

Self-supervised Semantic Texture Decomposition for Ultrasound Segmentation and Analysis

Tal Grutman¹ 

TALGRUTMAN@MAIL.TAU.AC.IL

Tali Ilovitsh^{1,2} 

ILOVITSH@TAUEX.TA.AC.IL

¹ School of Biomedical Engineering, Tel-Aviv University, Israel

² The Sagol School of Neuroscience, Tel-Aviv University, Israel

Editors: Accepted for publication at MIDL 2025

Abstract

Ultrasound is an especially challenging modality to interpret because it requires unique domain knowledge to draw conclusions on the anatomy based on analysis of B-mode intensity. For this reason, there is great value in transforming B-mode images to a color scheme that is more closely aligned with the anatomy and its unique tissue properties like speed-of-sound and scattering coefficient to simplify ultrasound analysis. In this work, we introduce texture ultrasound semantic analysis (TUSA), a self-supervised transformer model trained to decompose B-mode ultrasound into distinct channels that are defined by the texture they represent. We train our model on 10 freely available ultrasound datasets and demonstrate superior segmentation performance and consistency compared to training on B-mode intensity on an additional 11th dataset. We conclude that by incorporating TUSA into the training pipeline, downstream models can focus on recognizing the anatomy instead of extracting features from intensity.

Keywords: Ultrasound, self-supervised learning, segmentation

1. Introduction

Ultrasound imaging relies on B-mode intensity, which represents tissue properties but lacks direct anatomical correlation, making interpretation dependent on expert knowledge. In this work, we focus on the texture patterns in the image, which are closely intertwined with the anatomical structures. Previous efforts have explored classical texture analysis techniques to enhance ultrasound segmentation (Milko et al., 2008; Muzzolini et al., 1993; Id et al., 2019; Zhan and Shen, 2006), but these approaches often fail to generalize due to the high variability in B-mode images caused by low contrast and noise (Abhisheka et al., 2023), which has prompted great interest in deep learning applications for ultrasound analysis, as deep learning algorithms have been able to generalize well in a wide variety of ultrasound applications (Prevost et al., 2018; Park and Lee, 2020; Byra et al., 2018; Schein et al., 2023; Shin et al., 2024; Burgos-Artizzu et al., 2020; Simson et al., 2024). However, deep learning algorithms are bounded by the quantity and quality of their datasets, and although many ultrasound datasets are freely available online, semantic pixel-wise labels only exist for targeted applications rather than for general-purpose comprehension of images. This challenge can be alleviated by using contrastive learning methods based on SimCLR (Chen et al., 2020) to build powerful representations of data so that downstream models can focus on learning the task at hand rather than how to analyze the images. Inspired by the ability of self-supervised networks to learn deep intuition, our goal is to build a general-purpose ultrasound model that can pre-analyze B-mode images and semantically label the macro-scale texture groups to improve the accessibility of ultrasound.

2. Methods

2.1. General Framework

We propose the following texture ultrasound semantic analysis (TUSA) pipeline. First, a sliding window U-Net transformer (SwinUNETR) (Hatamizadeh et al., 2022) processes an input B-mode ultrasound image with an output dimension corresponding to the expected number of unique textures visible in the dataset and is numerically encouraged to select a single texture channel using the differentiable Sparsemax activation (Martins and Astudillo, 2016). Finally, a depth-wise separable convolution reproduces the B-mode image from its texture components. After training, the segmentation model can be used as a standalone pre-processing mechanism for downstream tasks.

2.2. Training Data

Our TUSA model is trained on freely available ultrasound data from various sources, countries, hardware and a uniform sampling of the following organs: knee (Yerich et al., 2020), fetal (Lu et al., 2022; Burgos-Artizzu et al., 2020), cardiac (Leclerc et al., 2019), carotid (Momot, 2022), breast (Al-Dhabyani et al., 2020; A., 2024), thyroid (Pedraza et al., 2015), liver (Byra et al., 2018), and uterine fibroid (Yang, 2023). An additional dataset featuring kidney ultrasound (Singla et al., 2023) was used for downstream validation experiments in addition to the JNU-IFM (Lu et al., 2022) dataset that was a part of the training set.

2.3. Downstream Segmentation

Downstream Segmentation. After selecting a model candidate, we trained segmentation models on the JNU-IFM and kidney datasets. The JNU-IFM dataset contains semantic labels of the fetal head and symphysis pubis. The kidney dataset contains two types of semantic labels: a binary mask encapsulating the entire kidney and a multi-class mask of the cortex, medulla, and central echogenic complex. Each of the downstream datasets was split into training and test data, each comprising 80% and 20% of the data accordingly. We trained Attention U-Net (Oktay et al.) 100 times, half of the train runs operated on the B-mode image and half of them learned to segment the kidney from TUSA’s texture logits.

3. Results and Discussion

Table 1: Dice scores for segmentation models trained on the kidney and JNU-IFM datasets

Dataset	Standard B-mode	TUSA
JNU-IFM (multi-class)	0.87 ± 0.02	0.88 ± 0.00
Kidney (multi-class)	0.54 ± 0.03	0.55 ± 0.01
Kidney (capsule)	0.8 ± 0.04	0.83 ± 0.01

The improvement in models trained on texture channels (Table 1) is in line with our expectation that simplifying B-mode intensity makes analysis of the image easier, as down-

stream models can skip straight to identifying target structures rather than extracting features from intensity patterns. This is reinforced by looking at individual channels of the validation data Figure 1, where certain features can be found just by isolating the correct channel and analyzing the shapes, as in the case of the carotid artery (1c), fetal brain ventricle (1c), and breast lesions (1g).

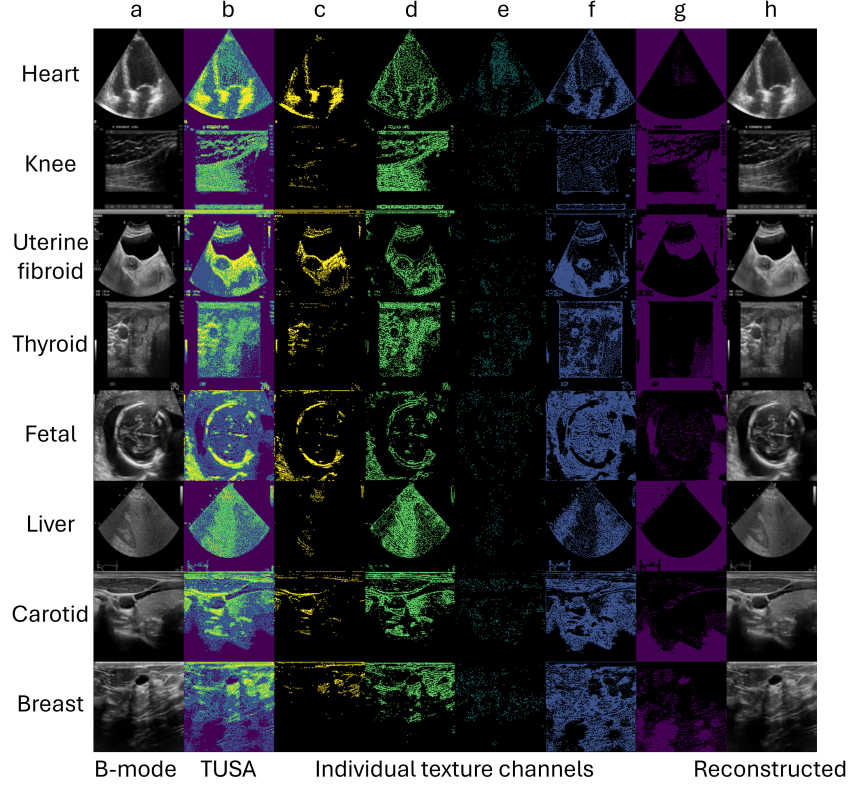


Figure 1: TUSA Validation Images– A sample image is shown for each organ in the validation set (a) alongside the corresponding TUSA decomposition (b). The individual texture channels (c-g) are presented in order of echo-genicity from left (most echogenic) to right (most anechoic). Output b-mode reconstruction (h) from TUSA feature channels.

4. Conclusion

We introduce TUSA, a self-supervised texture decomposition model for ultrasound segmentation and analysis. By learning a texture-based representation of ultrasound images, TUSA improves segmentation accuracy and consistency across datasets. Our approach provides a general-purpose tool for ultrasound analysis, reducing the reliance on expert annotations and enhancing model interpretability.

References

- A. Domalik A. Jaguś D. Kasprzak P. Matkowski R. Fura Ł. Nowicki A. Zolek N Pawłowska A., Ćwierz-Pieńkowska. A curated benchmark dataset for ultrasound based breast lesion analysis (breast-lesions-usg, 2024).
- Barsha Abhisheka, Saroj Kumar Biswas, Biswajit Purkayastha, Dolly Das, and Alexandre Escargueil. Recent trend in medical imaging modalities and their applications in disease diagnosis: a review. *Multimedia Tools and Applications*, 83:43035–43070, 10 2023. ISSN 1573-7721. doi: 10.1007/s11042-023-17326-1.
- Walid Al-Dhabyani, Mohammed Gomaa, Hussien Khaled, and Aly Fahmy. Dataset of breast ultrasound images. *Data in brief*, 28:104863, 2 2020. ISSN 2352-3409. doi: 10.1016/j.dib.2019.104863.
- Xavier P. Burgos-Artizzu, David Coronado-Gutiérrez, Brenda Valenzuela-Alcaraz, Elisenda Bonet-Carne, Elisenda Eixarch, Fatima Crispi, and Eduard Gratacós. Evaluation of deep convolutional neural networks for automatic classification of common maternal fetal ultrasound planes. *Scientific Reports*, 10:10200, 6 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-67076-5.
- Michał Byra, Grzegorz Styczynski, Cezary Szmigielski, Piotr Kalinowski, Łukasz Michałowski, Rafał Paluszkiwicz, Bogna Ziarkiewicz-Wróblewska, Krzysztof Zieniewicz, Piotr Sobieraj, and Andrzej Nowicki. Transfer learning with deep convolutional neural network for liver steatosis assessment in ultrasound images. *International Journal of Computer Assisted Radiology and Surgery*, 13:1895–1903, 12 2018. ISSN 1861-6410. doi: 10.1007/s11548-018-1843-2.
- Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. 2020. URL <https://github.com/google-research/simclr>.
- Ali Hatamizadeh, Vishwesh Nath, Yucheng Tang, Dong Yang, Holger R. Roth, and Daguang Xu. Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12962 LNCS:272–284, 2022. ISSN 1611-3349. doi: 10.1007/978-3-031-08999-2_22. URL https://link.springer.com/chapter/10.1007/978-3-031-08999-2_22.
- Alfredo Illanes Id, Nazila Esmaeili Id, Prabal Poudel, Sathish Balakrishnan, and Michael Friebeid. Parametrical modelling for texture characterization-a novel approach applied to ultrasound thyroid segmentation. 2019. doi: 10.1371/journal.pone.0211215. URL <https://doi.org/10.1371/journal.pone.0211215>.
- Sarah Leclerc, Erik Smistad, Joao Pedrosa, Andreas Ostvik, Frederic Cervenansky, Florian Espinosa, Torvald Espeland, Erik Andreas Rye Berg, Pierre-Marc Jodoin, Thomas Grenier, Carole Lartizien, Jan Dhooze, Lasse Lovstakken, and Olivier Bernard. Deep learning for segmentation using an open large-scale dataset in 2d echocardiography.

- IEEE Transactions on Medical Imaging*, 38:2198–2210, 9 2019. ISSN 0278-0062. doi: 10.1109/TMI.2019.2900516.
- Yaosheng Lu, Mengqiang Zhou, Dengjiang Zhi, Minghong Zhou, Xiaosong Jiang, Ruiyu Qiu, Zhanhong Ou, Huijin Wang, Di Qiu, Mei Zhong, Xiaoxing Lu, Gaowen Chen, and Jieyun Bai. The jnu-ifm dataset for segmenting pubic symphysis-fetal head. *Data in Brief*, 41:107904, 4 2022. ISSN 23523409. doi: 10.1016/j.dib.2022.107904.
- André F. T. Martins and Ramón Fernandez Astudillo. From softmax to sparsemax: A sparse model of attention and multi-label classification. 2 2016.
- Sergiy Milko, Eigil Samset, and Timor Kadir. Segmentation of the liver in ultrasound: A dynamic texture approach. *International Journal of Computer Assisted Radiology and Surgery*, 3:143–150, 5 2008. ISSN 18616429. doi: 10.1007/S11548-008-0217-6/METRICS. URL <https://link.springer.com/article/10.1007/s11548-008-0217-6>.
- Agata Momot Momot. Common carotid artery ultrasound images, 2022.
- Russell Muzzolini, Yee Hong Yang, and Roger Pierson. Multiresolution texture segmentation with application to diagnostic ultrasound images. *IEEE Transactions on Medical Imaging*, 12:108–123, 1993. ISSN 1558254X. doi: 10.1109/42.222674.
- Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven Mcdonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention u-net: Learning where to look for the pancreas.
- Jun Hong Park and Sang Joon Lee. Ultrasound deep learning for wall segmentation and near-wall blood flow measurement. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 67:2022–2032, 10 2020. ISSN 15258955. doi: 10.1109/TUFFC.2020.2995467.
- Lina Pedraza, Carlos Vargas, Fabián Narváez, Oscar Durán, Emma Muñoz, and Eduardo Romero. An open access thyroid ultrasound image database. page 92870W, 1 2015. doi: 10.1117/12.2073532.
- Raphael Prevost, Mehrdad Salehi, Simon Jagoda, Navneet Kumar, Julian Sprung, Alexander Ladikos, Robert Bauer, Oliver Zettinig, and Wolfgang Wein. 3d freehand ultrasound without external tracking using deep learning. *Medical Image Analysis*, 48:187–202, 8 2018. ISSN 1361-8415. doi: 10.1016/J.MEDIA.2018.06.003.
- Dror Schein, Tal Grutman, and Tali Ilovitsh. Deep learning-based ultrasound beam shaping for spatiotemporal acoustic holograms generation. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 70:551–561, 6 2023. ISSN 15258955. doi: 10.1109/TUFFC.2023.3264580.
- Yi Rang Shin, Matthew R. Lowerison, Yike Wang, Xi Chen, Qi You, Zhijie Dong, Mark A. Anastasio, and Pengfei Song. Context-aware deep learning enables high-efficacy localization of high concentration microbubbles for super-resolution ultrasound localization microscopy. *Nature Communications* 2024 15:1, 15:1–16, 4 2024. ISSN 2041-

1723. doi: 10.1038/s41467-024-47154-2. URL <https://www.nature.com/articles/s41467-024-47154-2>.
- Walter A. Simson, Magdalini Paschali, Vasiliki Sideri-Lampretsa, Nassir Navab, and Jeremy J. Dahl. Investigating pulse-echo sound speed estimation in breast ultrasound with deep learning. *Ultrasonics*, 137:107179, 2 2024. ISSN 0041-624X. doi: 10.1016/J.ULTRAS.2023.107179.
- Rohit Singla, Cailin Ringstrom, Grace Hu, Victoria Lessoway, Janice Reid, Christopher Ngan, and Robert Rohling. *The Open Kidney Ultrasound Data Set*, pages 155–164. 2023. doi: 10.1007/978-3-031-44521-7_15.
- Tiantian Yang. uterine fibroid ultrasound images. *Mendeley Data*, 2023.
- Nadiya V. Yerich, Carolina Alvarez, Todd A. Schwartz, Serena Savage-Guin, Jordan B. Renner, Catherine J. Bakewell, Minna J. Kohler, Janice Lin, Jonathan Samuels, and Amanda E. Nelson. A standardized, pragmatic approach to knee ultrasound for clinical research in osteoarthritis: The johnston county osteoarthritis project. *ACR Open Rheumatology*, 2:438–448, 7 2020. ISSN 2578-5745. doi: 10.1002/acr2.11159.
- Yiqiang Zhan and Dinggang Shen. Deformable segmentation of 3-d ultrasound prostate images using statistical texture matching method. *IEEE Transactions on Medical Imaging*, 25:256–272, 3 2006. ISSN 02780062. doi: 10.1109/TMI.2005.862744.