Self-Supervised Learning Meets Liver Ultrasound Imaging

Abder-Rahman Ali
Harvard Medical School
Massachusetts General Hospital
Boston, MA 02114 USA
aali25@mgh.harvard.edu

Anthony E. Samir
Harvard Medical School
Massachusetts General Hospital
Boston, MA 02114 USA
asamir@mgh.harvard.edu

Abstract

In the field of medical ultrasound imaging, conventional B-mode "grey scale" ultrasound and shear wave elastography (SWE) are widely used for chronic liver disease diagnosis and risk stratification. However, many abdominal ultrasound images do not include views of the liver, necessitating a pre-processing liver view detection step before feeding the image to the AI system. To address this, we propose a self-supervised learning method, SimCLR+LR, for image classification that utilizes a large set of unlabeled abdominal ultrasound images to learn image representations. These representations are then fine-tuned to the downstream task of liver view classification. This approach outperforms traditional supervised learning methods and achieves superior performance when compared to state-of-the-art (SOTA) models, ResNet-18 and MLP-Mixer. Once the liver view is detected, the next crucial phase involves the segmentation of the liver region, imperative for obtaining accurate and dependable results in SWE. For this, we present another self-supervised learning approach, SimCLR+ENet, which leverages the learned feature representations and fine-tunes them on the task of liver segmentation, followed by a refinement step using CascadePSP. The proposed approach outperforms the SOTA method U-Net. SimCLR+ENet was also used to detect poor probe contact (i.e., areas where the ultrasound probe/transducer does not have adequate contact with the patient's skin) in liver ultrasound images, an artifact that affects the reliability of SWE. The combination of the proposed self-supervised learning methods for liver view classification, liver segmentation, and poor probe contact detection not only reduces the time and cost associated with data labeling, but also optimizes the liver segmentation workflow and SWE reliability in a real-time setting.

1 Introduction

Non-alcoholic fatty liver disease (NAFLD) is a common chronic liver condition characterized by excess liver fat, affecting approximately 24% of the U.S. population and costing an estimated $103 billion annually (1; 2). Liver fibrosis stage at diagnosis is a key factor in patient outcomes (3). Current diagnostic methods, such as liver biopsy, are invasive and costly, with potential for errors. Non-invasive alternatives like shear wave elastography (SWE) have been developed to measure liver stiffness, a valuable biomarker for NAFLD (4). However, SWE's effectiveness depends on accurate liver segmentation and high-quality image acquisition. Not all abdominal ultrasound images contain liver views, necessitating a pre-processing step to confirm the liver presence before AI processing. The cost and time for image labeling and annotation can be high, especially for supervised deep learning algorithms. We propose two contrastive self-supervised learning methods, SimCLR+LR for liver view classification and SimCLR+ENet for liver segmentation and poor probe contact detection, which utilize a large dataset of unlabeled abdominal ultrasound images, achieving top performance

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

with minimal labeled/annotated data. The remainder of the paper is structured as follows: Section 2 provides a survey of related work. The proposed methods SimCLR+LR and SimCLR+ENet are described in Sections 3 and 4, respectively. The results of the proposed approaches are analyzed and compared with SOTA methods in Section 5. Finally, the paper is concluded in Section 6.

2 Related work

The literature on abdominal organ classification via ultrasound images is limited. Li et. al. (5) proposed an automatic abdominal organ recognition method using DNNs and k-NN, achieving 96.67% accuracy. Reddy et. al. (6) used a transfer learning framework with pre-trained models for the same task, achieving high performance. Dadoun et al. (7) developed a multi-label classification framework using deep clustering with PICA (8) and semi-supervised learning with FixMatch (9). The approach achieved an F-1 score of 0.89 in liver classification. The literature also features a limited number of studies focusing on the segmentation of the liver in ultrasound images using deep learning approaches, with most studies utilizing U-Net (10). Garcia et al. (11) trained U-Net on 500 B-mode images and tested on 125, with test accuracy surpassing 86%. Wu et al. (12) used U-Net for texture analysis of renal cortex and liver portions; liver intersection over union (IoU) evaluated to 0.643 ± 0.186. Ibrahim (13) used U-Net for liver tissue segmentation with Dice scores ranging between 78% and 89%. Shuhang et al. (14) used SDU-Net (15), a modification of U-Net, to segment the liver region, achieving a Dice score of 91.5%. Rhyou and Yoo (16) used the DeepLabv3+ architecture (17) pretrained through transfer learning to segment the liver region, resulting in an accuracy of 94.9%. It is worth mentioning that previous research has predominantly concentrated on liver segmentation, with limited exploration of poor probe contact detection.

3 SimCLR+LR

SimCLR+LR (18) (Fig.1) involves a two-stage training method. In the first stage, SimCLR (19) is used to learn effective feature representations from a large dataset of unlabeled abdominal ultrasound images. It maximizes agreement between differently augmented views of the same data example through a contrastive loss in the latent space. Using ResNet-18 as the base encoder network for SimCLR offers numerous advantages for self-supervised learning. Its computational efficiency, faster training, and reduced risk of overfitting make it suitable for real-world clinical applications. In the second stage, we use logistic regression (LR) as a supervised learning approach for the downstream task of liver view classification. Logistic regression fine-tunes the learned feature representations, enabling the model to predict the presence or absence of the liver in abdominal ultrasound images. The combination of SimCLR for unsupervised feature learning and LR for supervised fine-tuning allows the method to benefit from the large-scale unlabeled dataset in the first stage and achieve accurate classification results with a smaller labeled dataset in the second stage.

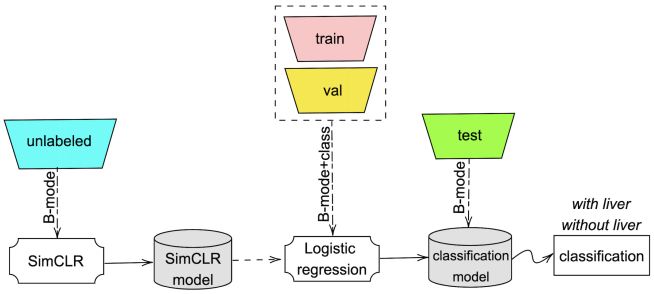

Figure 1: Liver view classification using SimCLR+LR

4 SimCLR+ENet

Similar to SimCLR+LR, the proposed liver segmentation approach, SimCLR+ENet (20), involves two stages: contrastive self-supervised learning using SimCLR, and supervised learning with ENet (21) for the downstream task of liver segmentation. To improve the accuracy of segmentation results,

we use CascadePSP (22) for refinement, which takes multiple imperfect segmentation masks at different scales and produces a refined segmentation. To further enhance the performance of our approach, we use physics-inspired augmentations for ultrasound imaging during training ENet. These augmentations include sector angle, penetration, gaussian blur, force control, and zooming. For sector angle, we generate B-mode images with a narrow sector angle, mirroring real-world line-per-frame scenarios. Penetration augmentation emulates the darkening of deeper areas, especially with high-frequency probes. Gaussian blur, acting as a low-pass filter, reduces higher spatial frequencies, introducing distortion. Conversely, forced control for focused ultrasound energy results in sharper images. Zooming manipulations simulate the changes in textural features during ultrasound scanning. SimCLR+ENet, depicted in Fig.2, was also used to detect poor probe contact as will be demonstrated in the results section; physics-inspired augmentations and CascadePSP refinement were not used in poor probe contact segmentation/detection.

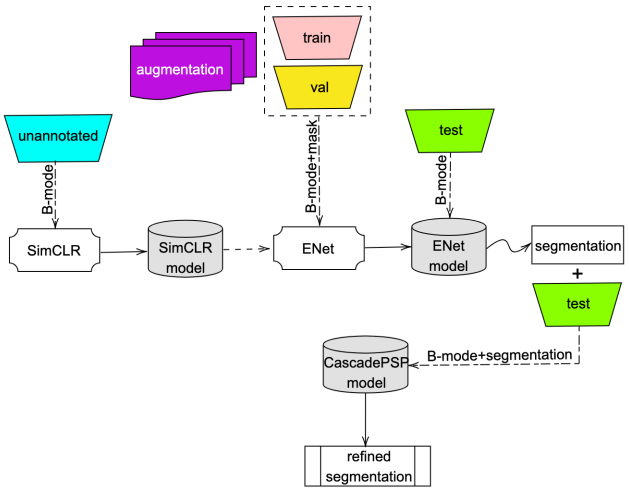

Figure 2: Liver segmentation using SimCLR+ENet

5 Results and discussion

Liver ultrasound DICOM exams were collected from the Massachusetts General Hospital's (MGH) electronic medical record (EMR). Specifically, the data for training, validation, and testing were drawn from 903 abdominal ultrasound examinations, each containing a sequence of DICOM files. The selected images were labeled/annotated by MGH radiologists for the supervised part of the proposed methods (i.e., LR and ENet). The self-supervised learning part, SimCLR, was trained on 17,441 unlabeled abdominal ultrasound images for 100 epochs, and the trained model was then fine-tuned for the relevant downstream task. The experiments were performed on a NVIDIA GeForce RTX 2080-Ti GPU.

5.1 SimCLR+LR

SimCLR+LR focuses on classifying liver ultrasound images into "with liver" and "without liver" classes. The LR (supervised) part was trained on a dataset of 854 labeled B-mode images (with liver: 495, without liver: 359), with the validation and testing datasets composed of 90 (with liver: 51, without liver: 39) and 150 (with liver: 101, without liver: 49) images, respectively. Fig.3 shows examples of the training data and corresponding classes. SimCLR+LR, ResNet-18, and MLP-Mixer (23) were trained for 100 epochs (i.e., in SimCLR+LR, each part was trained for 100 epochs). The SimCLR+LR model outperformed the other two models in both classifying "with liver" and "without liver" classes. For the "with liver" class, SimCLR+LR achieved an accuracy of 95.1%, while ResNet-18 and MLP-Mixer achieved an accuracy of 79.2% and 82.2%, respectively. For the "without liver" class, SimCLR+LR demonstrated an accuracy of 93.9%, higher than ResNet-18 (61.2%) and MLP-Mixer (89.8%). Overall, the accuracy for SimCLR+LR, ResNet-18, and MLP-Mixer was 94.5%, 70.2%, and 86% respectively. Notably, SimCLR+LR, trained with only "one" labeled image

for each class, achieved a performance similar to MLP-Mixer, which was trained on the entire dataset of 854 images.

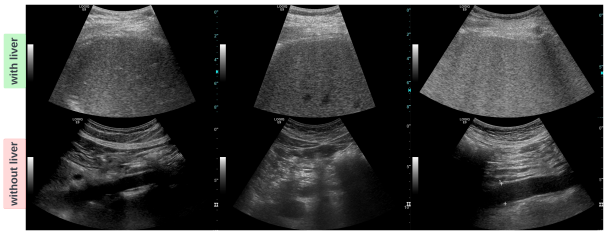

Figure 3: Samples of SimCLR+LR training images and their associated label/class

5.2 SimCLR+ENet

For liver segmentation, ENet was trained and validated on 495 and 51 images, respectively; the training process was carried out over 1050 epochs. SimCLR+ENet was tested on 95 images. The proposed method was compared with U-Net (10) for liver segmentation. SimCLR+ENet (with physics-inspired augmentations and refinement) and U-Net had an average Dice coefficient of 90.58% and 89.77%, and an average Hausdorff distance (i.e., takes into account the shape of the segmented regions, with lower values indicating better performance) of 21.71 and 29.53, respectively. Fig.4 provides a comparative illustration of liver segmentation using SimCLR+ENet and U-Net.

Figure 4: Liver ultrasound segmentation results: (a) B-mode image (b) ground truth (c) U-Net (d) SimCLR+ENet (e) liver segmentation contours (Ground truth, SimCLR+ENet, U-Net)

SimCLR+ENet was also used for poor probe contact detection. Training data was composed of 1564 images, with the validation and testing datasets the same as those used for liver segmentation. The performance was evaluated across four classes: none (no poor probe contact present), left (poor contact on the left), right (poor contact on the right), and both (poor contact on both sides). The model achieved an overall accuracy, sensitivity, and specificity of 65.4%, 49.1%, and 82.1%, respectively, indicating moderate detection of poor probe contact and better identification of its absence. Fig.5 shows a sample of the model's output. The model's limitation was not considering the zooming factor during training, which could be addressed in future work.

Figure 5: Poor probe contact segmentation/detection using SimCLR+ENet: (a) B-mode image (b) poor probe contact (c) bounding boxes indicating poor probe contact locations on the B-mode image.

6 Conclusion

This paper introduced two self-supervised learning based methods, SimCLR+LR and SimCLR+ENet, which were used for liver view classification and liver segmentation/poor probe contact detection in abdominal ultrasound images, respectively. The methods leverage a large dataset of unlabeled images, reducing the time and cost associated with data labeling. The proposed methods achieved superior performance when compared to SOTA methods. For future work, we aim to further extend our work to multi-class organ classification and improving training data quality with data-centric AI.

References

[1] T. Arshad, P. Golabi, L. Henry, and Z. M. Younossi, "Epidemiology of non-alcoholic fatty liver disease in north america," *Current Pharmaceutical Design*, vol. 26, no. 10, pp. 993–997, 2020.

[2] M. Witkowski, S. I. Moreno, J. Fernandes, P. Johansen, M. Augusto, and S. Nair, "The economic burden of non-alcoholic steatohepatitis: a systematic review," *PharmacoEconomics*, pp. 1–26, 2022.

[3] A. J. Sanyal, M. L. Van Natta, J. Clark, B. A. Neuschwander-Tetri, A. Diehl, S. Dasarathy, R. Loomba, N. Chalasani, K. Kowdley, B. Hameed, *et al.*, "Prospective study of outcomes in adults with nonalcoholic fatty liver disease," *New England Journal of Medicine*, vol. 385, no. 17, pp. 1559–1569, 2021.

[4] G. Xiao, S. Zhu, X. Xiao, L. Yan, J. Yang, and G. Wu, "Comparison of laboratory tests, ultrasound, or magnetic resonance elastography to detect fibrosis in patients with nonalcoholic fatty liver disease: a meta-analysis," *Hepatology*, vol. 66, no. 5, pp. 1486–1501, 2017.

[5] K. Li, Y. Xu, Z. Zhao, and M. Q.-H. Meng, "Automatic recognition of abdominal organs in ultrasound images based on deep neural networks and k-nearest-neighbor classification," in *2021 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pp. 1980–1985, IEEE, 2021.

[6] D. S. Reddy, P. Rajalakshmi, and M. Mateen, "A deep learning based approach for classification of abdominal organs using ultrasound images," *Biocybernetics and Biomedical Engineering*, vol. 41, no. 2, pp. 779–791, 2021.

[7] H. Dadoun, H. Delingette, A.-L. Rousseau, E. de Kerviler, and N. Ayache, "Deep clustering for abdominal organ classification in us imaging," 2022.

[8] J. Huang, S. Gong, and X. Zhu, "Deep semantic clustering by partition confidence maximisation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 8849–8858, 2020.

[9] K. Sohn, D. Berthelot, N. Carlini, Z. Zhang, H. Zhang, C. A. Raffel, E. D. Cubuk, A. Kurakin, and C.-L. Li, "Fixmatch: Simplifying semi-supervised learning with consistency and confidence," *Advances in neural information processing systems*, vol. 33, pp. 596–608, 2020.

[10] O. Ronneberger, P. Fischer, and T. Brox, "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.

[11] R. B. Garcia, A. Lightstone, M. N. Ibrahim, and A. El Kaffas, "Abstract PO-022: Automated liver tissue segmentation in point of care ultrasound b-mode images using U-Net," *Clinical Cancer Research*, vol. 27, pp. PO–022–PO–022, 03 2021.

[12] C.-H. Wu, C.-L. Hung, T.-Y. Lee, C.-Y. Wu, and W. C.-C. Chu, "Fatty liver diagnosis using deep learning in ultrasound image," in *2022 IEEE International Conference on Digital Health (ICDH)*, pp. 185–192, IEEE, 2022.

[13] M. Naim Ibrahim, *Automated Fatty Liver Disease Detection in Point-of-Care Ultrasound B-mode Images*. PhD thesis, University of Guelph, 2022.

[14] S. Wang, V. Kumar, S.-Y. Hu, X. Wang, L. Chen, J. Wang, A. Ozturk, Q. Li, T. T. Pierce, Y. Tsymbalenko, *et al.*, "Automated estimation of hepatorenal index: a biomarker for hepatic steatosis," 2022.

[15] S. Wang, S.-Y. Hu, E. Cheah, X. Wang, J. Wang, L. Chen, M. Baikpour, A. Ozturk, Q. Li, S.-H. Chou, *et al.*, "U-net using stacked dilated convolutions for medical image segmentation," *arXiv preprint arXiv:2004.03466*, 2020.

[16] S.-Y. Rhyou and J.-C. Yoo, "Cascaded deep learning neural network for automated liver steatosis diagnosis using ultrasound images," *Sensors*, vol. 21, no. 16, p. 5304, 2021.

[17] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 801–818, 2018.

[18] A.-R. Ali, A. E. Samir, and P. Guo, "Self-supervised learning for accurate liver view classification in ultrasound images with minimal labeled data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3086–3092, 2023.

[19] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*, pp. 1597–1607, PMLR, 2020.

[20] A.-R. Ali, P. Guo, and A. E. Samir, "Liver segmentation in ultrasound images using self-supervised learning with physics-inspired augmentation and global-local refinement," in *The 36th Canadian Conference on Artificial Intelligence*, 2023.

[21] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, "Enet: A deep neural network architecture for real-time semantic segmentation," *arXiv preprint arXiv:1606.02147*, 2016.

[22] H. K. Cheng, J. Chung, Y.-W. Tai, and C.-K. Tang, "Cascadepsp: Toward class-agnostic and very high-resolution segmentation via global and local refinement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8890–8899, 2020.

[23] I. O. Tolstikhin, N. Houlsby, A. Kolesnikov, L. Beyer, X. Zhai, T. Unterthiner, J. Yung, A. Steiner, D. Keysers, J. Uszkoreit, *et al.*, "Mlp-mixer: An all-mlp architecture for vision," *Advances in neural information processing systems*, vol. 34, pp. 24261–24272, 2021.