. Alternative input formats, such as text prompt (written text) or box prompt (bounding box) are also supported by this framework.

To better adapt to medical image applications, researchers have developed counterparts that are pre-trained on large datasets of medical images instead of natural images. MedSAM [46] and SAM-Med2D [47] are among the most popular variants.

Strictly speaking, SAM and its variants are not unsupervised learning methods. As a result, they still face the cross-domain generalization problem as previously mentioned, while their brute-force solution is to cover the entire data distribution with the huge training set. Despite their non-unsupervised nature, we decided to include them for comparison, since they are arguably the latest state-of-the-art segmentation framework.

3 Methods

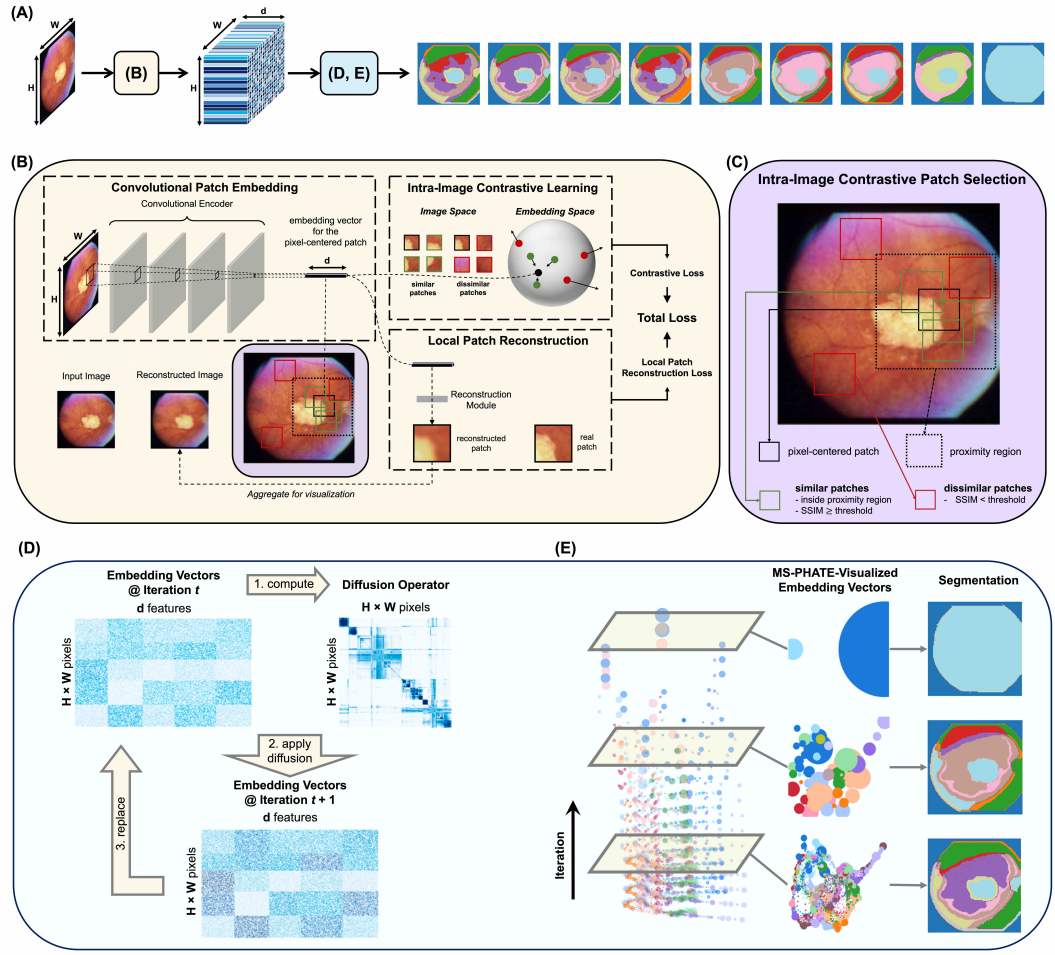

Figure 1: The CUTS Framework. **(A)** Overview. **(B)** Pixel-centered patches are mapped into the embedding space, jointly optimized by two objectives. **(C)** Positive and negative patch pairs are selected based on proximity and structural similarity. **(D)** Diffusion condensation coarse grains embedding 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) [48] 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 [49, 50]. Metastable granularities can be automatically identified from the condensation homology as granularities with zero topological activity [50]. Segmentation is performed by assigning labels to pixels that correspond to clusters arising from a particular metastable granularity (Fig. 1(D-E)).

3.1 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 [51] > 0.5) to the anchor patch are considered positive patches. The contrastive loss is defined by Eqn (1) where $sim(\cdot)$ denotes cosine similarity.

$$l_{contrast} = -\log \frac{\text{pos}}{\text{neg}}, \quad \text{pos} = \sum_{z^+_{ij}\in\Omega^+} e^{sim(z_{ij},z^+_{ij})/\tau}, \quad \text{neg} = \sum_{z^-_{ij}\in\Omega^-} e^{sim(z_{ij},z^-_{ij})/\tau} \tag{1}$$

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 is a weighted sum of the contrastive loss and reconstruction loss, balanced by a weighting coefficient $\lambda \in [0, 1]$.

$$loss = \lambda \cdot l_{contrast} + (1 - \lambda) \cdot l_{recon} \tag{2}$$

Hyperparameters Three key hyperparameters were tuned empirically (Fig. 2). First, we found that the optimal patch size for pixel-centered patches is 5×5. Then, we determined to sample 8 patches in each image for contrastive learning and reconstruction. Lastly, we set the contrastive loss coefficient at 0.0001. Note that $l_{contrast}$ is still nontrivial after weighing, because the numerical value of $l_{contrast}$ is more than 3 orders of magnitude higher than l_{recon} at convergence.

3.2 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.

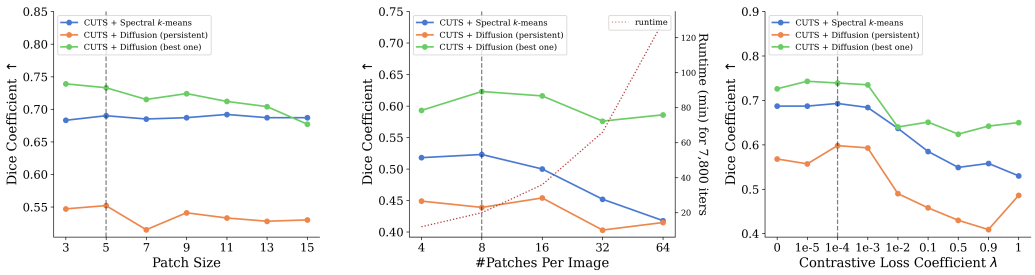

Figure 2: Effects of hyperparameters.

Diffusion condensation [49, 50] 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.

Formally, from a data matrix $X^{N\times d}$ with N observations (in our case, $N = W \times H =$ the number of pixels in an image) and d features, we can construct the local affinity among each observation pair $(m, n) \in \{1, ..., N\}$ using a Gaussian kernel

$$\mathbf{K}(x_m, x_n) = e^{-\frac{||x_m - x_n||^2}{\epsilon}} \tag{3}$$

$\mathbf{K}$ is a $N \times N$ Gram matrix whose (m, n) entry is denoted $\mathbf{K}(x_m, x_n)$ to emphasize the dependency on the data matrix X. x_m and x_n are both of dimension $\mathbb{R}^d$. The bandwidth parameter ϵ controls neighborhood sizes.

Given this affinity matrix $\mathbf{K}$, the diffusion operator is defined by Eqn (4a) where $\mathbf{D}$ is the diagonal degree matrix as shown in Eqn (4b).

$$\mathbf{P} = \mathbf{D}^{-1}\mathbf{K} \tag{4a}$$

$$\mathbf{D}(x_m, x_m) = \sum_n \mathbf{K}(x_m, x_n) \tag{4b}$$

The diffusion operator $\mathbf{P}$ defines the single-step transition probabilities for a diffusion process over the data, which can be viewed as a Markovian random walk. To perform multi-step diffusion, one way is to simulate a time-homogeneous diffusion process by raising the diffusion operator to a power of t which leads to $X_t = \mathbf{P}^t X$ [52]. On the other hand, as shown in [50], we could simulate a time-inhomogeneous diffusion process by iteratively computing the diffusion operator and the data matrix in the following manner.

$$\begin{aligned}
&X_0 \leftarrow X \\
&\textbf{for } t \in [1, ..., T]: \\
&\quad \mathbf{K}_{t-1} \leftarrow \mathcal{K}(X_{t-1}) && \text{/* using Eq. (3) */} \\
&\quad \mathbf{D}_{t-1} \leftarrow \mathcal{D}(\mathbf{K}_{t-1}) && \text{/* using Eq. (4b) */} \\
&\quad \mathbf{P}_{t-1} \leftarrow \mathbf{D}_{t-1}^{-1}\mathbf{K}_{t-1} && \text{/* using Eq. (4a) */} \\
&\quad X_t \leftarrow \mathbf{P}_{t-1}X_{t-1}
\end{aligned} \tag{5}$$

The process of diffusion condensation can be summarized as the alternation between the following two steps:

1. Computing a time-inhomogeneous diffusion operator from the data at iteration t.
2. Applying this operator to the data, moving points towards the local center of gravity, which forms the data in iteration $t + 1$.

More details on diffusion condensation can be found in [50]. In this paper, we used the official implementation (`https://github.com/KrishnaswamyLab/catch`).

We can identify the segments that occur consistently over the series of segmentations, called persistent structures. The terminology "persistence" is a measure defined in diffusion condensation as clusters that stay separated over many iterations. The discovery of persistent structures 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 [43, 45], 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 Empirical Results

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 [53, 54]. 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 [55]. 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.

4.1 Qualitative results on multigranular segmentation

As shown in Fig. 3, 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 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.

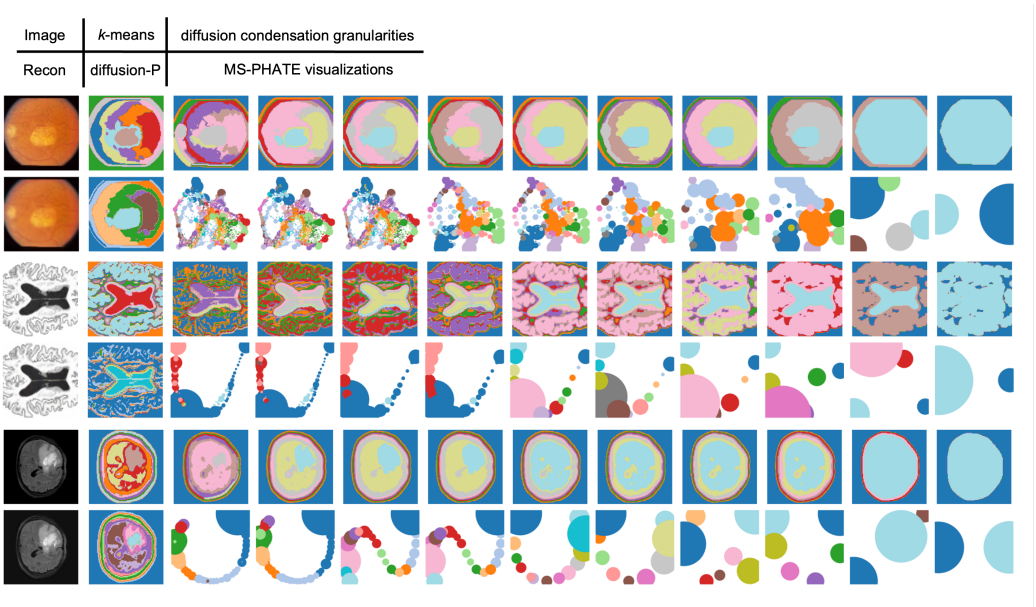

Figure 3: Multigranular segmentation (odd rows) captures distinctive patterns at various scales. Multiscale PHATE (even rows) is used to visualize the diffusion condensation process. The results of CUTS + spectral k-means clustering ("k-means") and CUTS + diffusion condensation persistent structures ("diffusion-P") are also shown for reference.

4.2 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 [5], Felzenszwalb [7], and SLIC [17]. We then compared with DFC [42] and STEGO [43], 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) [44, 45] which was pre-trained on 11 million images and 1.1 billion masks, as well as its medical-image variants (MedSAM [46]/SAM-Med2D [47]) 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 [45].

Lastly, for reference, we benchmarked a random labeler as the performance lower bound and two fully supervised methods, UNet [18] and nn-UNet [24], as the upper bound. For coarse-graining of the pixel embeddings, we also implemented a spectral k-means clustering [56] 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*.

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. 4). 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 con-

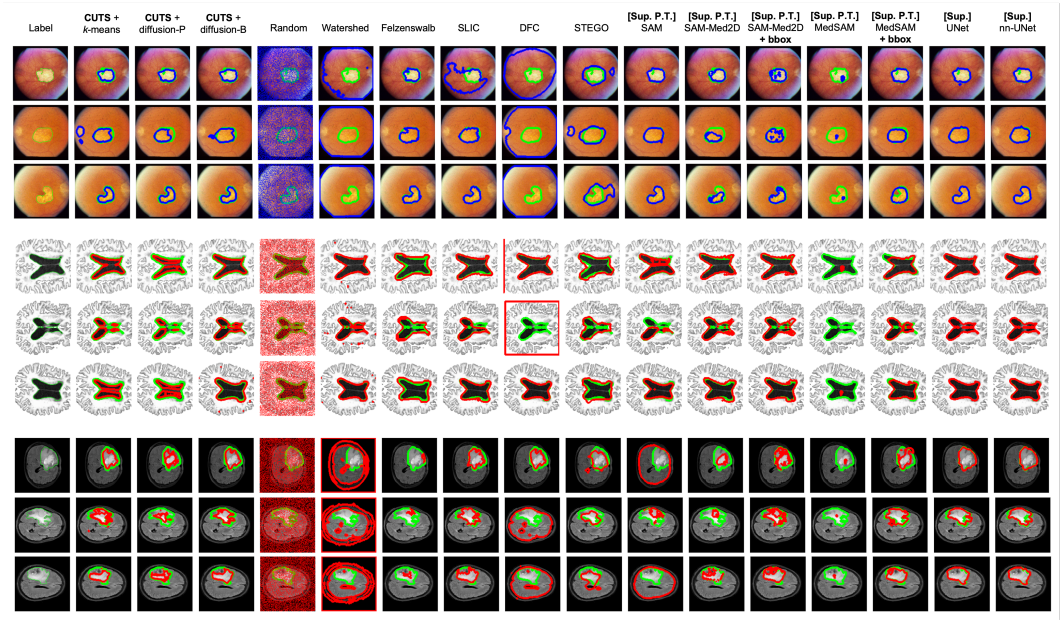

Figure 4: Qualitative segmentation comparison. Green curves outline the ground truth labels while blue or red curves outline the predictions. "diffusion-B": the best diffusion condensation granularity. "**Sup.**": supervised "**P.T.**": pre-training. "**+bbox**": using bounding boxes instead of points as input; included for completeness but would be unfair for comparison.

sidered clinically important because the volume of the brain ventricles can predict the progression of dementia [57, 58].

Qualitatively, CUTS delineated the brain ventricles in a wide variety of settings (Fig. 4). 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. 4. 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. CUTS led the other unsupervised methods by a larger margin compared to the less demanding task of ventricle segmentation.

Comparison with SAM, MedSAM and SAM-Med2D More impressively, as shown in Table 1, CUTS achieved better results than every SAM variant on at least 2 out of 3 datasets — under fair comparison of using a single point as input — 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).

Table 1: Quantitative comparisons from 3 random seeds. Among unsupervised methods, the best is **bolded** and runner-up is underscored. §Entries with "+bbox" use bounding boxes instead of points as input. They are included for completeness but would be unfair for comparison. ‡Diffusion condensation will not run since #features = 1 for each pixel in single-channel images. *The suboptimal performance of MedSAM is expected. According to the authors, "the point prompt is still an experimental function and the model was trained on a small abdomen CT organ segmentation dataset."

			Retinal Atrophy		Brain Ventricles		Brain Tumor	
	Deep learning?	Topological?	DSC ↑	HD ↓	DSC ↑	HD ↓	DSC ↑	HD ↓
Unsupervised, without learning								
Watershed (IEEE TPAMI'91 [5])	✗	✗	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 [7])	✗	✗	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 [17])	✗	✗	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 [42])	✓	✗	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 [43])	✓	✗	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 [44], MedIA'23 [45])	✓	✗	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 [47])	✓	✗	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 [46])	✓	✗	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 [18])	✓	✗	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 [24])	✓	✗	0.937±0.014	6.00±1.35	0.984±0.005	2.10±0.42	0.834±0.024	8.64±1.60

5 Conclusion

CUTS is a deep learning and topological framework that identifies and highlights important medical image structures using unsupervised learning. 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 like SAM necessitate domain-specific fine-tuning for tasks not covered by the initial supervised pre-training, which highlights the continued relevance of approaches like CUTS that investigate objectives, modules and techniques to inject the correct inductive biases. Therefore, CUTS offers a practical and effective alternative in the evolving landscape of medical imaging.

6 Discussion

Current state-of-the-art methods for medical image segmentation are primarily supervised and therefore require domain experts to annotate a large number of medical images. Moreover, it is often infeasible to collect enough images of rare diseases to train supervised learning models. For example, in this work we studied a retinal degeneration condition. The number of images available to any institution of this condition is usually within a hundred, several orders of magnitude fewer than popular natural image databases for deep learning with millions of images. Furthermore, another limitation of supervised learning approaches is the domain generalization problem. When a method is optimized for a specific type of image used for training, its performance may suffer if used on other types of image, even if they are only slightly dissimilar.

In contrast, unsupervised methods, like CUTS, while more challenging architect, do not require human expert grading and thereby circumvent this time-consuming, expensive, and labor-intensive initial step. Unsupervised methods can also be applied to much smaller datasets, ideal for rare diseases. Unfortunately, prior attempts to use unsupervised learning to segment medical images have not achieved the desired results. These unsupervised methods often yield subpar performance, despite having advantages including independence from labels and the ability to generalize to new datasets while preserving robustness.

CUTS bridges the difficulty of creating unsupervised images by using the key insight that, while an image as a whole may be hard to segment, pixels forming boundaries of image features may be

detectable by their local context. Thus, CUTS features carefully architected losses for local pixel-centered patch reconstruction and pixel-centered patch-based contrastive losses based on within-image augmentation of patches. With these unique penalties, CUTS learns an intermediate representation of a pixel-centered patch embedding for each image. The key advantage of this pixel-centered patch representation is that it is amenable to not one segmentation, but several multigranular segmentations of the same image via a topological coarse-graining scheme. The final output of CUTS is thus several segmentations of the image with features of different resolutions of interest for different clinical queries.

In our brain tumor image dataset, for example, CUTS enables: (1) brain extraction on the coarsest scale, (2) isolation of white matter, gray matter, and cerebrospinal fluid on the intermediate scale, and (3) small tumor segmentation on the finest scale. These different features can be important for different diagnostic purposes such as tumor placement identification for surgical purposes or small tumor size extraction for the analysis of metastases.

In this work, we demonstrate the application of CUTS to three medical image datasets from different medical domains. On the retinal fundus images, the watershed, Felzenszwalb, and DFC focus primarily on the contours of the retina without distinguishing the geographic atrophy regions, where the contrast is more subtle. SLIC and STEGO generally perform better, yet they tend to overestimate the region of interest. CUTS avoids all these caveats and consistently segments geographical atrophy. On brain MRI for ventricle segmentation, the watershed method often ignores the frontal or lateral half of the ventricles. Felzenszwalb, SLIC, and DFC have difficulty determining the segmentation boundary. STEGO tends to include tissues around the ventricles. CUTS is slightly more conservative on the boundary regions but nevertheless outperforms other unsupervised methods. On the brain MRI images for tumor segmentation, the watershed and Felzenszwalb methods merely isolate the entire brain from the background with no attention to detailed structures. SLIC, DFC, and STEGO either ignored the tumor region or merged it with the background. CUTS, on the other hand, is sensitive to tenuous contrast transitions in tumor regions and generates significantly better segmentations.

In conclusion, CUTS allows us to identify and highlight important medical image structures using an unsupervised learning approach. This has enormous implications for the expanding field of medical image evaluation and represents a step forward in identifying important and distinct information relevant to the clinical interpretation of images. Such interpretation of medical images is crucial to disease detection in asymptomatic individuals; examples include mammography for screening for breast cancer, teleophthalmology and automated image analysis for screening diabetes for eye disease, or vulnerable populations for macular degeneration and glaucoma, and screening of high-risk populations such as smokers for early lung cancer.

7 Acknowledgements

The authors would like to thank Mengyuan Sun, Aneesha Ahluwalia, Benjamin K. Young, and Michael M. Park for delineating geographic atrophy borders on fundus photographs.

This work was supported in part by the National Science Foundation (NSF DMS 2327211, NSF Career Grant 2047856) and the National Institute of Health (NIH 1R01GM130847-01A1, NIH 1R01GM135929-01).

References

[1] Yabo Fu, Yang Lei, Tonghe Wang, Walter J Curran, Tian Liu, and Xiaofeng Yang. A review of deep learning based methods for medical image multi-organ segmentation. *Physica Medica*, 85:107–122, 2021.

[2] Pinaki Pratim Acharjya, Ritaban Das, and Dibyendu Ghoshal. Study and comparison of different edge detectors for image segmentation. *Global Journal of Computer Science and Technology*, 2012.

[3] Xinjian Chen and Lingjiao Pan. A survey of graph cuts/graph search based medical image segmentation. *IEEE reviews in biomedical engineering*, 11:112–124, 2018.

[4] Issam El Naqa, Deshan Yang, Aditya Apte, Divya Khullar, Sasa Mutic, Jie Zheng, Jeffrey D Bradley, Perry Grigsby, and Joseph O Deasy. Concurrent multimodality image segmentation by active contours for radiotherapy treatment planning a. *Medical physics*, 34(12):4738–4749, 2007.

[5] Luc Vincent and Pierre Soille. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 13(06):583–598, 1991.

[6] Hyun-Hwa Oh, Kil-Taek Lim, and Sung-Il Chien. An improved binarization algorithm based on a water flow model for document image with inhomogeneous backgrounds. *Pattern Recognition*, 38(12):2612–2625, 2005.

[7] Pedro F Felzenszwalb and Daniel P Huttenlocher. Efficient graph-based image segmentation. *International journal of computer vision*, 59(2):167–181, 2004.

[8] Ivana Isgum, Marius Staring, Annemarieke Rutten, Mathias Prokop, Max A Viergever, and Bram Van Ginneken. 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.

[9] Paul Aljabar, Rolf A Heckemann, Alexander Hammers, Joseph V Hajnal, and Daniel Rueckert. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *Neuroimage*, 46(3):726–738, 2009.

[10] Juan Eugenio Iglesias and Mert R Sabuncu. Multi-atlas segmentation of biomedical images: a survey. *Medical image analysis*, 24(1):205–219, 2015.

[11] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[12] Andre Esteva, Alexandre Robicquet, Bharath Ramsundar, Volodymyr Kuleshov, Mark DePristo, Katherine Chou, Claire Cui, Greg Corrado, Sebastian Thrun, and Jeff Dean. A guide to deep learning in healthcare. *Nature medicine*, 25(1):24–29, 2019.

[13] Intisar Rizwan I Haque and Jeremiah Neubert. Deep learning approaches to biomedical image segmentation. *Informatics in Medicine Unlocked*, 18:100297, 2020.

[14] Shervin Minaee, Yuri Y Boykov, Fatih Porikli, Antonio J Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos. Image segmentation using deep learning: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 2021.

[15] Michael D Abràmoff, Philip T Lavin, Michele Birch, Nilay Shah, and James C Folk. 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.

[16] Andy Tsai, Anthony Yezzi, William Wells, Clare Tempany, Dewey Tucker, Ayres Fan, W Eric Grimson, and Alan Willsky. A shape-based approach to the segmentation of medical imagery using level sets. *IEEE transactions on medical imaging*, 22(2):137–154, 2003.

[17] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2274–2282, 2012.

[18] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[19] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pages 3–11. Springer, 2018.

[20] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.

[21] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention*, pages 424–432. Springer, 2016.

[22] Simon Kohl, Bernardino Romera-Paredes, Clemens Meyer, Jeffrey De Fauw, Joseph R Ledsam, Klaus Maier-Hein, SM Eslami, Danilo Jimenez Rezende, and Olaf Ronneberger. A probabilistic u-net for segmentation of ambiguous images. *Advances in neural information processing systems*, 31, 2018.

[23] Xin Yu, Qi Yang, Yinchi Zhou, Leon Y Cai, Riqiang Gao, Ho Hin Lee, Thomas Li, Shunxing Bao, Zhoubing Xu, Thomas A Lasko, et al. Unest: local spatial representation learning with hierarchical transformer for efficient medical segmentation. *Medical Image Analysis*, 90:102939, 2023.

[24] Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021.

[25] Arnab Kumar Mondal, Jose Dolz, and Christian Desrosiers. Few-shot 3d multi-modal medical image segmentation using generative adversarial learning. *arXiv preprint arXiv:1810.12241*, 2018.

[26] Cheng Ouyang, Konstantinos Kamnitsas, Carlo Biffi, Jinming Duan, and Daniel Rueckert. Data efficient unsupervised domain adaptation for cross-modality image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22*, pages 669–677. Springer, 2019.

[27] Hanchao Yu, Shanhui Sun, Haichao Yu, Xiao Chen, Honghui Shi, Thomas S Huang, and Terrence Chen. Foal: Fast online adaptive learning for cardiac motion estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4313–4323, 2020.

[28] Chen Chen, Chen Qin, Huaqi Qiu, Cheng Ouyang, Shuo Wang, Liang Chen, Giacomo Tarroni, Wenjia Bai, and Daniel Rueckert. Realistic adversarial data augmentation for mr image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part I 23*, pages 667–677. Springer, 2020.

[29] Qi Yang, Xin Yu, Ho Hin Lee, Yucheng Tang, Shunxing Bao, Kristofer S Gravenstein, Ann Zenobia Moore, Sokratis Makrogiannis, Luigi Ferrucci, and Bennett A Landman. Label efficient segmentation of single slice thigh ct with two-stage pseudo labels. *Journal of Medical Imaging*, 9(5):052405–052405, 2022.

[30] Cheng Ouyang, Carlo Biffi, Chen Chen, Turkay Kart, Huaqi Qiu, and Daniel Rueckert. 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*, pages 762–780. Springer, 2020.

[31] Yang Nan, Peng Tang, Guyue Zhang, Caihong Zeng, Zhihong Liu, Zhifan Gao, Heye Zhang, and Guang Yang. Unsupervised tissue segmentation via deep constrained gaussian network. *IEEE Transactions on Medical Imaging*, 41(12):3799–3811, 2022.

[32] Amy Zhao, Guha Balakrishnan, Fredo Durand, John V Guttag, and Adrian V Dalca. Data augmentation using learned transformations for one-shot medical image segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8543–8553, 2019.

[33] Shuxin Wang, Shilei Cao, Dong Wei, Renzhen Wang, Kai Ma, Liansheng Wang, Deyu Meng, and Yefeng Zheng. 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*, pages 9162–9171, 2020.

[34] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations, 2020.

[35] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. *Advances in Neural Information Processing Systems*, 33:9912–9924, 2020.

[36] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9729–9738, 2020.

[37] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, et al. Bootstrap your own latent-a new approach to self-supervised learning. *Advances in neural information processing systems*, 33:21271–21284, 2020.

[38] Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. Barlow twins: Self-supervised learning via redundancy reduction. In *International Conference on Machine Learning*, pages 12310–12320. PMLR, 2021.

[39] Xinlei Chen and Kaiming He. Exploring simple siamese representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15750–15758, 2021.

[40] Krishna Chaitanya, Ertunc Erdil, Neerav Karani, and Ender Konukoglu. Contrastive learning of global and local features for medical image segmentation with limited annotations, 2020.

[41] Ke Yan, Jinzheng Cai, Dakai Jin, Shun Miao, Adam P. Harrison, Dazhou Guo, Youbao Tang, Jing Xiao, Jingjing Lu, and Le Lu. Self-supervised learning of pixel-wise anatomical embeddings in radiological images, 2020.

[42] Wonjik Kim, Asako Kanezaki, and Masayuki Tanaka. Unsupervised learning of image segmentation based on differentiable feature clustering. *IEEE Transactions on Image Processing*, 29:8055–8068, 2020.

[43] Mark Hamilton, Zhoutong Zhang, Bharath Hariharan, Noah Snavely, and William T Freeman. Unsupervised semantic segmentation by distilling feature correspondences. In *International Conference on Learning Representations*, 2022.

[44] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4026, 2023.

[45] Maciej A Mazurowski, Haoyu Dong, Hanxue Gu, Jichen Yang, Nicholas Konz, and Yixin Zhang. Segment anything model for medical image analysis: an experimental study. *Medical Image Analysis*, page 102918, 2023.

[46] Jun Ma, Yuting He, Feifei Li, Lin Han, Chenyu You, and Bo Wang. Segment anything in medical images. *Nature Communications*, 15(1):654, 2024.

[47] Junlong Cheng, Jin Ye, Zhongying Deng, Jianpin Chen, Tianbin Li, Haoyu Wang, Yanzhou Su, Ziyan Huang, Jilong Chen, Lei Jiang, et al. Sam-med2d. *arXiv preprint arXiv:2308.16184*, 2023.

[48] Manik Kuchroo, Jessie Huang, Patrick Wong, Jean-Christophe Grenier, Dennis Shung, Alexander Tong, Carolina Lucas, Jon Klein, Daniel B Burkhardt, Scott Gigante, et al. Multiscale phate identifies multimodal signatures of covid-19. *Nature biotechnology*, 40(5):681–691, 2022.

[49] Nathan Brugnone, Alex Gonopolskiy, Mark W Moyle, Manik Kuchroo, David van Dijk, Kevin R Moon, Daniel Colon-Ramos, Guy Wolf, Matthew J Hirn, and Smita Krishnaswamy. Coarse graining of data via inhomogeneous diffusion condensation. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 2624–2633. IEEE, 2019.

[50] Manik Kuchroo, Marcello DiStasio, Eric Song, Eda Calapkulu, Le Zhang, Maryam Ige, Amar H Sheth, Abdelilah Majdoubi, Madhvi Menon, Alexander Tong, et al. Single-cell analysis reveals inflammatory interactions driving macular degeneration. *Nature Communications*, 14(1):2589, 2023.

[51] Alain Hore and Djemel Ziou. Image quality metrics: Psnr vs. ssim. In *2010 20th international conference on pattern recognition*, pages 2366–2369. IEEE, 2010.

[52] Ronald R Coifman and Stéphane Lafon. Diffusion maps. *Applied and computational harmonic analysis*, 21(1):5–30, 2006.

[53] Matthew D Davis, Ronald E Gangnon, Li Yin Lee, Larry D Hubbard, BE Klein, Ronald Klein, Frederick L Ferris, Susan B Bressler, Roy C Milton, 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.

[54] Liangbo L Shen, Mengyuan Sun, Aneesha Ahluwalia, Benjamin K Young, Michael M Park, Cynthia A Toth, Eleonora M Lad, and Lucian V Del Priore. Relationship of topographic distribution of geographic atrophy to visual acuity in nonexudative age-related macular degeneration. *Ophthalmology Retina*, 5(8):761–774, 2021.

[55] Karen L Crawford, Scott C Neu, and Arthur W Toga. The image and data archive at the laboratory of neuro imaging. *Neuroimage*, 124:1080–1083, 2016.

[56] Hongyuan Zha, Xiaofeng He, Chris Ding, Ming Gu, and Horst Simon. Spectral relaxation for k-means clustering. *Advances in neural information processing systems*, 14, 2001.

[57] Owen T Carmichael, Lewis H Kuller, Oscar L Lopez, Paul M Thompson, Rebecca A Dutton, Allen Lu, Sharon E Lee, Jessica Y Lee, Howard J Aizenstein, Carolyn Cidis Meltzer, et al. Ventricular volume and dementia progression in the cardiovascular health study. *Neurobiology of aging*, 28(3):389–397, 2007.

[58] Brian R Ott, Ronald A Cohen, Assawin Gongvatana, Ozioma C Okonkwo, Conrad E Johanson, Edward G Stopa, John E Donahue, Gerald D Silverberg, Alzheimer's Disease Neuroimaging Initiative, et al. Brain ventricular volume and cerebrospinal fluid biomarkers of alzheimer's disease. *Journal of Alzheimer's disease*, 20(2):647–657, 2010.