



This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

# Leveraging Multi-Representation Features from Diffusion Models for Unsupervised 3D Medical Image Segmentation

Tanatta Chaichakan<sup>1</sup>[0009–0004–5827–209X], Fani Deligianni<sup>1</sup>[0000–0003–1306–5017], and Paul Henderson<sup>1</sup>[0000–0002–5198–7445]

School of Computing Science, University of Glasgow, Glasgow, UK  
2831588c@student.gla.ac.uk

**Abstract.** We propose a complementary feature integration framework for unsupervised 3D medical image segmentation that repurposes a pre-trained 3D latent diffusion model as a feature extractor. The framework integrates raw image intensity with multiple representations extracted from different components of the diffusion model, including VAE latent representations, diffusion U-Net features, and self-attention features. These representations are integrated into a unified feature space together with spatial coordinates and subsequently segmented using recursive 3D normalized cuts. We evaluate the proposed framework on BraTS2020 brain tumour MRI and FLARE22 abdominal CT datasets using identical hyperparameters without dataset-specific tuning. Experimental results show that integrating multiple representations improves segmentation performance across both datasets compared with individual representations and baseline methods. The proposed framework achieves the best overall average performance across the evaluated datasets while demonstrating the ability to generalise across MRI and CT modalities. These findings suggest that representations derived from different components of the diffusion framework capture complementary anatomical information that benefits unsupervised 3D medical image segmentation. Code is publicly available at <https://doi.org/10.5281/zenodo.21375116>.

**Keywords:** Unsupervised medical image segmentation · diffusion representations · normalized cuts · feature integration

## 1 Introduction

Unsupervised segmentation in medical imaging has shown potential as an alternative to supervised approaches [1,9,12] by eliminating the need for costly and time-consuming expert annotations [18]. However, accurate segmentation remains challenging because medical images often contain low-contrast regions, noise, and substantial anatomical variability [14,7,10]. Furthermore, the performance of unsupervised methods depends heavily on the underlying feature representations, and poorly discriminative features may lead to suboptimal segmentation results [13].

Existing studies on unsupervised segmentation have explored a wide range of techniques and demonstrated promising results across various biomedical imaging datasets [4,5,6,8,16,19]. However, several recent approaches are designed primarily for 2D images, whereas medical imaging data are inherently three-dimensional and require volumetric consistency across slices. A further limitation is that many methods rely on a single feature representation, which may not fully capture the diverse characteristics of medical images. Relevant information exists at multiple levels, including local appearance, global context, multi-scale structures, and spatial relationships. Therefore, integrating complementary representations may improve unsupervised 3D medical image segmentation.

In this work, we propose an unsupervised framework for 3D medical image segmentation that repurposes a pretrained 3D latent diffusion model [15] as a feature extractor. Rather than using the model for medical image synthesis, we leverage representations extracted from different components of the diffusion framework as informative features for graph-based segmentation. The proposed framework combines image intensity (local appearance), latent representations from a VAE encoder (global contextual information), hierarchical features from the diffusion U-Net (multi-scale structural information), and self-attention features (spatial relationships). These features are augmented with spatial coordinates and fused to construct an affinity graph, which is subsequently partitioned using recursive 3D normalized cuts (N-Cuts) to produce the final segmentation. We evaluate the proposed framework on both MRI and CT datasets under consistent experimental settings. Our contributions are:

- An unsupervised graph-based framework for 3D medical image segmentation that leverages a pretrained 3D latent diffusion model for feature extraction and recursive 3D normalized cuts for segmentation.
- An empirical analysis of complementary representations, including intensity, VAE, diffusion, and self-attention features, for unsupervised segmentation.
- A unified evaluation across MRI and CT datasets using identical hyperparameter settings to assess generalisation across imaging modalities.

## 2 Related Work

**Classical Methods.** Unsupervised segmentation has been studied using clustering-based, probabilistic, and graph-based approaches [4,5,6,8]. Among these, normalized cuts [22] formulate segmentation as a graph partitioning problem based on pairwise affinities and form the basis of our approach.

**Diffusion-based Methods.** DiffCut [11] extracts self-attention features from pretrained diffusion models and applies recursive normalized cuts for zero-shot segmentation. ADZUS [16] similarly utilizes diffusion self-attention representations for biomedical image segmentation.

**Representation-based Methods.** CUTS [19] learns patch-level representations using contrastive learning and diffusion condensation to generate hierarchical segmentations. Unlike our approach, it relies on learned representations and operates on 2D slices.

**Pretrained Generative Models.** MAISI [15] is a 3D latent diffusion framework for medical image synthesis. Its volumetric design enables the learning of rich anatomical representations from CT and MRI data.

### 3 Proposed Method

The proposed method consists of two main components (Fig. 1): (1) multi-representation feature construction using a pretrained generative model [15], and (2) segmentation via 3D recursive N-Cuts. Given a 3D input image  $I$ , a VAE is first used to encode the image into a latent representation  $A$ . Gaussian noise is then injected at a fixed noise level into the latent representation to obtain a noisy latent. This noisy latent is passed through a diffusion U-Net in a single forward pass, producing diffusion features  $B$ . In parallel, self-attention weights are extracted from the self-attention (SA) blocks of the network. These attention maps are averaged across attention heads, reshaped into spatial feature maps, and further reduced using principal component analysis (PCA) to obtain attention-derived features  $C$ . Finally, raw intensity  $I$ , latent representations  $A$ , diffusion features  $B$ , and attention-derived features  $C$  are resampled to a common voxel grid. These feature representations, together with spatial coordinates, are then integrated to construct an affinity graph. Segmentation is subsequently performed using recursive N-Cuts to generate the final segmentation masks.

#### 3.1 Latent Representation using VAE

We use the encoder of a pretrained VAE [15] to map an input volume  $I \in \mathbb{R}^{H \times W \times D}$  to a compact latent representation  $A \in \mathbb{R}^{C \times h \times w \times d}$ . The latent representation preserves anatomical information while reducing spatial resolution, and is used for both affinity graph construction and diffusion feature extraction.

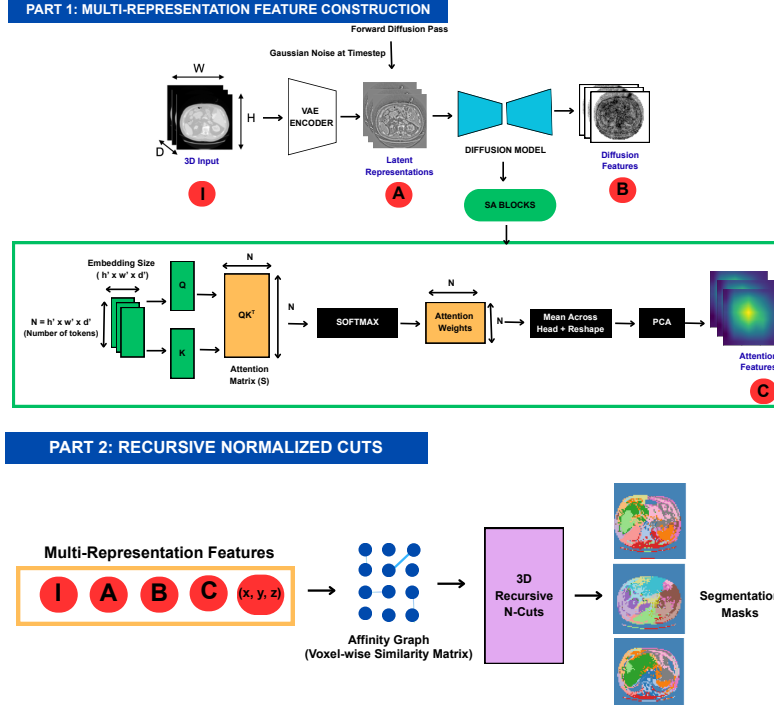
#### 3.2 Diffusion-Based Feature Extraction

Given the latent representation  $A \in \mathbb{R}^{C \times h \times w \times d}$ , we repurpose a pretrained MAISI latent diffusion model [15] as a feature extractor. Gaussian noise is injected at a fixed noise level  $t \in [0, 1]$  to obtain a noisy latent:

$$z_t = (1 - t)A + t\alpha\epsilon, \quad \epsilon \sim \mathcal{N}(0, I), \quad (1)$$

where  $\alpha$  controls the noise magnitude.

The noisy latent  $z_t$  is passed through the pretrained diffusion U-Net in a single forward pass. Instead of performing iterative denoising, we extract the feature map from the fourth decoder upsampling block. This feature map is used as the diffusion representation  $B \in \mathbb{R}^{C' \times h' \times w' \times d'}$ .



**Fig. 1.** Overview of the proposed framework. Latent representations (Sec. 3.1), diffusion features (Sec. 3.2), self-attention features (Sec. 3.3), image intensity, and spatial coordinates are integrated to construct an affinity graph, followed by recursive normalized cuts for segmentation (Sec. 3.4).

### 3.3 Self-Attention Feature Extraction

Using the pretrained diffusion U-Net, we derive self-attention features from a selected self-attention layer. Following the standard attention formulation, attention weights are computed as

$$S = \text{softmax} \left( \frac{QK^T}{\sqrt{d}} \right), \quad (2)$$

where  $S$  denotes the multi-head attention weights. The attention maps are averaged across heads and symmetrized to capture mutual relationships between tokens. The symmetrized attention matrix is then reshaped into token-wise spatial feature maps, reduced using PCA, and upsampled to the target spatial size to form the attention representation  $C$ .

### 3.4 Segmentation via Recursive Normalized Cuts

We formulate segmentation as a graph partitioning problem using N-Cuts on a 3D voxel-wise graph, where each voxel is treated as a node and pairwise affinities

define edge weights. Affinities are constructed from the feature representations described in Sec. 3.1–3.3, including image intensity  $I$ , latent representation  $A$ , diffusion feature  $B$ , and attention feature  $C$ .

All feature representations are resampled to a common voxel grid. A foreground mask is applied to select valid voxels, resulting in feature matrices  $X_i \in \mathbb{R}^{C_i \times M}$  for each feature type  $i \in \{I, A, B, C\}$ . Spatial coordinates  $X_{xyz} \in \mathbb{R}^{3 \times M}$  are incorporated as additional features. The augmented feature representation is defined as

$$\tilde{X}_i = [X_i; \alpha_i X_{xyz}], \quad (3)$$

where  $\alpha_i$  controls the contribution of spatial information.

Affinity matrices are computed using cosine similarity:

$$W_{pq}^{(i)} = \frac{1}{2} \left( \cos(\tilde{x}_p^{(i)}, \tilde{x}_q^{(i)}) + 1 \right), \quad (4)$$

where  $\tilde{x}_p^{(i)}$  and  $\tilde{x}_q^{(i)}$  denote the augmented feature vectors of voxels  $p$  and  $q$  in  $\tilde{X}_i$ , respectively.

The final affinity matrix is obtained by weighted fusion:

$$W = w_I W^{(I)} + w_A W^{(A)} + w_B W^{(B)} + w_C W^{(C)}. \quad (5)$$

where  $w_I$ ,  $w_A$ ,  $w_B$ , and  $w_C$  are fusion weights.

The symmetric normalized graph Laplacian is then computed as

$$L_{\text{sym}} = I - D^{-1/2} W D^{-1/2}, \quad (6)$$

where  $D$  is the degree matrix with diagonal entries  $D_{pp} = \sum_q W_{pq}$  for voxel  $p$ .

Segmentation is obtained from the eigenvector corresponding to the second smallest eigenvalue of  $L_{\text{sym}}$  [22]. Recursive bipartitioning is applied until a minimum region size or maximum recursion depth is reached, yielding the final segmentation mask.

## 4 Experiments and Results

### 4.1 Datasets and Preprocessing

We evaluate the proposed framework on two public medical imaging datasets: BraTS2020 [21,2,3] and FLARE22 [20]. For BraTS2020, we use the T2-FLAIR modality and perform binary segmentation by merging all tumour subregions into a single foreground class. For FLARE22, we use abdominal CT scans for multi-organ segmentation. All volumes are preprocessed following the MAISI protocol [15], including orientation standardization, modality-specific intensity normalization, and spatial resampling. Volumes are resized to a fixed resolution of  $256 \times 256 \times 256$  for BraTS2020 and  $256 \times 256 \times 128$  for FLARE22.

## 4.2 Implementation Details

We use a pretrained MAISI model [15] for feature extraction, with the VAE encoder and diffusion U-Net kept frozen during inference. A diffusion noise timestep of  $t = 0.2$  is used. Diffusion features are extracted from the fourth decoder upsampling block, while self-attention features are extracted from the third downsampling block. PCA dimensionality is set to 512 for BraTS2020 and 64 for FLARE22. Spatial coordinates are incorporated into all feature representations with a weighting factor of 0.01. The fusion weights are set to  $(w_I, w_A, w_B, w_C) = (1.0, 0.5, 1.0, 0.2)$ . All feature representations are resampled to a common spatial resolution of  $64 \times 64 \times 64$  for graph construction. Segmentation is performed using recursive normalized cuts with a maximum recursion depth of 5 and a minimum region size of 50 voxels. All hyperparameters were selected based on ablation experiments and fixed throughout the experiments.

## 4.3 Post-processing and Evaluation

After recursive N-Cuts segmentation (Sec. 3.4), post-processing is applied to suppress small fragmented regions and reduce segmentation noise. For BraTS2020, connected component filtering and a  $3 \times 3 \times 3$  mode filter are applied. For FLARE22, a  $2 \times 2 \times 2$  mode filter is used for lightweight smoothing. The resulting segmentation masks are then upsampled to the original image resolution. Ground-truth annotations are used exclusively for evaluation and cluster-to-label matching, and are not involved in any stage of the segmentation process. Since cluster identities are arbitrary in unsupervised segmentation, predicted segments are matched to reference labels for metric computation. For BraTS2020, all tumour subregions are merged into a binary lesion mask, and clusters are matched based on overlap. For FLARE22, connected components are matched to organ labels using a greedy Dice-based assignment strategy.

## 4.4 Representation Analysis

Table 1 shows that the contribution of each representation varies across datasets. On BraTS2020, image intensity alone achieves strong performance, reflecting the distinctive appearance of tumour regions in MRI. In contrast, no single representation performs consistently well on FLARE22, indicating that abdominal CT segmentation requires more diverse anatomical cues. Integrating multiple representations generally improves performance over individual representations. Full integration achieves the best results on FLARE22 while remaining competitive with Intensity + VAE + Diffusion on BraTS2020. Since our objective is to develop a unified framework that generalises across imaging modalities without dataset-specific tuning, the fully integrated configuration is adopted because it provides the most consistent performance across both datasets.

#### 4.5 Segmentation Results

We evaluate segmentation performance on two datasets against three baselines using Dice score (Dice; unitless in  $[0, 1]$ ), Intersection over Union (IoU; unitless in  $[0, 1]$ ), and Hausdorff distance (HD; mm). Higher Dice and IoU indicate better overlap with the reference segmentation, whereas lower HD indicates more accurate boundary localisation. We consider both 2D and 3D evaluation protocols. For our method and the intensity-based baseline, cluster-to-ground-truth matching is performed in both slice-wise and volume-wise settings. The intensity-based baseline applies 3D normalized cuts directly to the input image intensity and spatial coordinates, without using VAE, diffusion, or self-attention features. For DiffCut and CUTS, slice-wise matching is first performed, followed by global relabeling across adjacent slices using Hungarian matching [17] based on Dice overlap. Reconstructed volumes are then evaluated using both 2D and 3D metrics. All results are reported as mean  $\pm$  standard deviation, where  $\uparrow$  and  $\downarrow$  denote higher-is-better and lower-is-better performance, respectively.

Table 2 and Fig. 2 present the quantitative and qualitative results on the BraTS2020 dataset. Our method achieves the best performance under the 3D evaluation setting across all metrics, indicating improved volumetric consistency compared with the baselines. Although the intensity-based baseline performs competitively under 2D evaluation, incorporating complementary representations improves consistency across slices, resulting in higher Dice and IoU scores together with lower Hausdorff distance under 3D evaluation. DiffCut and CUTS show consistently lower performance across all metrics, likely because their slice-wise formulation does not explicitly exploit 3D spatial information.

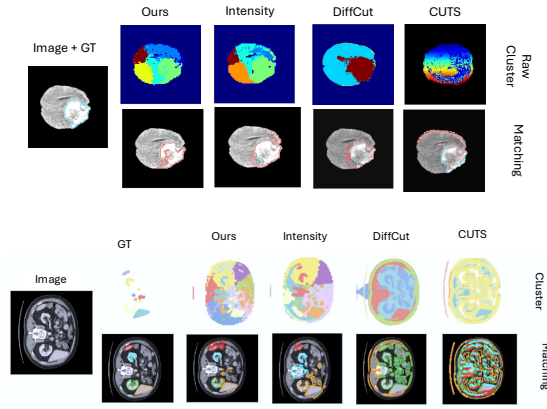
To assess generalisation, we further evaluate the proposed framework on the FLARE22 abdominal CT dataset using the same experimental setting. Table 3 presents the quantitative results under the 3D evaluation protocol. Overall performance is lower than on BraTS2020, reflecting the greater difficulty of unsupervised abdominal CT segmentation, where organs often exhibit similar intensity distributions and ambiguous boundaries.

Our method achieves the best overall performance, obtaining the highest average Dice and IoU scores across organs. While DiffCut performs better on some large organs, such as the liver and stomach, our method shows stronger performance on smaller and anatomically complex structures, including the adrenal glands and spleen. This suggests that integrating multiple representations provides complementary anatomical cues that are beneficial for challenging multi-organ segmentation tasks. Qualitative examples are shown in Fig. 2.

CUTS consistently achieves lower performance under our evaluation protocol, possibly because its representations are learned from a specific training dataset and applied in a zero-shot manner. Overall, the results across BraTS2020 and FLARE22 demonstrate that the proposed framework generalises across MRI and CT datasets under a unified unsupervised setting.

**Table 1.** Representation analysis. Segmentation performance is reported as the average Dice score over 10 patients for each dataset.

| Representation                                             | BraTS2020 $\uparrow$ | FLARE22 $\uparrow$ |
|------------------------------------------------------------|----------------------|--------------------|
| Intensity                                                  | 0.709                | 0.048              |
| VAE                                                        | 0.288                | 0.077              |
| Diffusion                                                  | 0.270                | 0.072              |
| Attention                                                  | 0.431                | 0.046              |
| Intensity + VAE                                            | 0.715                | 0.091              |
| Intensity + VAE + Diffusion                                | <b>0.721</b>         | 0.087              |
| Intensity + VAE + Diffusion + Attention (Full Integration) | 0.703                | <b>0.092</b>       |

**Fig. 2.** Qualitative results on BraTS2020 and FLARE22. For BraTS2020, **blue** contours denote ground truth and **red** contours denote matched predictions. For FLARE22, organ boundaries are shown in different colors. Our method produces more coherent segmentations than the baselines.

## 5 Conclusion

In this work, we propose an unsupervised framework for 3D medical image segmentation that combines multiple feature representations with recursive normalized cuts. The proposed framework integrates image intensity, VAE latent representations, diffusion features, and self-attention features for unsupervised segmentation. Experiments demonstrate generalisation across imaging modalities and show that integrating complementary representations improves segmentation performance, particularly for anatomically complex structures. We further observe a trade-off between preserving global anatomical structures and capturing fine-grained details. Future work will investigate organ-aware adaptive partitioning strategies and evaluate the framework on additional medical imaging datasets.

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

**Table 2.** Segmentation performance on BraTS2020 (30 subjects).

| Method      | 2D evaluation                     |                                   |                                  | 3D evaluation                     |                                   |                                  |
|-------------|-----------------------------------|-----------------------------------|----------------------------------|-----------------------------------|-----------------------------------|----------------------------------|
|             | Dice $\uparrow$                   | IoU $\uparrow$                    | HD $\downarrow$                  | Dice $\uparrow$                   | IoU $\uparrow$                    | HD $\downarrow$                  |
| <b>Ours</b> | <b>0.597<math>\pm</math>0.137</b> | 0.442 $\pm$ 0.135                 | <b>22.79<math>\pm</math>5.58</b> | <b>0.704<math>\pm</math>0.136</b> | <b>0.558<math>\pm</math>0.152</b> | <b>26.44<math>\pm</math>6.95</b> |
| Intensity   | <b>0.597<math>\pm</math>0.137</b> | <b>0.443<math>\pm</math>0.138</b> | 23.50 $\pm$ 5.42                 | 0.692 $\pm$ 0.136                 | 0.544 $\pm$ 0.149                 | 28.95 $\pm$ 8.60                 |
| DiffCut     | 0.339 $\pm$ 0.189                 | 0.269 $\pm$ 0.182                 | 69.77 $\pm$ 21.91                | 0.259 $\pm$ 0.143                 | 0.157 $\pm$ 0.098                 | 82.77 $\pm$ 13.94                |
| CUTS        | 0.311 $\pm$ 0.211                 | 0.241 $\pm$ 0.190                 | 69.45 $\pm$ 22.44                | 0.281 $\pm$ 0.176                 | 0.176 $\pm$ 0.131                 | 84.21 $\pm$ 13.97                |

**Table 3.** Per-organ segmentation performance on FLARE22 (3D evaluation) over 38 subjects.

| Organ          | Ours            |                |                 | Intensity       |                |                 | DiffCut         |                |                 | CUTS            |                |                 |
|----------------|-----------------|----------------|-----------------|-----------------|----------------|-----------------|-----------------|----------------|-----------------|-----------------|----------------|-----------------|
|                | Dice $\uparrow$ | IoU $\uparrow$ | HD $\downarrow$ | Dice $\uparrow$ | IoU $\uparrow$ | HD $\downarrow$ | Dice $\uparrow$ | IoU $\uparrow$ | HD $\downarrow$ | Dice $\uparrow$ | IoU $\uparrow$ | HD $\downarrow$ |
| Liver          | 0.666           | 0.501          | 92.522          | 0.657           | 0.492          | 93.363          | <b>0.745</b>    | <b>0.596</b>   | <b>91.874</b>   | 0.264           | 0.155          | 177.752         |
| Right Kidney   | 0.287           | 0.182          | 54.418          | <b>0.318</b>    | <b>0.202</b>   | <b>54.346</b>   | 0.232           | 0.134          | 112.176         | 0.051           | 0.026          | 185.078         |
| Spleen         | <b>0.489</b>    | <b>0.342</b>   | <b>51.800</b>   | 0.426           | 0.283          | 59.835          | 0.157           | 0.089          | 109.293         | 0.035           | 0.018          | 204.861         |
| Pancreas       | 0.102           | 0.057          | <b>46.262</b>   | 0.100           | 0.057          | 47.559          | <b>0.160</b>    | <b>0.089</b>   | 69.222          | 0.053           | 0.028          | 106.982         |
| Aorta          | 0.125           | 0.070          | <b>58.202</b>   | 0.067           | 0.036          | 62.921          | <b>0.172</b>    | <b>0.095</b>   | 77.477          | 0.001           | 0.000          | 185.949         |
| IVC            | <b>0.058</b>    | <b>0.031</b>   | <b>55.354</b>   | 0.027           | 0.014          | 60.599          | 0.050           | 0.026          | 72.062          | 0.003           | 0.001          | 173.019         |
| Right Adrenal  | <b>0.031</b>    | <b>0.017</b>   | <b>16.848</b>   | 0.025           | 0.014          | 18.546          | 0.005           | 0.003          | 86.869          | 0.005           | 0.003          | 120.705         |
| Left Adrenal   | 0.039           | 0.022          | <b>14.026</b>   | <b>0.052</b>    | <b>0.029</b>   | 18.415          | 0.003           | 0.001          | 60.841          | 0.005           | 0.003          | 129.172         |
| Gallbladder    | 0.062           | 0.037          | 39.982          | 0.027           | 0.016          | <b>35.853</b>   | <b>0.076</b>    | <b>0.044</b>   | 109.441         | 0.060           | 0.032          | 142.078         |
| Esophagus      | 0.037           | 0.020          | 40.287          | 0.042           | 0.024          | <b>21.580</b>   | <b>0.071</b>    | <b>0.038</b>   | 64.933          | 0.002           | 0.001          | 175.393         |
| Stomach        | 0.222           | 0.130          | 54.069          | 0.219           | 0.127          | <b>52.200</b>   | <b>0.396</b>    | <b>0.250</b>   | 87.749          | 0.042           | 0.022          | 184.195         |
| Duodenum       | 0.051           | 0.027          | <b>46.794</b>   | 0.050           | 0.027          | 46.890          | <b>0.082</b>    | <b>0.044</b>   | 66.802          | 0.022           | 0.011          | 144.506         |
| Left Kidney    | 0.344           | 0.226          | <b>45.050</b>   | <b>0.388</b>    | <b>0.261</b>   | 57.369          | 0.242           | 0.139          | 107.723         | 0.076           | 0.040          | 160.929         |
| <b>Average</b> | <b>0.193</b>    | <b>0.128</b>   | <b>47.355</b>   | 0.184           | 0.122          | 48.421          | 0.184           | 0.119          | 85.882          | 0.048           | 0.026          | 160.817         |

## References

1. Acharjya, P.P., Das, R., Ghoshal, D.: Study and comparison of different edge detectors for image segmentation. Global Journal of Computer Science and Technology (2012)
2. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., et al.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. Scientific Data **4**, 170117 (2017). <https://doi.org/10.1038/sdata.2017.117>
3. Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., et al.: Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. arXiv preprint arXiv:1811.02629 (2018)
4. Benkarim, O.M., Piella, G., González Ballester, M., Sanroma, G.: Discriminative confidence estimation for probabilistic multi-atlas label fusion. Medical Image Analysis **42**, 274–287 (2017)
5. Bhagwat, N., Pipitone, J., Winterburn, J.L., Guo, T., Duerden, E.G., Voineskos, A.N., Lepage, M., Miller, S.P., Pruessner, J.C., Chakravarty, M.M.: Manual-protocol inspired technique for improving automated mr image segmentation during label fusion. Frontiers in Neuroscience **10**, 325 (2016)
6. Bijar, A., Khayati, R., Peñalver Benavent, A., Benavent, A.P.: Increasing the contrast of the brain mr flair images using fuzzy membership functions and structural similarity indices in order to segment ms lesions. PLoS One **8**(6), e65469 (2013)

7. Bougourzi, F., Hadid, A.: Recent advances in medical imaging segmentation: A survey. *arXiv preprint arXiv:2505.09274* (2025)
8. Brosch, T., Tang, L.Y.W., Yoo, Y., et al.: Deep 3d convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation. *IEEE Transactions on Medical Imaging* **35**(5), 1229–1239 (2016)
9. Chen, X., Pan, L.: A survey of graph cuts/graph search based medical image segmentation. *IEEE Reviews in Biomedical Engineering* **11**, 112–124 (2018)
10. Chen, Z., Nan, X., Li, J., Zhao, J., Li, H., Lin, Z., Li, H., Chen, H., Liu, Y., Tang, L., et al.: Pam: a propagation-based model for segmenting any 3d objects across multi-modal medical images. *npj Digital Medicine* (2025)
11. Couairon, P., Shukor, M., Haugeard, J.E., Cord, M., Thome, N.: Diffcut: Catalyzing zero-shot semantic segmentation with diffusion features and recursive normalized cut. *Advances in Neural Information Processing Systems* **37**, 13548–13578 (2024)
12. El Naqa, I., Yang, D., Apte, A., Khullar, D., Mutic, S., Zheng, J., Bradley, J.D., Grigsby, P., Deasy, J.O.: Concurrent multimodality image segmentation by active contours for radiotherapy treatment planning. *Medical Physics* **34**(12), 4738–4749 (2007)
13. Fu, J., Li, H., Lu, T., Zhang, S., Wang, G.: Um-sam: Unsupervised medical image segmentation using knowledge distillation from segment anything model. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 616–626. Springer (2025)
14. Gao, Y., Jiang, Y., Peng, Y., Yuan, F., Zhang, X., Wang, J.: Medical image segmentation: A comprehensive review of deep learning-based methods. *Tomography* **11**(5), 52 (2025)
15. Guo, P., Zhao, C., Yang, D., Xu, Z., Nath, V., Tang, Y., Simon, B., Belue, M., Harmon, S., Turkbey, B., et al.: Maisi: Medical ai for synthetic imaging. In: *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. pp. 4430–4441. IEEE (2025)
16. Hamrani, A., Godavarty, A.: Self-attention diffusion models for zero-shot biomedical image segmentation: Unlocking new frontiers in medical imaging. *Bioengineering* **12**(10), 1036 (2025)
17. Kuhn, H.W.: The hungarian method for the assignment problem. *Naval research logistics quarterly* **2**(1-2), 83–97 (1955)
18. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. *Medical image analysis* **42**, 60–88 (2017)
19. Liu, C., Amodio, M., Shen, L.L., Gao, F., Avesta, A., Aneja, S., Wang, J.C., Del Priore, L.V., Krishnaswamy, S.: Cuts: A deep learning and topological framework for multigranular unsupervised medical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 155–165. Springer (2024)
20. Ma, J., Zhang, Y., Gu, S., Ge, C., Ma, S., Young, A., Zhu, C., Yang, X., Meng, K., Huang, Z., Zhang, F., Pan, Y., Huang, S., Wang, J., Sun, M., Zhang, R., Jia, D., Choi, J.W., Alves, N., de Wilde, B., Koehler, G., Lai, H., Wang, E., Wiesenfarth, M., Zhu, Q., Dong, G., He, J., He, J., Yang, H., Huang, B., Lyu, M., Ma, Y., Guo, H., Xu, W., Maier-Hein, K., Wu, Y., Wang, B.: Unleashing the strengths of unlabelled data in deep learning-assisted pan-cancer abdominal organ quantification: the flare22 challenge. *The Lancet Digital Health* **6**(11), e815–e826 (2024). [https://doi.org/10.1016/S2589-7500\(24\)00154-7](https://doi.org/10.1016/S2589-7500(24)00154-7)

21. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., et al.: The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging* **34**(10), 1993–2024 (2015). <https://doi.org/10.1109/TMI.2014.2377694>
22. Shi, J., Malik, J.: Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence* **22**(8), 888–905 (2000)