Deep Spectral Methods for Unsupervised Ultrasound Image Interpretation

Oleksandra Tmenova[1*], Yordanka Velikova[1,2*], Mahdi Saleh[1], and Nassir Navab[1,2]

[1] Computer Aided Medical Procedures, Technical University of Munich, Germany
[2] Munich Center for Machine Learning, Munich, Germany

Abstract. Ultrasound imaging is challenging to interpret due to non-uniform intensities, low contrast, and inherent artifacts, necessitating extensive training for non-specialists. Advanced representation with clear tissue structure separation could greatly assist clinicians in mapping underlying anatomy and distinguishing between tissue layers. Decomposing an image into semantically meaningful segments is mainly achieved using supervised segmentation algorithms. Unsupervised methods are beneficial, as acquiring large labeled datasets is difficult and costly, but despite their advantages, they still need to be explored in ultrasound. This paper proposes a novel unsupervised deep learning strategy tailored to ultrasound to obtain easily interpretable tissue separations. We integrate key concepts from unsupervised deep spectral methods, which combine spectral graph theory with deep learning methods. We utilize self-supervised transformer features for spectral clustering to generate meaningful segments based on ultrasound-specific metrics and shape and positional priors, ensuring semantic consistency across the dataset. We evaluate our unsupervised deep learning strategy on three ultrasound datasets, showcasing qualitative results across anatomical contexts without label requirements. We also conduct a comparative analysis against other clustering algorithms to demonstrate superior segmentation performance, boundary preservation, and label consistency.

Keywords: Spectral Methods · Unsupervised Learning · Ultrasound.

1 Introduction

Ultrasound is commonly used in diagnostic medicine, valued for its real-time imaging capabilities and non-invasive nature, which enables regular health check-ups and screenings without ionizing radiation [20]. However, despite its advantages, the interpretation of ultrasound images often presents a significant challenge, necessitating specialized training or years of experience for clinicians [3]. The complexity of these images makes the apparent separation and identification of tissue structures difficult. Improved representation techniques can aid in the

*Shared first authorship.

interpretation process as understanding the underlying anatomy and differentiation between tissue layers is complex

Decomposing images into semantically meaningful regions has predominantly been tackled using supervised deep learning (DL) algorithms for segmentation [16]. Particularly, convolutional neural networks (CNNs) and architectures like U-net [22] have significantly advanced supervised ultrasound segmentation [25]. These methods excel in delineating target anatomical structures and have shown promise in automating the identification of structures across a range of screening applications, leading to improved diagnostic accuracy [16]. Despite their efficacy, these methods depend on the availability of large, annotated datasets for training and are often tailored to specific anatomical structures, limiting their scalability and adaptability [24]. Consequently, unsupervised learning approaches, which do not necessitate expert-reliant labeled data for training, emerge as an alternative.

Relavant works utilize graph-based methods [8,23], gradient-ascent-based algorithms [26], SLIC-K-means-based methods [1] and intermediate representations [28,27]. Such techniques have proven helpful for computer vision tasks like semantic instance segmentation and have found their application in the ultrasound domain too [11,10]. However, they demand careful parameter selection to avoid the loss of critical edge information [19]. Furthermore, the resultant segments have class-agnostic labels and are primarily used as an initial step for further DL-based frameworks.

Following traditional clustering methods, spectral clustering emerges as another unsupervised approach for identifying image segments. It is done by constructing a similarity graph representing the relationships between data points. For spectral clustering, an affinity matrix is built, and by utilizing eigenvalues and eigenvectors, similar data points are grouped into clusters [29]. Spectral clustering excels in handling complex cluster shapes that are non-convex or consist of disjoint convex sets, making it particularly advantageous for applications where conventional clustering methods fail. When applied to images where the affinity matrix mirrors the adjacency matrix of a graph, spectral clustering is used to identify normalized graph cuts, which can divide images into meaningful segments without the need for predefined labels [23].

Recent works leverage the strengths of self-supervised Vision Transformer (ViT) models, such as DINO [5], which utilize self-distillation techniques to learn rich visual feature representations from unlabeled data. Those features are then applied to spectral clustering techniques to construct an affinity matrix and identify distinct segments within an image [31,32,17,30]. In particular, deep spectral segmentation (DSS) [17] employs multiple eigenvectors to obtain per-image segments and introduces additional spatial and color affinities for improved consistency. It then utilizes DINO features from all dataset segments and clusters them to obtain semantic labels. Combining self-supervised ViT features with spectral clustering has become a powerful approach for unsupervised object discovery and segmentation. Those unsupervised segmentation techniques have drawn attention for their ability to provide label-free representations eas-

ily adaptable for downstream tasks [2], making them particularly suitable for medical imaging applications where labeled data is scarce.

Contributions This work introduces an unsupervised deep-learning framework specifically designed for enhancing ultrasound image analysis. Utilizing self-supervised transformer-based features, we implement spectral clustering to derive semantically meaningful segments. We incorporate ultrasound-specific metrics together with shape and geometric priors to ensure consistency across diverse anatomical contexts. This provides clear tissue structure separation without the need for labeled datasets. We validate our framework across three ultrasound datasets, showcasing its capability, and provide qualitative results that adeptly preserve the contours of the underlying anatomical structures. Our comparative analysis with other clustering algorithms underscores our method's superior segmentation accuracy, boundary preservation, and label consistency performance. The source code is publicly available at `https://github.com/alexaatm/UnsupervisedSegmentor4Ultrasound.git` [1]

2 Method

Our approach builds upon the deep spectral family of unsupervised segmentation methods [32,17,30], particularly the deep spectral segmentation (DSS) for multiple-object semantic segmentation [17]. The proposed method's architecture, shown in Figure 1, includes two major steps: spectral decomposition for obtaining per-image segments and clustering them into semantically consistent classes. As an addition to the duo of self-supervised transformers with classic spectral clustering [31,17,30], we propose several adaptations to enhance segment separation in ultrasound images. In the first step (Figure 1, top), we introduce ultrasound affinities and add a preprocessing step to address the domain gap between natural and ultrasound data. In the second step (Figure 1, bottom), we incorporate additional shape and position priors to add extra information to the final clustering of segments.

2.1 Spectral decomposition

Data Preprocessing Different from real-world images with diverse colors and distinct borders, ultrasound data is infamously challenging to analyze. That is why US image analysis benefits from proper preprocessing [18,6]. To take this into account, we add a preprocessing block to the pipeline and explore different strategies for enhancing the image quality, including classical approaches (gaussian blurring, histogram equalization) and pretrained denoising models like MPRNet [33].

[1] All implementation and experiments were conducted by O. Tmenova as part of her master's thesis at TUM.

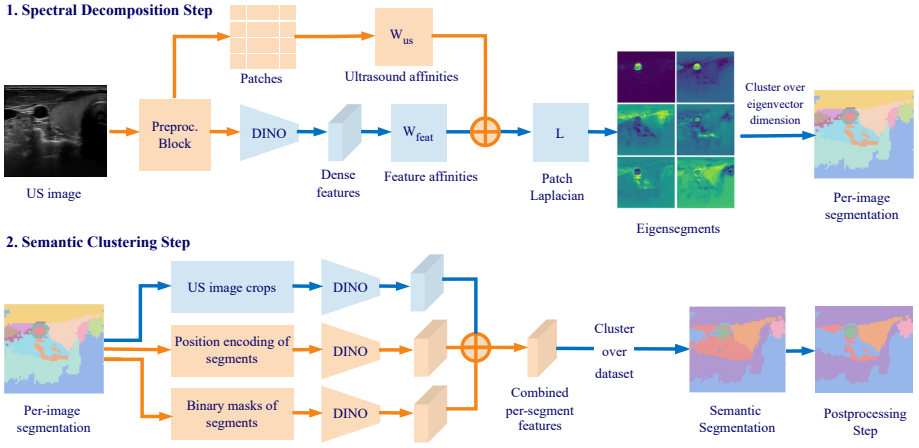

Fig. 1. In our unsupervised semantic segmentation pipeline, ultrasound images undergo preprocessing and dense feature extraction to derive feature affinities. Ultrasound-specific affinities are then calculated using similarity metrics (MI, SSD) and combined with initial affinities for spectral clustering, yielding pseudo masks. Subsequently, images are cropped to focus on detected segments, and dense features alongside positional and shape priors refine clustering across the dataset. This two-step process enhances semantic consistency, transitioning from class-agnostic to more meaningful segmentations, all without relying on labels.

Affinity Matrix Construction Self-supervised attention-based architectures like DINO [5] serve as a good base for extracting rich features. Following [17], we use the features from the keys of the last attention layer of the pre-trained DINO. An essential step in spectral clustering is treating image segmentation as a graph-cutting problem [23]. Images are represented as graphs $G = (V, E)$ where nodes correspond to either pixels (for color affinities) or patches (for DINO affinities). Edge weights between nodes indicate their similarity. The self-correlation of DINO features provides an effective affinity matrix, enabling successful graph partitioning and meaningful image segments. Like in DSS [17], the features are thresholded at 0 to exclude anti-correlations:

$$W_{\text{feat}} = \mathrm{f} \cdot \mathrm{f}^T \odot (\mathrm{f} \cdot \mathrm{f}^T > 0) \tag{1}$$

Since color affinities cannot be leveraged from ultrasound greyscale data, we integrate ultrasound patch-wise affinities employing standard pixel-based metrics that proved successful in the task of both rigid and non-rigid ultrasound image registration [6]. In particular, we employ two common metrics: Sum of Squared Differences (SSD) $SSD(P_1, P_2) = \sum_{i=1}^{X} \sum_{j=1}^{Y} (P_1(i,j) - P_2(i,j))^2$ where X and Y represent the dimensions of the patches P_1 and P_2, and Mutual Information (MI) $MI(P_1, P_2) = (H(P_1) + H(P_2))/(H(P_1, P_2))$, where $H(P_1)$ and $H(P_2)$ are the entropies of the individual patches, and $H(P_1, P_2)$ is their joint entropy. To

build the affinity matrix, an image is partitioned into patches of size $k \times k$. We chose k to match the patch size of the used transformer backbone.

The dissimilarity matrix $D_{\text{patchwise}}$ is then constructed by comparing each patch P_i to every other patch P_j using the specified distance metric d (Eq. 2). It is then transformed into an affinity matrix using a Gaussian kernel (Eq. 3).

$$D_{\text{patchwise}}(P_i, P_j) = d(P_i, P_j) = \begin{cases} \text{SSD}(P_i, P_j), & \text{if } d = \text{SSD} \\ 1 - \text{MI}(P_i, P_j), & \text{if } d = \text{MI} \end{cases}, \tag{2}$$

$$W_{\text{patchwise}} = \exp\left(-\delta \cdot D_{\text{patchwise}}\right) \tag{3}$$

Additionally, we explore position-based affinities using linear interpolation from 0 to 1 for the N_{height} and N_{width}, where $N_{height} = H//k$, $N_{width} = W//k$, where k is the size of a patch, which results in patch feature vectors $\psi(u) = (x_{\text{pos}}, y_{\text{pos}})$, which are then used to construct a positional affinity matrix (Eq. 4).

$$W_{\text{pos}}(P_i, P_j) = \begin{cases} 1 - \|\psi(P_i) - \psi(P_j)\|, & \text{if } P_i \in \text{KNN}_\psi(P_j), \\ 0, & \text{otherwise}, \end{cases} \tag{4}$$

where $P_i \in \text{KNN}_\psi(P_j)$ are the k-nearest neighbors of patch P_j under the SSD distance of feature vectors ψ. Finally, we linearly combine DINO, ultrasound, and positional affinities, controlled by coefficients $C_{feat}, C_{mi}, C_{pos}$, to obtain the final affinity matrix needed for spectral clustering (Eq. 5).

$$W_{\text{comb}} = W_{\text{feat}} + C_{\text{ssd}} \cdot W_{\text{ssd}} + C_{\text{mi}} \cdot W_{\text{mi}} + C_{\text{pos}} \cdot W_{\text{pos}}, \tag{5}$$

Spectral Clustering From the obtained affinity matrix W_{comb}, its Laplacian matrix is calculated (Eq. 6). Then the objective function for spectral clustering can be expressed using the graph Laplacian: $\min \text{Tr}(E^\top L E)$ s.t. $E^\top E = I$, where Tr denotes the trace norm of a matrix, and $E = \{a_{ij}\}$ is a matrix whose rows represent the low-dimensional embedding of the original data points.

$$L = D^{-1/2}(D - W)D^{-1/2}, \text{ where } D \text{ has values } d_{ii} = \sum_j a_{ij} \quad \text{for all } i \tag{6}$$

The Laplacian matrix is decomposed into eigensegments, $e_0, \ldots, e_{n-1}$, where only positive eigenvectors ($e > 0$) are used as per-image segments. K-means clustering is then applied to obtain these segments, following the approach in DSS [17]. We refer to this step as Oversegmentation, with the number of eigensegments set to 15.

2.2 Semantic Clustering

In the second clustering step, bounding boxes of segments are calculated to extract per-segment features, which are then clustered using K-means [17]. We refine this process for ultrasound data and optimize the segment feature extraction step by addressing the challenge of textural similarity in different anatomical

areas. To achieve this, we employ a dual embedding strategy that enhances the differences between features of segments, such as vessels and features of other areas with similar textures, while minimizing the overall segment count. We construct a mask embedding to capture shape features via binary masks and positional embedding to encode spatial locations of segments. This streamlined approach ensures that segments are grouped not only by similar features but also by shape and position, resulting in better spatial and structural consistency across ultrasound sweeps. The resulting feature vectors from the image crop $\mathrm{f}_{\mathrm{image}} = \phi(\mathrm{s}_{\mathrm{crop}})$, from the mask $\mathrm{f}_{\mathrm{mask}} = \phi(\mathrm{s}_{\mathrm{mask}})$, and from the position encoding $\mathrm{f}_{\mathrm{pos}} = \phi(\mathrm{s}_{\mathrm{pos}})$ are then linearly combined before clustering.

Postprocessing Results obtained after the two clustering steps can already serve as a coarse segmentation. However, for sharper boundaries, we include additional postprocessing. We upscale and apply CRF [13], as also commonly done in other segmentation pipelines [17,30,9].

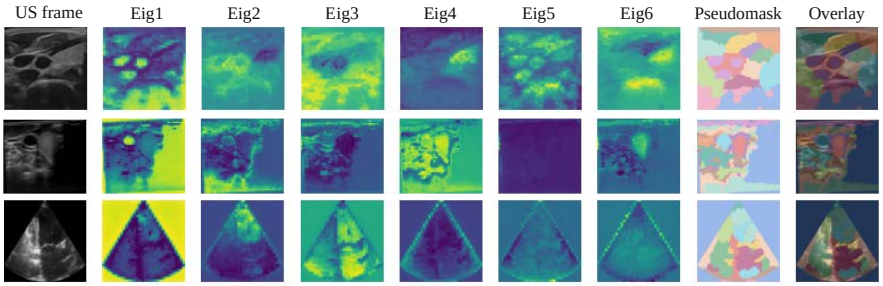

Fig. 2. Combining ultrasound-based affinities and deep features leads to meaningful image separation.

3 Experimental Setup

CCA Common Carotid Artery dataset consists of ultrasound images from four different machines from 24 adults with single labels of the carotid artery [4,21], which was sampled to remove repetitive slices, totaling 349 images for testing.
Thyroid dataset contains annotated 3D ultrasound images of the thyroid [14]. It includes scans from 28 healthy volunteers using a Siemens Acuson NX-3 US machine with a VF12-4 probe. The 3D ultrasound scans were post-processed to remove empty labels, extract the 2D slices with corresponding labels, and remove repetitive slices, in total 634 images.
CAMUS dataset includes 400 cardiac patient images for training and 50 for test [15]. For our evaluation, we used end-systole (ES) and end-diastole (ED) images from the test set - in total 500 validation images (5 for each ES and

ED) from 50 patients with 3 labels from manual expert annotations of the left ventricle endocardium, the myocardium and the left atrium.

Evaluation Methodology The evaluation methodology includes two main aspects: per-image mask evaluation (step I) and semantic evaluation post-clustering (step II). For step I, we assess the quality of individual segments before clustering using DICE score. For step II, we evaluate the segments obtained after semantic clustering. Ground truth and pseudo-labels are matched using Hungarian matching or majority vote [12,17], and only matched masks are evaluated. To assess semantic consistency of label mappings, we identify the most prevalent pseudo-label class assigned to each ground truth class across the entire dataset. The label consistency (LC) metric is then computed as the percentage of times the final pseudo-label class has been consistently assigned to a particular ground truth class across the entire dataset. We use DSS [17] as a baseline for comparisons, which aligns with our goal of multi-class segmentation. TokenCut [32] and CutLER [30], while conceptually similar, focus on single-object and instance segmentation, respectively, making direct comparisons challenging. Therefore, we focus on zero-shot unsupervised methods that segment images without prior training: SLIC [1] and FZ [8], baselines for superpixel evaluation, including ultrasound [7]. We report superpixel metrics such as Boundary Recall (BR) and Undersegmentation Error (UE) [19], setting the distance parameter d to 3 to accommodate the imprecise boundaries in ultrasound images.

4 Results and Discussion

In Tables 1 and 2, we compare the performance of our proposed method against the DSS baseline [17] with added preprocessing ($ours_{preproc}$), affinities ($ours_{aff}$), and their combined effect ($ours_{comb}$) in terms of the DICE score. In Table 2 we additionally assess their effect together with positional and mask priors in the semantic clustering step and evaluate label consistency (LC). Our proposed methods ($ours_{preproc}$, $ours_{aff}$ $ours_{comb}$) show improvements in segmentation quality compared to the baseline method [17] across all three datasets. Specifically, $ours_{aff}$ consistently achieves the highest DICE scores of 63.72 ± 14.31 for the Carotid dataset, 62.52 ± 8.62 for the Thyroid dataset, and 45.32 ± 9.16 for Cardiac dataset. The improvement in segmentation quality suggests that the preprocessing steps and additional affinities positively enhance segmentation performance. Figure 2 depicts eigensegments obtained from combining ultrasound

Table 1. Comparison with baseline method STEP 1. *CRF postprocessing

Method	N seg	Carotid DICE, std	Thyroid DICE, std	Cardiac DICE, std
DSS baseline *	15	32.33 ± 11.38	43.75 ± 9.87	36.98 ± 8.49
Ours_{proc}*	15	56.31 ± 12.89	62.45 ± 10.91	42.13 ± 6.78
Ours_{Aff}*	15	**63.72 ± 14.31**	**62.52 ± 8.62**	40.44 ± 9.07
Ours_{comb}*	15	46.21 ± 8.43	61.43 ± 10.03	**45.32 ± 9.16**

Table 2. Evaluation on downstream task STEP II

Method	Carotid		Thyroid1		Cardiac	
	DICE, std	LC, std	DICE, std	LC, std	DICE, std	LC, std
DSS baseline	39.24 ± 9.1	**70.25 ±16.0**	39.57 ± 8.6	47.77 ± 15.4	26.12 ± 7.7	78.33 ± 16.5
Ours$_{proc} + DSS_{step2}$	32.25 ± 6.7	47.69 ± 14.8	50.44 ± 7.3	67.61 ± 9.8	25.38 ± 7.3	77.06 ± 10.8
Ours$_{Aff} + DSS_{step2}$	42.56 ± 10.0	55.72 ± 10.8	54.95 ± 10.4	**71.99 ±3.5**	**37.53 ±6.5**	85.00 ± 10.8
Ours$_{comb} + DSS_{step2}$	30.50 ± 7.3	55.50 ± 13.3	**59.86 ±8.7**	59.79 ± 8.9	29.25 ± 10.6	87.38 ± 11.3
Ours$_{proc} + Ours_{step2}$	32.32 ± 6.9	52.50 ± 14.6	50.26 ± 17.9	67.01 ± 10.1	26.82 ± 11.3	**93.75 ±8.2**
Ours$_{Aff} + Ours_{step2}$	**44.98 ±14.5**	52.18 ± 15.9	47.62 ± 11.1	63.77 ± 14.0	30.92 ± 9.4	81.56 ± 11.9
Ours$_{comb} + Ours_{step2}$	19.30 ± 11.2	45.14 ± 5.1	52.90 ± 10.7	74.63 ± 8.9	28.11 ± 13.3	85.71 ± 9.2

MI, SSD, and positional affinities (with coefficients 1.0, 1.0, and 0.1, respectively) and the resulting pseudo mask. It can be observed how different eigensegments capture distinct areas from the original image, for example, the vessels in the carotid image (top, Eig1), the thyroid lobe (middle, Eig4), or the heart chamber (bottom, Eig2), which then get assigned a distinct label. In Table 2, we observe the positive effects of preprocessing, affinities, and shape priors on semantic clustering, with DICE scores of 44.98 ± 14.5, 59.86 ± 8.7 and 37.53 ± 6.5, consistently outperforming the baseline. However, there is a trade-off between mask quality and label consistency: methods that preserve finer details (higher DICE) result in more complex and varied segment shapes, making it harder to achieve consistent clustering labels (lower label consistency) across similar structures.

Finally, we compare the best results from Tables 1 and 2 to SLIC [1] and Felzenszwalb [8], common baselines for superpixel evaluation. The results are reported in Tables 3 and 4. In Table 3, we observe that the performance of 'deep spectral segments' is on par with SLIC and Felzenszwalb, exhibiting a lower UE of 0.0158 and 0.2125 for the Carotid and Cardiac datasets, respectively, and a higher BR of 0.677 for the Thyroid dataset. SLIC has a better BR for the other two datasets, which can be explained by the fact that it is not possible to enforce a specific number of segments for fair comparisons, making the SLIC images being even more oversegmented, leading to higher BR. In Table 4, we compare our eigensegments with SLIC and Felzenszwalb for a downstream task of semantic segmentation and observe that our method has both lower UE and higher BR for two out of three of our datasets.

Table 3. Comparison with other methods - UE and BR of Step I masks

Method	Carotid		Thyroid		Cardiac	
	UE	BR	UE	BR	UE	BR
SLIC	0.018	**0.907**	**0.035**	0.589	0.224	**0.492**
Fz	0.026	0.578	0.035	0.475	0.302	0.434
Ours best	**0.016**	0.679	0.051	**0.677**	**0.213**	0.479

Table 4. Comparison with other methods - UE and BR of Step II masks

Method	Carotid		Thyroid		Cardiac	
	UE	BR	UE	BR	UE	BR
SLIC + DSS$_{step2}$	0.046	0.352	0.126	0.287	**0.139**	**0.649**
Fz + DSS$_{step2}$	0.046	0.335	0.111	0.314	0.238	0.339
Ours best	**0.030**	**0.433**	**0.042**	**0.589**	0.275	0.379

Our analysis reveals that masks derived from spectral decomposition (step I) outperform the baseline by a large margin, showing the benefits of ultrasound-

based affinities. At the same time, final segmentation masks (step II) fall behind in DICE scores, highlighting a quality gap and the need for ensuring semantic consistency. Although mask and position embeddings have marginally improved segmentation performance, challenges such as segment merging persist, indicating the need for further exploration into feature space enhancement and self-training techniques.

5 Conclusions

We present an adapted deep spectral segmentation method tailored for B-mode ultrasound data, utilizing self-supervised transformers to create affinity graphs for segment extraction. We integrate image preprocessing and leverage ultrasound-specific patchwise affinities in spectral clustering to mitigate semantic inconsistencies through mask and positional embeddings. Through extensive ablation studies, we underscore the efficacy of our approach. Our results highlight the significant potential of deep spectral methods for unsupervised ultrasound segmentation and suggest a promising direction for future investigations.

Acknowledgements

We would like to thank ImFusion for support and collaboration within the ForNero Project funded by BFS, AZ-1592-23.

Disclosure of Interests. The authors have no competing interests.

References

1. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: Slic superpixels compared to state-of-the-art superpixel methods. IEEE transactions on pattern analysis and machine intelligence **34**(11), 2274–2282 (2012)
2. Balestriero, R., Ibrahim, M., Sobal, V., Morcos, A., Shekhar, S., Goldstein, T., Bordes, F., Bardes, A., Mialon, G., Tian, Y., et al.: A cookbook of self-supervised learning. arXiv preprint arXiv:2304.12210 (2023)
3. zu Berge, C.S., Baust, M., Kapoor, A., Navab, N.: Predicate-based focus-and-context visualization for 3d ultrasound. IEEE Transactions on Visualization and Computer Graphics **20**(12), 2379–2387 (2014)
4. Bi, Y., Jiang, Z., Clarenbach, R., Ghotbi, R., Karlas, A., Navab, N.: Mi-segnet: Mutual information-based us segmentation for unseen domain generalization (03 2023)
5. Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., Joulin, A.: Emerging properties in self-supervised vision transformers. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 9650–9660 (2021)
6. Che, C., Mathai, T.S., Galeotti, J.: Ultrasound registration: A review. Methods **115**, 128–143 (2017)

7. Daoud, M.I., Atallah, A.A., Awwad, F., Al-Najjar, M., Alazrai, R.: Automatic superpixel-based segmentation method for breast ultrasound images. Expert Systems with Applications **121**, 78–96 (2019)
8. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient graph-based image segmentation. International journal of computer vision **59**, 167–181 (2004)
9. Hamilton, M., Zhang, Z., Hariharan, B., Snavely, N., Freeman, W.T.: Unsupervised semantic segmentation by distilling feature correspondences. arXiv preprint arXiv:2203.08414 (2022)
10. Huang, Q., Huang, Y., Luo, Y., Yuan, F., Li, X.: Segmentation of breast ultrasound image with semantic classification of superpixels. Medical Image Analysis **61**, 101657 (2020)
11. Ilesanmi, A.E., Idowu, O.P., Makhanov, S.S.: Multiscale superpixel method for segmentation of breast ultrasound. Computers in Biology and Medicine **125**, 103879 (2020)
12. Kirillov, A., He, K., Girshick, R., Rother, C., Dollár, P.: Panoptic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9404–9413 (2019)
13. Krähenbühl, P., Koltun, V.: Efficient inference in fully connected crfs with gaussian edge potentials. Advances in neural information processing systems **24** (2011)
14. Kroenke, M., Eilers, C., Dimova, D., Köhler, M., Buschner, G., Schweiger, L., Konstantinidou, L., Makowski, M., Nagarajah, J., Navab, N., Weber, W., Wendler, T.: Tracked 3d ultrasound and deep neural network-based thyroid segmentation reduce interobserver variability in thyroid volumetry. PLOS ONE **17**, e0268550 (07 2022). https://doi.org/10.1371/journal.pone.0268550
15. Leclerc, S., Smistad, E., Pedrosa, J., Østvik, A., Cervenansky, F., Espinosa, F., Espeland, T., Berg, E.A.R., Jodoin, P.M., Grenier, T., Lartizien, C., D'hooge, J., Løvstakken, L., Bernard, O.: Deep learning for segmentation using an open large-scale dataset in 2d echocardiography. IEEE Transactions on Medical Imaging **38**, 2198–2210 (2019), https://api.semanticscholar.org/CorpusID:73510235
16. Liu, S., Wang, Y., Yang, X., Lei, B., Liu, L., Li, S.X., Ni, D., Wang, T.: Deep learning in medical ultrasound analysis: a review. Engineering **5**(2), 261–275 (2019)
17. Melas-Kyriazi, L., Rupprecht, C., Laina, I., Vedaldi, A.: Deep spectral methods: A surprisingly strong baseline for unsupervised semantic segmentation and localization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8364–8375 (2022)
18. Mounica, S., Ramakrishnan, S., Thamotharan, B.: A study on preprocessing techniques for ultrasound images of carotid artery. In: Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB). pp. 1725–1738. Springer (2019)
19. Neubert, P., Protzel, P.: Superpixel benchmark and comparison. In: Proc. Forum Bildverarbeitung. vol. 6, pp. 1–12 (2012)
20. Noble, J.A., Navab, N., Becher, H.: Ultrasonic image analysis and image-guided interventions. Interface focus **1**(4), 673–685 (2011)
21. Riha, K., Mašek, J., Burget, R., Beneš, R., Zavodna, E.: Novel method for localization of common carotid artery transverse section in ultrasound images using modified viola-jones detector. Ultrasound in medicine biology **39** (07 2013). https://doi.org/10.1016/j.ultrasmedbio.2013.04.013
22. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. pp. 234–241. Springer (2015)

23. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Transactions on pattern analysis and machine intelligence **22**(8), 888–905 (2000)
24. Siddique, N., Paheding, S., Elkin, C.P., Devabhaktuni, V.: U-net and its variants for medical image segmentation: A review of theory and applications. Ieee Access **9**, 82031–82057 (2021)
25. Van Sloun, R.J., Cohen, R., Eldar, Y.C.: Deep learning in ultrasound imaging. Proceedings of the IEEE **108**(1), 11–29 (2019)
26. Vedaldi, A., Soatto, S.: Quick shift and kernel methods for mode seeking. In: Computer Vision–ECCV 2008: 10th European Conference on Computer Vision. pp. 705–718. Springer (2008)
27. Velikova, Y., Azampour, M.F., Simson, W., Gonzalez Duque, V., Navab, N.: Lotus: learning to optimize task-based us representations. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 435–445. Springer Nature Switzerland Cham (2023)
28. Velikova, Y., Simson, W., Azampour, M.F., Paprottka, P., Navab, N.: Cactuss: Common anatomical ct-us space for us examinations. International Journal of Computer Assisted Radiology and Surgery pp. 1–9 (2024)
29. Von Luxburg, U.: A tutorial on spectral clustering. Statistics and computing **17**, 395–416 (2007)
30. Wang, X., Girdhar, R., Yu, S.X., Misra, I.: Cut and learn for unsupervised object detection and instance segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3124–3134 (2023)
31. Wang, Y., Shen, X., Hu, S.X., Yuan, Y., Crowley, J.L., Vaufreydaz, D.: Self-supervised transformers for unsupervised object discovery using normalized cut. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14543–14553 (2022)
32. Wang, Y., Shen, X., Yuan, Y., Du, Y., Li, M., Hu, S.X., Crowley, J.L., Vaufreydaz, D.: Tokencut: Segmenting objects in images and videos with self-supervised transformer and normalized cut. IEEE Transactions on Pattern Analysis and Machine Intelligence (2023)
33. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Multi-stage progressive image restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 14821–14831 (2021)