MICCAI 2024 Marrakesh MOROCCO

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

Chen Liu*, Matthew Amodio*, Liangbo L. Shen, Feng Gao, Arman Avesta, Sanjay Aneja, Jay C. Wang, Lucian V. Del Priore, Smita Krishnaswamy
Please direct correspondence to: **smita.krishnaswamy@yale.edu** or **lucian.delpriore@yale.edu**.
GitHub: https://github.com/ChenLiu-1996/CUTS and https://github.com/KrishnaswamyLab/CUTS.

1. Motivation and background

Unsupervised Segmentation

Why unsupervised?

- Supervised
  - Pros: SOTA, easy to train
  - Cons: Issues with cross-domain generalization, and requires labeling ***per dataset*** and ***per target***.
- Unsupervised
  - Pros: Lower demand on labeling.
  - Cons: Less mature and needs improvement.

CUTS

What is CUTS named after?

1. To honor 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. This inspired us that meaningful image segmentations are comprised of generally contiguous regions of similar color and texture.
2. **CUTS** stands for "**C**ontrastive and **U**nsupervised **T**raining for **S**egmentation".

2. Methods (Figure 1)

Stage 1

Encode "pixel-centered patches" into **rich embeddings**, with *intra-image contrastive learning* and *local patch reconstruction* (Figure 1 (B-C)).

Stage 2

Coarse grain the rich embeddings into segmentation maps at **various levels of granularities**, with diffusion condensation (Figure 1 (D-E)).

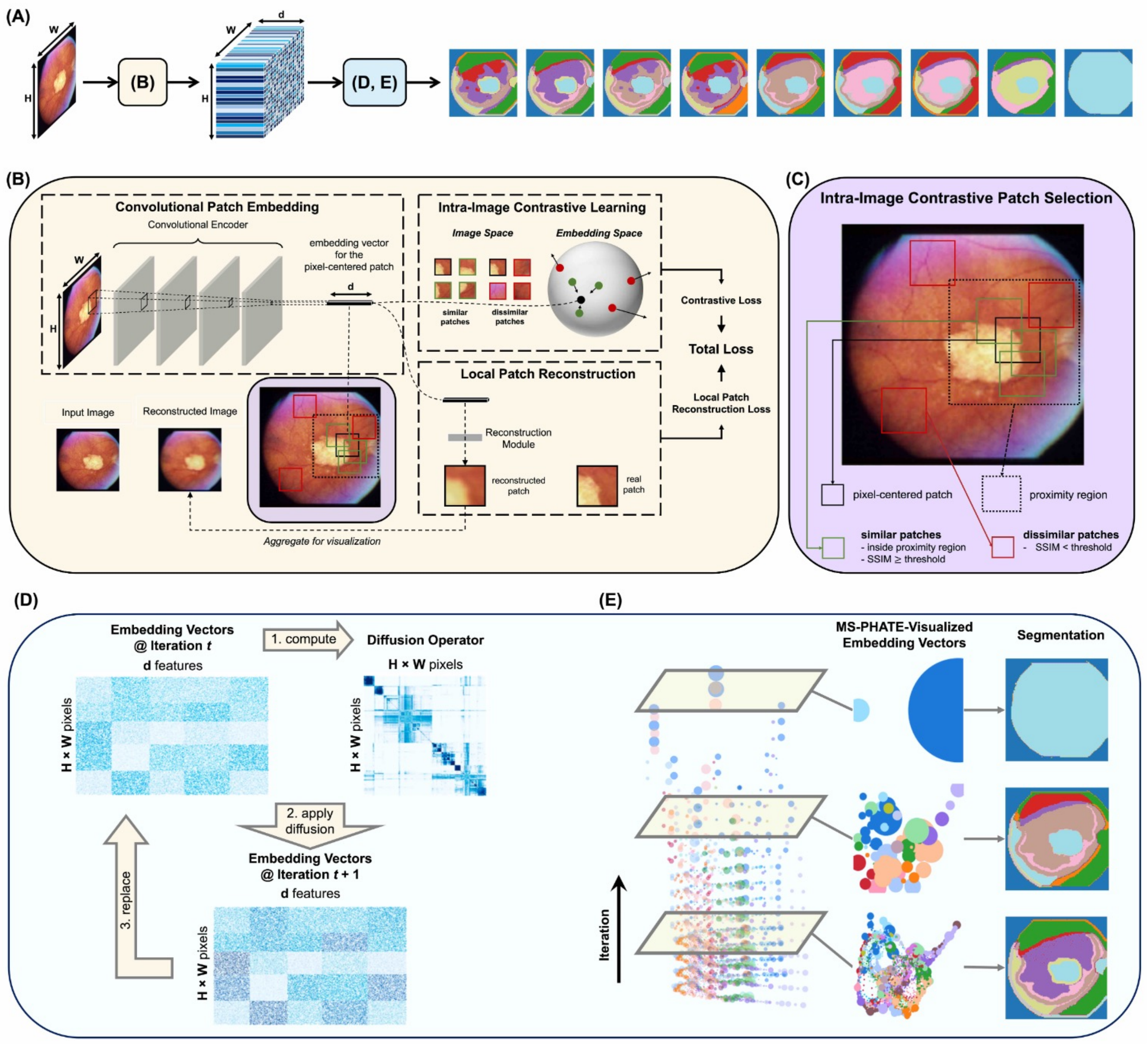


Figure 1: The CUTS Framework. **(A)** Overview. **(B)** Pixel-centered patches are mapped into the embedding space, jointly optimized by two objectives. **(C)** Positive and negative patch pairs are selected based on proximity and structural similarity. **(D)** Diffusion condensation coarse grains embedding 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.

3. Results

Qualitative Results on Multiscale Segmentation

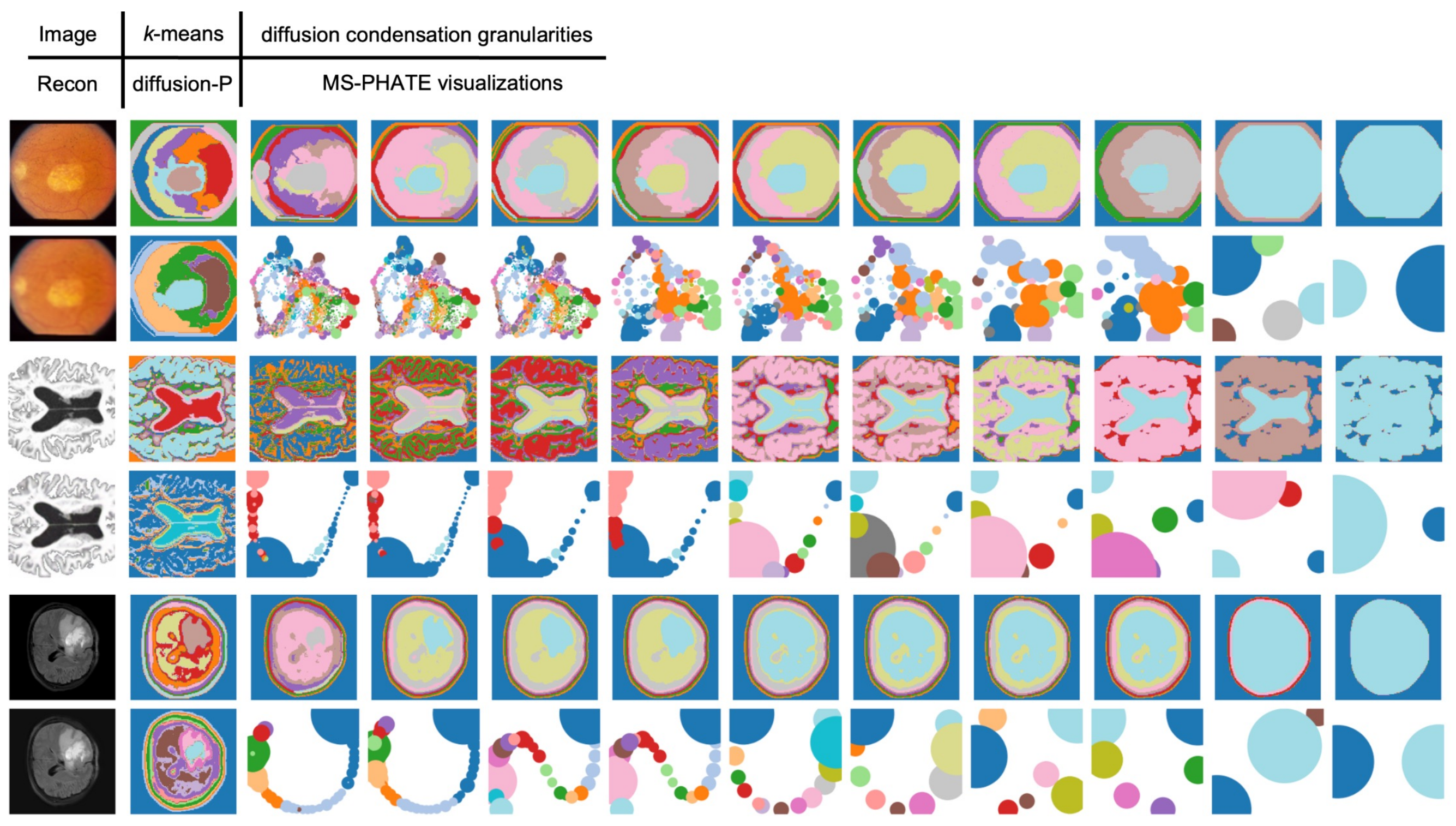


Figure 3: Multigranular segmentation (odd rows) captures distinctive patterns at various scales. Multiscale PHATE (even rows) is used to visualize the diffusion condensation process. The results of CUTS + spectral k-means clustering ("k-means") and CUTS + diffusion condensation persistent structures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

Qualitative Results on Binary Segmentation

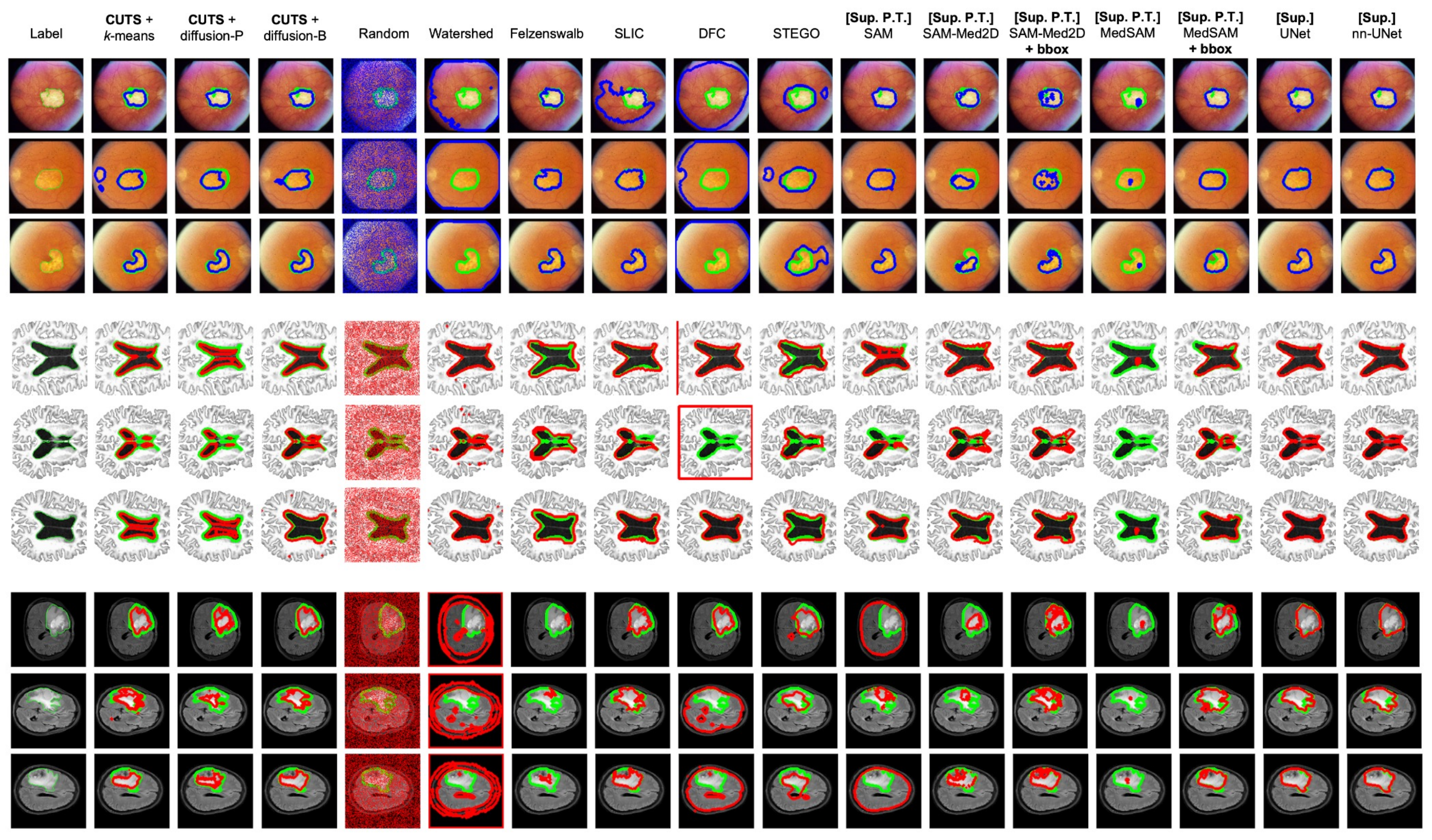


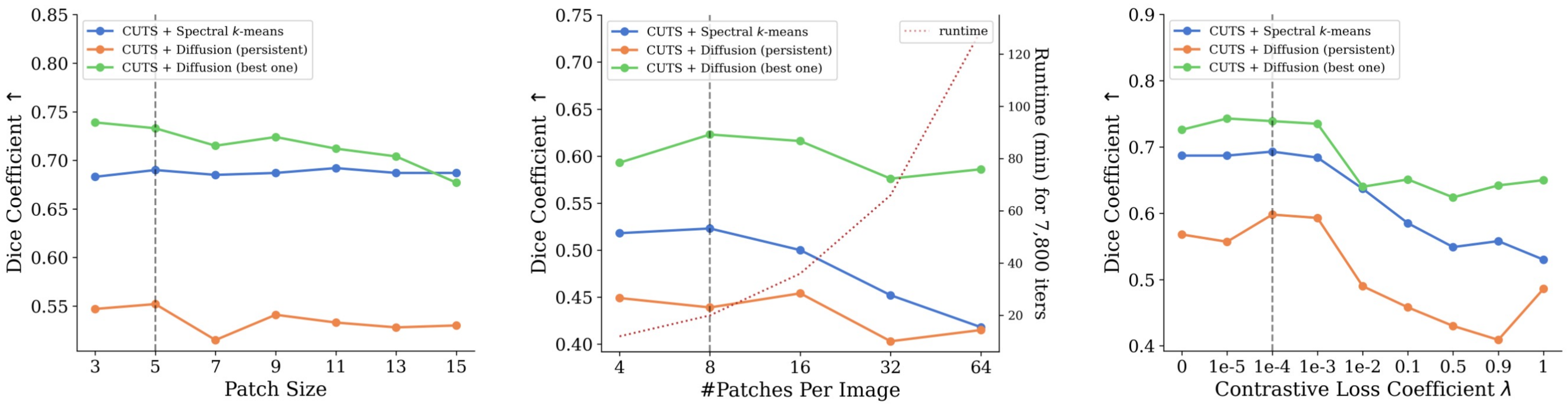
Figure 4: Qualitative segmentation comparison. Green curves outline the ground truth labels while blue or red curves outline the predictions. "diffusion-B": the best diffusion condensation granularity. "**Sup.**": supervised "**P.T.**": pre-training. "**+bbox**": using bounding boxes instead of points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

Quantitative Results on Binary Segmentation & Ablation Studies

Table 1: Quantitative comparisons from 3 random seeds. Among unsupervised methods, the best is **bolded** and runner-up is <u>underscored</u>. §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	<u>0.781</u>±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	✓	✗	<u>0.675</u>±0.014	26.82±0.88	0.774±0.008	<u>8.31</u>±0.23	<u>0.432</u>±0.010	33.94±0.65
(Ours) CUTS + Diffusion (pers.)	✓	✓	0.604±0.003	<u>21.69</u>±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	<u>25.16</u>±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

Selection of Hyper-parameters


CUTS + Spectral k-means
CUTS + Diffusion (persistent)
CUTS + Diffusion (best one)
Dice Coefficient ↑
Patch Size
#Patches Per Image
runtime
Runtime (min) for 7,800 iters
1e-5
1e-4
1e-3
1e-2
Contrastive Loss Coefficient λ