



# Ultrasound-Based Detection of Lung Abnormalities Using Single Shot Detection Convolutional Neural Networks

Sourabh Kulhare<sup>1</sup>(✉), Xinliang Zheng<sup>1</sup>, Courosh Mehanian<sup>1</sup>,  
Cynthia Gregory<sup>2</sup>, Meihua Zhu<sup>2</sup>, Kenton Gregory<sup>1,2</sup>, Hua Xie<sup>2</sup>,  
James McAndrew Jones<sup>2</sup>, and Benjamin Wilson<sup>1</sup>

<sup>1</sup> Intellectual Ventures Laboratory, Bellevue, WA 98007, USA  
skulhare@intven.com

<sup>2</sup> Oregon Health Sciences University, Portland, OR 97239, USA

**Abstract.** Ultrasound imaging can be used to identify a variety of lung pathologies, including pneumonia, pneumothorax, pleural effusion, and acute respiratory distress syndrome (ARDS). Ultrasound lung images of sufficient quality are relatively easy to acquire, but can be difficult to interpret as the relevant features are mostly non-structural and require expert interpretation. In this work, we developed a convolutional neural network (CNN) algorithm to identify five key lung features linked to pathological lung conditions: B-lines, merged B-lines, lack of lung sliding, consolidation and pleural effusion. The algorithm was trained using short ultrasound videos of *in vivo* swine models with carefully controlled lung conditions. Key lung features were annotated by expert radiologists and sonographers. Pneumothorax (absence of lung sliding) was detected with an Inception V3 CNN using simulated M-mode images. A single shot detection (SSD) framework was used to detect the remaining features. Our results indicate that deep learning algorithms can successfully detect lung abnormalities in ultrasound imagery. Computer-assisted ultrasound interpretation can place expert-level diagnostic accuracy in the hands of low-resource health care providers.

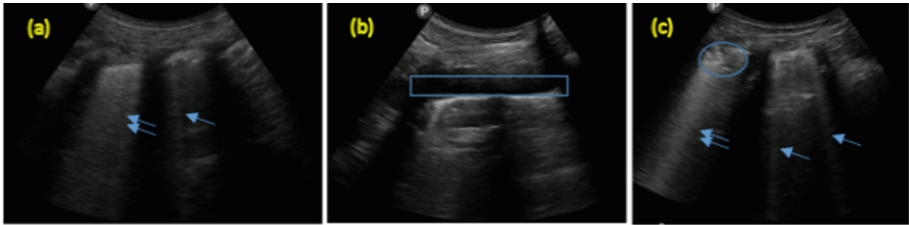
**Keywords:** Lung ultrasound · Deep learning · Convolutional neural networks

## 1 Introduction

Ultrasound imaging is a versatile and ubiquitous imaging technology in modern healthcare systems. Ultrasound enables skilled sonographers to diagnose a diverse set of conditions and can guide a variety of interventions. Low cost ultrasound systems are becoming widely available, many of which are portable and have user-friendly touch displays. As ultrasound becomes more available and easier to operate, the limiting factor for adoption of diagnostic ultrasound will become the lack of training in interpreting images rather than the cost and complexity of ultrasound hardware. In remote settings like small health centers, combat medicine, and developing-world health care systems, the lack of experienced radiologists and skilled sonographers is already a key limiting factor for the effectiveness of ultrasound imaging. Recent advances in artificial

intelligence provide a potential route to improve access to ultrasound diagnostics in remote settings. State of the art computer vision algorithms such as convolutional neural networks have demonstrated performance matching that of humans on a variety of image interpretation tasks [1].

In this work, we demonstrate the feasibility of computer-assisted ultrasound diagnosis by using a CNN-based algorithm to identify abnormal pulmonary conditions. Ultrasound in most cases does not show any structural information from within the lung due to the high impedance contrast between the lung, which is mostly air, and the surrounding soft tissue. Despite this, lung ultrasound has gained popularity in recent years as a technique to detect pulmonary conditions such as pneumothorax, pneumonia, pleural effusion, pulmonary edema, and ARDS [2, 3]. Skilled sonographers can perform these tasks if they have been trained to find the structural features and non-structural artifacts correlated with disease. These include abstract features such as A-lines, B-lines, air bronchograms, and lung sliding. Pleural line is defined in ultrasound as a thin echogenic line at the interface between the superficial soft tissues and the air in the lung. A-line is a horizontal artifact indicating a normal lung surface. The B-line is an echogenic, coherent, wedge-shaped signal with a narrow origin in the near field of the image. Figure 1 shows examples of ultrasound lung images.



**Fig. 1.** Ultrasound images from swine modeling lung pathologies that demonstrate (a) single (single arrow) and merged B-lines (double arrow), (b) pleural effusion (box), and (c) single and merged B-lines along with consolidation (circle).

Lung ultrasound is an ideal target for computer-assisted diagnosis because imaging the lung is relatively straightforward. The lungs are easy to locate in the thorax and precise probe placement and orientation is not necessary to visualize key features. By selecting a target that is relatively easy to image but complicated to interpret, we maximize the potential benefit of the algorithm to an unskilled user.

Computer processing of ultrasound images is a well-established field. Most methods focus on tools that assist skilled users with metrology, segmentation, or tasks that expert operators perform inconsistently, unaided [4]. Methods for detecting B-lines have previously been reported [5–7]. A recent survey [8] outlines deep learning work on ultrasound lesion detection but there has been less work on consolidation and effusion. Other examples include segmentation and measurement of muscle and bones [9], carotid artery [10], and fetus orientation [11]. Note that while these efforts utilize CNNs, their goal is segmentation and metrology, as opposed to computer-assisted diagnosis.

To show the effectiveness of CNN-based computer vision algorithms for interpreting lung ultrasound images, this work leverages swine models with various lung pathologies, imaged with a handheld ultrasound system. We include an overview of the swine models and image acquisition and annotation procedures. We provide a description of our algorithm and its performance on swine lung ultrasound images. Our detection framework is based on single shot detection (SSD) [12], an efficient, state-of-the-art deep learning system suitable for embedded devices such as smart phones and tablets.

## 2 Approach

### 2.1 Animal Model, Data Collection and Annotation

All animal studies and ultrasound imaging were performed at Oregon Health & Science University (OHSU), following Institutional Animal Care and Use Committee (IACUC) and Animal Care and Use Review Office (ACURO) approval. Ultrasound data from swine lung pathology models were captured for both normal and abnormal lungs. Normal lung features included pleural lines and A-lines. Abnormal lung features included B-lines (single and merged), pleural effusion, pneumothorax, and consolidation. Models of 3 different lung pathologies were used to generate ultrasound data with one or more target features. For normal lung data collection (i.e. pleural line and A-line data collection), all animals were scanned prior to induction of lung pathology. For pneumothorax and pleural effusion ultrasound features, swine underwent percutaneous thoracic puncture of one hemithorax followed by injection of air and infusion with saline into the pleural space of the other hemithorax, respectively. For consolidation, single and merged B-line ultrasound features, in separate swine, acute respiratory distress syndrome (ARDS) was induced by inhalation of nebulized lipopolysaccharide. Examples of ultrasound images acquired from the animal studies are shown in Figs. 1 and 2.

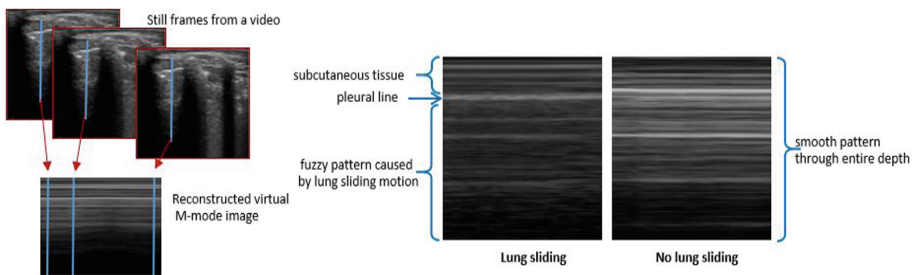


Fig. 2. Reconstruction of simulated M-mode images (left) and examples images (right).

Ultrasound data were acquired using a Lumify handheld system with a C5-2 broadband curved array transducer (Philips, Bothell, WA, USA). All images were acquired after selecting the Lumify app’s lung preset. Per the guidelines for point-of-care lung ultrasound [13], the swine chest area was divided into eight zones. For each

zone, at least two 3-s videos were collected at a frame rate of approximately 20 per second. One exam was defined as the collection of videos from all eight zones at each time point. Therefore, at least 16 videos were collected in each exam. For each swine, the lung pathology was induced incrementally and therefore, multiple exams were performed on each swine. Approximately 100 exams were performed with 2,200 videos collected in total. Lung ultrasound experts annotated target features frame-by-frame using a custom Matlab-based annotation tool.

## 2.2 Data Pre-processing

Input data for pre-processing consisted of either whole videos or video frames (images). Frame-level data was used to locate A-lines, single B-lines, merged B-lines, pleural line, pleural effusion, and consolidation. Video-level data was used for representation of pneumothorax. Raw ultrasound data collected from a curvilinear probe take the form of a polar coordinate image. These raw data were transformed from polar coordinates to Cartesian, which served to eliminate angular variation among B-lines and accelerate learning. The transformed images were cropped to remove uninformative data, such as dark borders and text, resulting in images with a resolution of  $801 \times 555$  pixels.

Video data were similarly transformed to Cartesian coordinates. Each transformed video was used to generate simulated M-mode images. An M-mode image is a trace of a vertical line (azimuthal, in the original polar image) over time. The vertical sum threshold-based method [7] was used to detect intercostal spaces. Each intercostal space was sampled to generate ten M-mode images at equally spaced horizontal locations.

Ultrasound video of a healthy lung displays lung sliding, caused by the relative movement of parietal and visceral pleura during respiration. This can readily be observed in M-mode images, where there is a transition to a “seashore” pattern below the pleural line. Pneumothorax prevents observation of the relative pleural motion and causes the M-mode image to appear with uniform horizontal lines as shown in Fig. 2.

## 2.3 Single Shot CNN Model for Image-Based Lung Feature Detection

Single Shot Detector (SSD) is an extension of the family of regional convolutional neural networks (R-CNNs) [14–16]. Previous object detection methods used a de-facto two network approach, with the first network responsible for generating region proposals followed by a CNN to classify each proposal into target classes. SSD is a single network that applies small convolutional filters (detection filters) to the output feature maps of a base network to predict object category scores and bounding box offsets. The convolutional filters are applied to feature maps at multiple spatial scales to enable detection of objects of various sizes. Furthermore, multiple filters representing default bounding boxes of various aspect ratios are applied at each spatial location to detect objects of varying shapes. This architecture renders SSD an efficient and accurate object detection framework [17], making it a suitable choice for on-device inference tasks. Figure 3 provides an overview of the SSD architecture. Details can be found in [12].

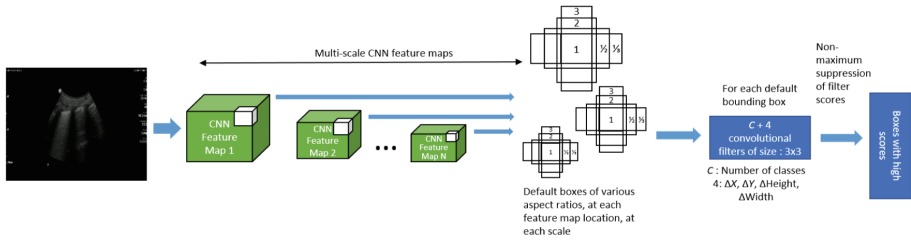


Fig. 3. SSD network schematic

**Training.** Each detection filter in SSD corresponds to a default bounding box at a particular location, at a particular scale, and aspect ratio. Prior to training, each ground truth bounding box is matched against the default bounding box with maximum Jaccard overlap. It is also matched against any default bounding box with Jaccard overlap greater than a threshold (usually 0.5). Thus, each ground truth box may be matched to more than one default box, which makes the learning problem smoother. The training objective of SSD is to minimize an overall loss that is a weighted sum of localization loss and confidence loss. Localization loss is Smooth L1 loss between location parameters of the predicted box and the ground truth box. Confidence loss is the softmax over multiple class confidences for each predicted box. We used horizontal flip, random crop, scale, and object box displacement as augmentations for training the lung features CNN models. For training the lung sliding model, we used Gaussian blur, random pixel intensity and contrast enhancement augmentations.

**Hyperparameters.** We use six single-class SSD networks as opposed to a multi-class network because the training data is small and unbalanced. Pleural lines and A-lines are abundant as they are normal lung features, whereas pathological lung features are rare. Furthermore, pleural line and pleural effusion features are in close proximity, thus there is significant overlap between their bounding boxes. Closely located features, combined with an unbalanced, small training set compromises performance when trained on multi-class SSD. We plan to address these issues in future work.

The train and test set sizes for each detection model are shown in Table 1. Feature models were trained for 300k iterations with batch size of 24, momentum 0.9, and initial learning rate of 0.004 (piece-wise constant learning rate that is reduced by 0.95 after every 80k iterations). We used the following aspect ratios for default boxes: 1, 2, 3, 1/2, 1/3, and 1/4. The base SSD network, Inception V2 [18], started with pre-trained ImageNet [19] weights and was fine-tuned for lung feature detection. The training process required 2–3 days per feature with the use of one GeForce GTX 1080Ti graphics card.

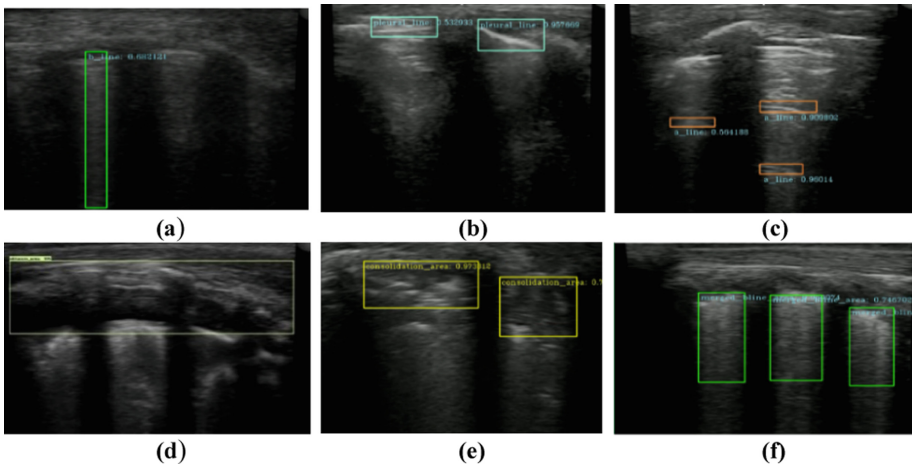
## 2.4 Inception V3 Architecture for Video-Based Lung Sliding Detection

Lung sliding was detected using virtual M-mode images that were generated by the process described in Sect. 2.2. We trained a binary classifier based on the Inception V3 CNN architecture [18]. Compared to V2, Inception V3 reduces the number of

**Table 1.** Training statistics and testing performance

Feature	Training set (frames)	Testing set (videos)	Sensitivity (%)	Specificity (%)
B-line	16,300	212	28.0	93.0
Merged B-line	14,961	337	85.0	96.5
B-line (combined)	–	521	88.4	93.0
A-line	10,510	580	87.2	89.0
Pleural line	48,429	640	85.6	93.1
Pleural effusion	21,200	143	87.5	92.2
Consolidation	18,713	444	93.6	86.3
Pneumothorax	13,255*	35	93.0	93.0

\*6,743 M-mode images with lung sliding, 6,512 M-mode images without lung sliding



**Fig. 4.** Sample results for SSD detection models. Detected features are highlighted by bounding boxes and confidence scores. (A) B-line, (B) pleural line, (C) A-line, (D) pleural effusion, (E) consolidation, (F) merged B-line.

convolutions, limiting maximum filter size to  $3 \times 3$ , increases the depth of the network and uses an improved feature combination technique at each inception module. We initialized Inception V3 with pre-trained ImageNet weights and fine-tuned only the last two classification layers with virtual M-mode images. The network was trained for 10k iterations with batch size 100 and a constant learning rate of 0.001.

### 3 Results

We compare single class SSD performance with threshold-based detection methods [7, 20], which are effective only for pleural line and B-line features. The SSD framework is applicable to all lung ultrasound features and our SSD detection model detects pleural lines with 89% accuracy compared to 67% with threshold-based methods.

Our CNN models were evaluated against holdout test dataset acquired from two swine. Table 1 shows the final test results and Fig. 4 shows sample outputs for features other than lung sliding. The pleural effusion model detected effusion at all fluid volumes from 50 mL to 600 mL (300 mL shown). Pleural line was the most common lung feature, present in most ultrasound videos. Videos without pleural line were uncommon, making the specificity calculation unreliable. The absence of an intercostal space in a video was treated as a pleural line negative sample. Note that for consolidation, pleural effusion and merged B-lines, sensitivity and specificity metrics are defined on a per video basis, rather than per object.

The algorithm achieved at least 85% in sensitivity and specificity for all features, with the exception of B-line sensitivity. There exists a continuum of B-line density from single B-lines, to dense B-lines, to merged B-lines. We observed that in many cases, dense B-lines that were not detected by the B-line detection model were detected by the merged B-line model. We combined the B-line and merged B-line output with the idea that the distinction between these two classes may be poorly defined. The combined B-line model achieved 88.4% sensitivity and 93% specificity, which was significantly better than B-lines alone. The video-based pneumothorax model had the highest overall accuracy with 93% sensitivity and specificity.

### 4 Conclusions and Future Work

In summary, we demonstrated that a CNN-based computer vision algorithm can achieve a high level of concordance with an expert's observation of lung ultrasound images. Seven different lung features critical for diagnosing abnormal lung conditions were detected with greater than 85% accuracy. The algorithm in its current form would allow an ultrasound user with limited skill to identify the abnormal lung conditions outlined here. This work with swine models is an important step toward clinical trials with human patients, and an important proof of concept for the ability of computer vision algorithms to effect automated ultrasound image interpretation.

In the future, we will continue this work using clinical patient data. This will help validate the method's efficacy in humans while providing a sufficient diversity of patients and quantity of data to determine patient-level diagnostic accuracy. We are also working to implement this algorithm on tablets and smartphones. To help with runtime on mobile devices, we are streamlining the algorithm to combine the six parallel SSD models into a single multi-class model, while eliminating the need for coordinate transformations, which represents the bulk of the computational time during inference.

## References

1. Litjens, G., et al.: A survey on deep learning in medical image analysis. *Med. Image Anal.* **42**, 60–88 (2017)
2. Testa, A., Soldati, G., Copetti, R., Giannuzzi, R., Portale, G., Gentiloni-Silveri, N.: Early recognition of the 2009 pandemic influenza A (H1N1) pneumonia by chest ultrasound. *Crit. Care* **16**(1), R30 (2011)
3. Parlamento, S., Copetti, R., Bartolomeo, S.D.: Evaluation of lung ultrasound for the diagnosis of pneumonia in the ED. *Am. J. Emerg.* **27**(4), 379–384 (2009)
4. Weitzel, W., Hamilton, J., Wang, X., Bull, J., Vollmer, A.: Quantitative lung ultrasound comet measurement: method and initial clinical results. *Blood Purif.* **39**, 37–44 (2015)
5. Anantrasirichai, N., Allinovi, M., Hayes, W., Achim, A.: Automatic B-line detection in paediatric lung ultrasound. In: 2016 IEEE International Ultrasonics Symposium (IUS), Tours, France (2016)
6. Moshavegh, R., et al.: Novel automatic detection of pleura and B-lines (comet-tail artifacts) on in vivo lung ultrasound scans. In: *SPIE Medical Imaging 2016* (2016)
7. Fang, S., Wang, Y.R.B.: Automatic detection and evaluation of B-lines by lung ultrasound. NYU, New York City
8. Huang, Q., Zhang, F., Li, X.: Machine learning in ultrasound computer-aided diagnostic systems: a survey. *BioMed Res. Int.* (2018)
9. Jabbar, S., Day, C., Heinz, N., Chadwick, E.: Using Convolutional Neural Network for edge detection in musculoskeletal ultrasound images. In: *International Joint Conference on Neural Networks*, pp. 4619–4626 (2016)
10. Shin, J., Tajbakhsh, N., Hurst, R., Kendall, C., Liang, J.: Automating carotid intima-media thickness video interpretation with convolutional neural networks. In: *Conference on Computer Vision and Pattern Recognition, Las Vegas*, pp. 2526–2535 (2016)
11. Chen, H., et al.: Standard plane localization in fetal ultrasound via domain transferred deep neural networks. *IEEE J. Biomed. Health Inform.* **19**(5), 1627–1636 (2015)
12. Liu, W., et al.: SSD: Single Shot MultiBox Detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) *ECCV 2016*. LNCS, vol. 9905, pp. 21–37. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-46448-0\\_2](https://doi.org/10.1007/978-3-319-46448-0_2)
13. Volpicelli, G., et al.: International evidence-based recommendations for point-of-care lung ultrasound. *Intensive Care Med.* **38**(4), 577–591 (2012)
14. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 580–587 (2014)
15. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788 (2016)
16. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: *Advances in Neural Information Processing Systems (NIPS)*, pp. 91–99 (2015)
17. Huang, J., et al.: Speed/accuracy trade-offs for modern convolutional object detectors. In: *The Conference on Computer Vision and Pattern Recognition* (2017)
18. Szegedy, C., Vanhoucke, V., Loffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: *The Conference on Computer Vision and Pattern Recognition* (2015)



19. Jia, D., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: The Conference on Computer Vision and Pattern Recognition (2009)
20. Omar, Z., et al.: An explorative childhood pneumonia analysis based on ultrasonic imaging texture features. In: 11th International Symposium on Medical Information Processing and Analysis, vol. 9681 (2015)