Can Crowdsourced Annotations Improve AI-based Congestion Scoring For Bedside Lung Ultrasound?

Ameneh Asgari-Targhi[1 [0000−0002−1971−3962]], Tamas Ungi[2 [0000−0003−4743−0609]], Mike Jin[1,3 [0000−0001−7237−6697]], Nicholas Harrison[3 [0000−0002−8331−7833]], Nicole Duggan[1 [0000−0003−4829−4979]], Erik Duhaime[3 [0000−0001−8026−4206]], Andrew Goldsmith[1 [0000−0003−0979−7178]], and Tina Kapur[1 [0000−0003−3646−9508]]

[1] Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
aasgari@bwh.harvard.edu
[2] Queen's University, Kingston, ON, Canada
[3] Centaur Labs, Boston, MA
[4] University of Indiana School of Medicine, Indianapolis, IN

Abstract. Lung ultrasound (LUS) has become an indispensable tool at the bedside in emergency and acute care settings, offering a fast and non-invasive way to assess pulmonary congestion. Its portability and cost-effectiveness make it particularly valuable in resource-limited environments where quick decision-making is critical. Despite its advantages, the interpretation of B-line artifacts, which are key diagnostic indicators for conditions related to pulmonary congestion, can vary significantly among clinicians and even for the same clinician over time. This variability, coupled with the time pressure in acute settings, poses a challenge. To address this, our study introduces a new B-line segmentation method to calculate congestion scores from LUS images, aiming to standardize interpretations. We utilized a large dataset of 31,000 B-line annotations synthesized from over 550,000 crowdsourced opinions on LUS images of 299 patients to improve model training and accuracy. This approach has yielded a model with 94% accuracy in B-line counting (within a margin of 1) on a test set of 100 patients, demonstrating the potential of combining extensive data and crowdsourcing to refine lung ultrasound analysis for pulmonary congestion.

Keywords: B-line segmentation · lung ultrasound · crowdsourcing · low cost imaging

1 Introduction

Lung ultrasonography (LUS) has gained increasing importance in bedside diagnostic assessments and therapeutic management within acute care settings [18]. B-line artifacts in LUS, hyperechoic lines originating from the pleura and extending radially to the bottom of the screen, move synchronously with respiration [15]. Despite their artifactual nature, B-lines play a significant role in

detecting and evaluating the severity of various lung diseases. B-lines appear in pulmonary congestion due to decompensated heart failure, viral and bacterial pneumonia, or increased collagen deposition in the interstitial space due to autoimmune diseases or prolonged exposure to dust. The quantity of B-lines often serves as a biomarker for disease severity, influencing treatment decisions. However, studies have noted considerable variability in B-line quantification among observers, attributed to differences in expertise, operator dependence, and acquisition settings. In clinical practice, physicians commonly rely on estimating disease severity scores based on visual B-line quantification.

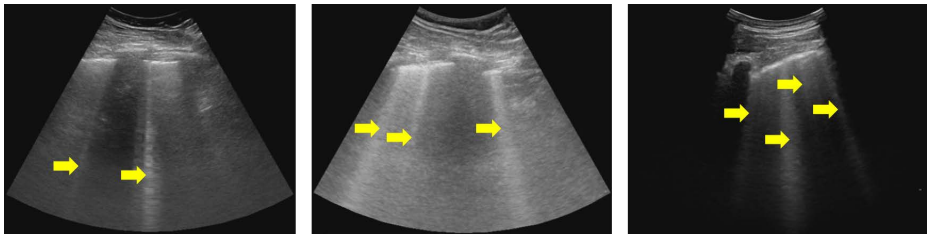


Fig. 1. Arrows pointing at B-lines in example images. Acoustic artifacts observed in lung ultrasound (LUS) imaging, characterized by hyperechoic vertical lines originating from the pleura line and extending to the bottom of the screen.

B-lines have been studied using classical image processing techniques as well as contemporary deep learning methods. One of the earliest classical image processing approaches transformed curvilinear LUS frames from polar coordinates to a Cartesian grid, where B-lines appear as vertical lines, and then performed column-wise handcrafted intensity-based features to detect B-lines [3]. B-lines detection was posed as an inverse line-detection problem and solved using the Radon transform [1], to which sparsity-enforcing and Cauchy-based penalty process was added to regularize the solution [13]. More recently, wavelets were used to denoise prior to using the Radon transforms to extract B-lines [7].

Among the earliest deep learning applications to B-line detection trained convolutional neural networks (CNNs) with gradient-based class activation mapping on both phantom and patient LUS videos [24]. Spatial transformer networks were used for COVID-19 scoring on a dataset with 277 LUS videos from 35 patients with a weighted F1 score of 65.1% [22]. CNNs were trained on 153 exams of adults containing 4651 videos and 122 exams of pediatric patients with 3022 videos to classify discrete and merged (confluent) B-lines, reporting an area under the ROC curve of 0.92 [23]. CNNs were used for disease classification on a dataset of 202 video clips and 59 images from 261 patients, achieving an F1 score of 0.92 [2]. Networks with a temporal shift module (TSM) were employed to detect A-lines and B-lines in 665 LUS videos from 172 subjects with intraclass correlation coefficients of 0.73 and 0.66, respectively [8].

U-nets with domain adaptation were applied to B-line segmentation in COVID and pneumonia on a dataset of 13 patients (1303 images) with a reported sensitivity of 0.84 and specificity of 0.95 [17]. Mask R-CNN followed by a tracking algorithm were employed to localize and count B-lines in dialysis patients. They used data from 46 patients (1,003 images) for training and 15 patients (382 images) for validation and reported intraclass correlation of 0.9 with physicians [25]. U-Nets were used for segmentation of B-lines using a training set of 450 simulated phantom images supplemented by 41 patients (57 images) for transfer learning, reporting a Dice Score of 0.7 [11].

The effectiveness of deep learning methodologies is significantly influenced by access to the right quantity and quality of training data, which can be challenging to acquire due to limited availability of expert clinicians for generating annotations. The strategies that have been explored to mitigate this challenge in LUS image analysis include the use of publicly accessible annotated datasets [2], the application of simulated LUS data [26], and the implementation of semi- and self-supervised learning techniques [20, 14, 5]. As a method for increasing access to large amounts of high quality training data, a gamified crowdsourcing approach with extensive quality control metrics was recently introduced to collect opinions from users on segmenting B-lines on frames from LUS videos. By comparing the concordance of the crowdsourced segmentations to those from clinical experts, they showed that crowdsourced consensus B-line segmentations achieved expert-level quality [12]. This paper aims to investigate the potential of utilizing crowdsourced data beyond the amount that expert clinicians can provide for training U-net based models for B-line segmentation. Our approach involves gradually increasing the crowdsourced training data and testing the effect on segmentation accuracy and automatic B-line counting.

2 Methods

2.1 Data Description

This study utilized data from an IRB-approved retrospective review of medical records for patients who presented to our emergency department with symptoms indicative of pulmonary congestion—such as shortness of breath, cough, fever, weight gain—between October 9, 2020, and March 15, 2022. A cohort of 299 patients was selected based on the presence of a bedside lung ultrasound examination in their institutional medical records, conducted in the emergency department. The cohort had a mean age of 64.4 years, with 46% female sex and 66% White race. LUS exams are acquired using the BLUE protocol [16], which recommends recordings from 12 specific zones across the patient's chest, encompassing anterior, lateral, and posterior areas of both lungs. Each video is recorded for six seconds to cover at least one respiratory cycle. The number of clips per patient varied from 5 to 20, influenced by clinical constraints such as patient comfort and stability. Low-frequency curvilinear transducers (C5 and C5s) on a Mindray M9 ultrasound machine were used, with frame rates of 15-46 Hz and variable settings for depth, focal point, and gain. No enhancement

modes were applied. All LUS exams were downloaded for analysis in the Digital Imaging and Communications in Medicine (DICOM) format.

2.2 Data De-Identification and Scan Geometry Extraction

A custom software module, AnonymizeUltrasound, was created on the 3D Slicer platform to de-identify the LUS images. In addition to clearing the known DICOM tags associated with patient information, the de-identification process included masking all pixels outside the fan-shaped ultrasound area. This removes any burnt-in text, machine settings, or potentially identifiable information that may compromise patient privacy or influence model training. Additionally, the four corners of the ultrasound fan were recorded with each LUS video to enable computation of ultrasound scanlines. Extracting scanlines for data analysis standardizes images acquired with different transducers and geometries for subsequent model development.

2.3 Collection of annotations

Overview We obtained B-line annotations for all our available data in two stages. First, we obtained classifications of all LUS clips from all patients as having B-lines anywhere in the clip or not. Then, we annotated (i.e., segmented using two points $[(x_1, y_1), (x_2, y_2)]$) B-lines on every frame of every clip classified as having B-lines. For both the classification and segmentation stages, first we collected expert annotations on a small subset of data to obtain a high-quality expert consensus, and then we used the expert consensus as training material to crowdlabel the remainder of the data.

Collection of B-line segmentations First, consensus expert classifications of B-line presence on LUS clips were obtained for 400 randomly selected clips from a 50/50 train/test split of patients, with 200 from each 50/50 group of patients. Six lung ultrasound experts classified all 400 LUS clips as having B-lines anywhere in the clip or not, and an expert consensus classification was formed by taking the majority opinion of the six experts with ties broken arbitrarily. The 200 training clips were then used to seed a crowdlabeling task to obtain high-quality crowd consensus labels on all remaining clips [6].

Consensus expert B-line segmentations were obtained for 400 randomly selected LUS frames from clips where the expert or crowd consensus found B-lines present, sampled 50/50 by the same train/test patient split as above. Five experts annotated all 400 frames by segmenting all B-lines found in the frame. To form the consensus annotation from five individual expert opinions, all B-lines (in [0-100]-scaled coordinates by image size) were first clustered by agglomerative clustering up to a Hausdorff distance of 10, and clusters containing B-lines from at least three experts had their B-lines averaged to form an expert consensus B-line. The 200 training frames were similarly used to obtain high-quality crowd consensus segmentations on all other frames, verified in terms of both B-line

count and segmentation accuracy using the 21,154 crowd opinions collected over the first two days [12]. Over 530,000 opinions were then additionally collected over the course of three weeks to obtain crowd consensus segmentations for all 31,168 frames.

Mechanics of crowdsourcing Crowd opinions were collected via DiagnosUs, an iOS application on which thousands of participants compete daily for cash prizes on a variety of medical data annotation tasks. Crowd annotators were given an unlimited stream of randomly interspersed labeled (i.e., having a correct annotation) and unlabeled cases with a relative frequency of 1:2. Initially, only the 200 cases with expert consensus from the training set were labeled. After submitting an opinion on a labeled case, annotators would be shown 1) the correct annotation on the case as feedback for learning, and 2) their score on the case, computed as the soft F1 score between their annotation and the correct annotation. On unlabeled cases, annotators received neither. Their case scores on labeled cases determined their leaderboard score and thus their eligibility for cash prizes.

The skill level of individual crowd annotators was dynamically tracked using a quality score ("Qscore"), computed as their trailing average (accuracy for classification, soft F1 score for segmentation) on their last 50 submitted opinions on labeled cases. Opinions from annotators whose Qscore at the time was less than 0.8 were ignored, and the rest were considered "qualified". For the classification crowdsourcing task, cases became labeled when they reached a minimum difference of 3 votes in the top two most common qualified opinions, and the correct answer was set to the majority qualified opinion. For the segmentation crowdsourcing task, cases became labeled once they received five qualified opinions, and the correct answer was set to the result of combining the qualified opinions in the same manner as for expert consensus segmentations.

2.4 Data Pre-processing for Deep Learning

The dataset used in this study consists of 31,168 curvilinear (fan-shaped) ultrasound image frames with crowd consensus B-line annotations from 1,109 ultrasound clips of 299 patients. This data was split into a test set of 10,896 frames from 383 clips of 100 patients, a validation set of 3,861 frames from 139 clips of 40 patients, and a training set of 16,411 frames from 587 clips of 159 patients.

The fan shape of the ultrasound was used to define a polar coordinate system for each image where the origin of the polar coordinates is the intersection point of the two sides of the fan (first and last scan line). The sound beams define the radial direction, while top and bottom circles of the ultrasound area define the angular direction of the polar coordinate system. All images were down-sampled with a 128×128 grid in polar coordinates (Fig. 2). Training and evaluation of AI models were performed on images in polar coordinates to standardize image size and shape across different types of ultrasound transducers, and to keep acoustic image features uniform across all images.

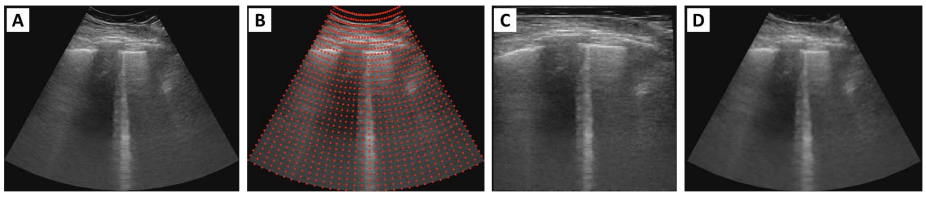


Fig. 2. Original B-mode ultrasound image of size 500×520 pixels (A). Polar coordinate grid shown as red dots (B). Same image in polar coordinate systems sampled at 128×128 pixels (C). Image C transformed back to original cartesian coordinates to verify no significant information loss (D).

2.5 Model selection and optimization

The first step in AI optimization was to select a model architecture for segmentation. We trained four previously published models on a subset of the training data, 3,392 frames from 26 patients for 100 epochs. The trained models were evaluated on the validation dataset. Implementations of the models in the MONAI framework were used [4]. All training and testing experiments were performed on the same computer with NVIDIA RTX 4080 GPU, 32 GB RAM, in a Python version 3.9.18 environment with PyTorch version 2.0.1. The four candidate model architectures were the U-Net [21], Attention U-Net [19], UNETR [10], and Swin UNETR [9]. For these initial experiments, we used the default parameters of each model in their implementations. The transformer-based models (UNETR and Swin UNETR) exhibited considerably slower inference speeds, achieving under 35 frames per second, compared to over 100 frames per second in other models. No single model demonstrated significantly better accuracy on the validation dataset. However, Attention U-Net showed slightly higher Dice scores compared to the others, leading to its selection for subsequent phases of model optimization.

The second step was hyperparameter optimization by experimenting with random parameter adjustments for training. As a result, we configured the Attention U-Net to include five contracting and five expanding stages, with 8, 16, 32, 64, 128, and 256 channels at the model depth levels, respectively. 10% drop-out rate was applied during training. The loss function for training was $L = 0.1 \times (Dice) + 0.9 \times (CrossEntropy)$. The maximum learning rate was 6.4×10^{-4} with linear warm-up and cosine annealing. The $AdamW$ optimizer was used for 100-200 epochs depending on the training data amount.

2.6 Experiments

Following hyperparameter optimization, we trained five Attention U-Net models with progressively larger labeled datasets. Dataset1: 5,233 frames (196 clips, 37 patients), Dataset2: 7,703 frames (285 clips, 53 patients), Dataset3: 10,207 frames (369 clips, 71 patients), Dataset4: 13,298 frames (483 clips, 114 patients),

and Dataset5: 16,411 frames (587 clips, 159 patients). The batch size remained constant at 64 while we adjusted the number of epochs upward for smaller datasets to balance the reduced number of optimizer steps.

Trained models were evaluated using conventional pixel-based metrics and B-line counting. Automatic B-line counting was implemented as counting connected components in the images after applying a threshold on the segmentation outputs at 1% intensity and resizing along the scanlines direction to an image height of 4 pixels to avoid counting B-lines multiple times (Figure 3). The highest B-line count from each frame within a video clip determined the B-line count for that clip. Clinically, this clip-level count, representing the peak number of B-lines observed during 1-2 breathing cycles at a lung area, holds significance for each clip.

As a reference for comparison to the models trained on Datasets 1 through 5, we also trained an Attention U-Net model using only the 400 frames with expert consensus annotations (no crowdlabeling augmentation).

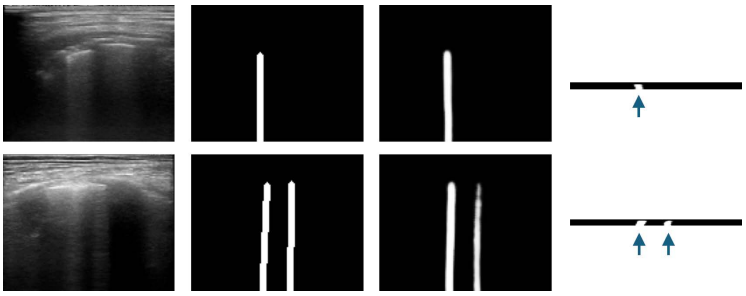
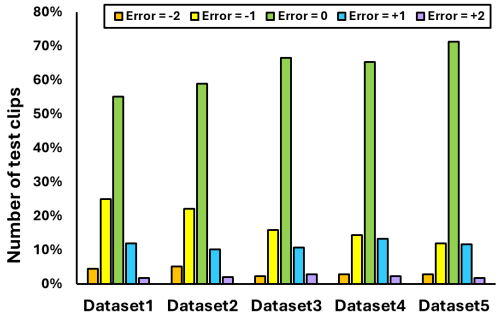


Fig. 3. Example images with one B-line (upper row) and two B-lines (lower row). Columns show ultrasound image converted to parallel scan lines, crowd consensus annotation, segmentation output, and resized output for B-line counting

3 Results and Discussion

The percentage of correctly estimated B-lines increased with training data size. The model trained only on 400 frames with expert annotations output the correct B-line count on only 18% of test clips, and was within ±1 B-line or ±2 B-lines on 24% and 29% of test clips, respectively. The number of test clips with correct B-line count was 55% for Dataset1, 59% for Dataset2, 66% for Dataset3, 65% for Dataset4, and 71% for Dataset5, and the number of clips within 1 B-line of the correct value increased to 84% for Dataset1, 91% for Dataset2, 93% for Dataset3, 93% for Dataset4, and 94% for Dataset5. Figure 4 illustrates the errors in the B-line counts that are computed from post-processing the results of the AUNet outputs. Compared to the Dataset 1 model accuracies, the Dataset3 model significantly improved accuracy ($p=0.001$ for exact B-line count, $p<0.0001$ for ±1

and ±2 B-line counts). Results for Datasets 4 and 5 were similarly significant at the α=0.05 level with FWER control.



Dataset	B-line count error		
Frames (N)	0	≤ 1	≤ 2
5,233	55.1%	84.3%	90.6%
7,703	59.0%	91.4%	98.7%
10,207	66.6%	93.2%	97.9%
13,298	65.3%	93.9%	98.2%
16,411	71.3%	94.5%	99.2%

Fig. 4. B-line count errors in the different models decrease by increasing training data amount from 5k to 16k: Dataset1 (5k), Dataset2 (7.5k), Dataset3(10k), Dataset4(13k), and Dataset5(16k). The bars show the number of test clips assessed correctly (green), within 1 B-line error (yellow and blue), and within 2 B-lines error (orange and purple).

Our study introduces an innovative approach to B-line segmentation in lung ultrasound (LUS) analysis. By using a novel gamified crowdsourcing technique, we collected over 550,000 B-line annotation opinions within three weeks, resulting in more than 30,000 high-quality segmentations which replicate the accuracy of clinician-produced segmentations at a much larger scale and greatly reduced cost. Each expert took an average of 15 seconds per frame annotation, totaling 8.3 hours across 400 frames. By crowdsourcing annotations for our full dataset (31K frames), we saved 650 expert hours.

We found that using more crowd-augmented data significantly improved model accuracy: our Dataset5 model achieved 94% accuracy in providing B-line counts for LUS videos within a margin of ±1 B-line, presenting a substantial improvement in both reliability and scalability for automatic B-line detection and scoring. One notable challenge encountered was annotation ambiguity from distinguishing between single and merged B-lines. Future improvements could involve using sectors to denote B-lines and calculating the occupied percentage of the intercostal space for better consistency in B-line quantification.

The practical implications of this research extend into clinical practice, especially in the management of heart failure patients, where accurate monitoring of pulmonary congestion is essential. By accurately counting and locating B-lines, the AI model we have developed has the potential to streamline diagnostic processes, reduce variability among clinicians, and support more consistent patient care in acute settings. While promising, these advancements represent a major step forward in lung ultrasound analysis rather than a complete solution, indicating the ongoing need for technological and methodological improvements to fully realize the benefits of AI in enhancing patient outcomes.

Acknowledgments. This study was funded by a Bits to Bytes grant from The Massachusetts Life Sciences Center (MLSC) and NIH grant 1R21EB034075-01A1.

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

References

1. Anantrasirichai, N., Hayes, W., Allinovi, M., Bull, D., Achim, A.: Line detection as an inverse problem: application to lung ultrasound imaging. IEEE Transactions on Medical Imaging **36**(10), 2045–2056 (2017)
2. Born, J., et al.: Accelerating detection of lung pathologies with explainable ultrasound image analysis. Applied Sciences **11**(2), 672 (2021)
3. Brattain, L.J., Telfer, B.A., Liteplo, A.S., Noble, V.E.: Automated B-line scoring on thoracic sonography. Journal of Ultrasound in Medicine **32**(12), 2185–2190 (2013)
4. Cardoso, M.J., Li, W., Brown, R., Ma, N., Kerfoot, E., Wang, Y., Murrey, B., Myronenko, A., Zhao, C., Yang, D., et al.: Monai: An open-source framework for deep learning in healthcare. arXiv preprint arXiv:2211.02701 (2022)
5. Chen, L., Rubin, J., Ouyang, J., Balaraju, N., Patil, S., Mehanian, C., Kulhare, S., Millin, R., Gregory, K.W., Gregory, C.R., et al.: Contrastive self-supervised learning for spatio-temporal analysis of lung ultrasound videos. In: 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI). pp. 1–5. IEEE (2023)
6. Duggan, N.M., Jin, M., Duran Mendicuti, M.A., Hallisey, S., Bernier, D., Selame, L.A., Asgari-Targhi, A., Fischetti, C.E., Lucassen, R., Samir, A.E., Duhaime, E.P., Kapur, T., Goldsmith, A.J.: Gamified crowdsourcing as a novel approach to lung ultrasound dataset labeling: Prospective analysis. J Med Internet Res **26**, e51397 (2024). https://doi.org/10.2196/51397
7. Farahi, M., Aranda, J., Habibian, H., Casals, A.: Automatic feature detection in lung ultrasound images using wavelet and radon transforms. arXiv preprint arXiv:2306.12780 (2023)
8. Fox, T.H., Gare, G.R., Hutchins, L.E., Perez, V.S., Rodriguez, R., Smith, D.L., Brito-Encarnacion, F.X., Danrad, R., Tran, H.V., Lowery, P.B., et al.: Artificial intelligence neural network consistently interprets lung ultrasound artifacts in hospitalized patients: A prospective observational study. medRxiv pp. 2023–03 (2023)
9. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In: International MICCAI Brainlesion Workshop. pp. 272–284. Springer (2021)
10. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H.R., Xu, D.: Unetr: Transformers for 3d medical image segmentation. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 574–584 (2022)
11. Howell, L., Ingram, N., Lapham, R., Morrell, A., McLaughlan, J.R.: Deep learning for real-time multi-class segmentation of artefacts in lung ultrasound. Ultrasonics p. 107251 (2024)
12. Jin, M., Duggan, N.M., Bashyakarla, V., Mendicuti, M.A.D., Hallisey, S., Bernier, D., Stegeman, J., Duhaime, E., Kapur, T., Goldsmith, A.J.: Expert-level annotation quality achieved by gamified crowdsourcing for b-line segmentation in lung ultrasound. arXiv preprint arXiv:2312.10198 (2023)

13. Karakuş, O., Anantrasirichai, N., Aguersif, A., Silva, S., Basarab, A., Achim, A.: Detection of line artifacts in lung ultrasound images of COVID-19 patients via nonconvex regularization. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control **67**(11), 2218–2229 (2020)
14. Li, G.Y., Chen, L., Zahiri, M., Balaraju, N., Patil, S., Mehanian, C., Gregory, C., Gregory, K., Raju, B., Kruecker, J., et al.: Weakly semi-supervised detector-based video classification with temporal context for lung ultrasound. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2483–2492 (2023)
15. Lichtenstein, D., Mezière, G., Biderman, P., Gepner, A., Barre, O.: The comet-tail artifact: an ultrasound sign of alveolar-interstitial syndrome. American Journal of Respiratory and Critical Care Medicine **156**(5), 1640–1646 (1997)
16. Lichtenstein, D.A., Meziere, G.A.: Relevance of lung ultrasound in the diagnosis of acute respiratory failure*: the blue protocol. Chest **134**(1), 117–125 (2008)
17. Mason, H., et al.: Lung ultrasound segmentation and adaptation between COVID-19 and community-acquired pneumonia. In: Int. Workshop Adv. Simplifying Med. Ultrasound (ASMUS). pp. 45–53. Springer (2021)
18. Mongodi, S., De Luca, D., Colombo, A., Stella, A., Santangelo, E., Corradi, F., Gargani, L., Rovida, S., Volpicelli, G., Bouhemad, B., et al.: Quantitative lung ultrasound: technical aspects and clinical applications. Anesthesiology **134**(6), 949–965 (2021)
19. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., Glocker, B., Rueckert, D.: Attention u-net: Learning where to look for the pancreas. In: Medical Imaging with Deep Learning (2018), https://openreview.net/forum?id=Skft7cijM
20. Ouyang, J., Chen, L., Li, G.Y., Balaraju, N., Patil, S., Mehanian, C., Kulhare, S., Millin, R., Gregory, K.W., Gregory, C.R., et al.: Weakly semi-supervised detection in lung ultrasound videos. In: International Conference on Information Processing in Medical Imaging. pp. 195–207. Springer (2023)
21. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI). pp. 234–241. Springer (2015)
22. Roy, S., et al.: Deep learning for classification and localization of COVID-19 markers in point-of-care lung ultrasound. IEEE Transactions on Medical Imaging **39**(8), 2676–2687 (2020)
23. Shea, D.E., Kulhare, S., Millin, R., Laverriere, Z., Mehanian, C., Delahunt, C.B., Banik, D., Zheng, X., Zhu, M., Ji, Y., et al.: Deep learning video classification of lung ultrasound features associated with pneumonia. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3102–3111 (2023)
24. van Sloun, R.J., Demi, L.: Localizing B-lines in lung ultrasonography by weakly supervised deep learning, in-vivo results. IEEE Journal of Biomedical and Health Informatics **24**(4), 957–964 (2019)
25. Tan, G.F.L., Du, T., Liu, J.S., Chai, C.C., Nyein, C.M., Liu, A.Y.L.: Automated lung ultrasound image assessment using artificial intelligence to identify fluid overload in dialysis patients. BMC nephrology **23**(1), 410 (2022)
26. Zhao, L., Fong, T.C., Bell, M.A.L.: Covid-19 feature detection with deep neural networks trained on simulated lung ultrasound b-mode images. In: 2022 IEEE International Ultrasonics Symposium (IUS). pp. 1–3. IEEE (2022)