Contents lists available at ScienceDirect

Software Impacts

journal homepage: www.journals.elsevier.com/software-impacts

Original software publication

vid-SAMGRAH: A PyTorch framework for multi-latent space reinforcement learning driven video summarization in ultrasound imaging

Roshan P. Mathews [a], Mahesh Raveendranatha Panicker [a,*], Abhilash R. Hareendranathan [b]

[a] *Indian Institute of Technology Palakkad, Kerala, India*
[b] *University of Alberta, Alberta, Canada*

ARTICLE INFO

Keywords:
Ultrasound
Video summarization
Unsupervised reinforcement learning
Attention ensemble encoders
Python
PyTorch

ABSTRACT

The COVID-19 pandemic has accelerated the need for automatic triaging and summarization of ultrasound videos for fast access to pathologically relevant information in the Emergency Department and lowering resource requirements for telemedicine. In this work, a PyTorch based unsupervised reinforcement learning methodology which incorporates multi feature fusion to output classification labels, segmentation maps and summary videos for lung ultrasound is presented. The use of unsupervised training eliminates tedious manual labeling of key-frames by clinicians opening new frontiers in scalability in training using unlabeled or weakly labeled data. Our approach was benchmarked against expert clinicians from different geographies displaying superior Precision and F1 scores (over 80% and 44%).

Code metadata

Current code version	v1
Permanent link to code/repository used for this code version	https://github.com/SoftwareImpacts/SIMPAC-2021-164
Permanent link to Reproducible Capsule	https://codeocean.com/capsule/8503804/tree/v1
Legal Code License	MIT License
Code versioning system used	git
Software code languages, tools, and services used	python (Anaconda distribution), Spyder (IDE)
Compilation requirements, operating environments & dependencies	python 3.8, PyTorch 1.10, segmentation_models_pytorch, torchvision, albumentations, opencv, h5py, numpy, matplotlib, shutil, tkinter, streamlit. Please see https://github.com/rpm1412/LUS_Video_Summarization/blob/main/requirements.txt
If available Link to developer documentation/manual	https://arxiv.org/abs/2109.01309
Support email for questions	mahesh@iitpkd.ac.in

1. Introduction

The COVID-19 pandemic has amplified the diagnostic potential of ultrasound imaging for continuous monitoring as it is free from ionizing radiation, portable and hardly requires dedicated facilities making it an economical diagnostic tool when compared to other imaging modalities [1–3]. But the implementation of it (or any imaging modality) in a practical scenario faces a two fold challenge viz, (1) the dearth of human experts (clinicians) to interpret large volumes of data generated, and (2) the time constraint of clinicians to provide inferences for large amounts of data. To bring a perspective of the volume of data generated, consider a typical ultrasound video for lung assessment at 30 fps for 30 s produces about a thousand frames for just a single assessment. This large volume of data generated poses two major challenges — (1) as the time of clinicians is limited and manually combing through huge volumes of data becomes impractical and, (2) large data files necessitate larger bandwidth and storage requirements for telemedicine over the internet for expert annotation and diagnosis.

With an aim to solve the above bottleneck of growing disproportionate volumes of data versus available human experts to analyze them,

The code (and data) in this article has been certified as Reproducible by Code Ocean: (https://codeocean.com/). More information on the Reproducibility Badge Initiative is available at https://www.elsevier.com/physical-sciences-and-engineering/computer-science/journals.

[*] Corresponding author.

E-mail addresses: rpmedu22@gmail.com (R.P. Mathews), mahesh@iitpkd.ac.in (M.R. Panicker), hareendr@ualberta.ca (A.R. Hareendranathan).

https://doi.org/10.1016/j.simpa.2021.100185
Received 14 November 2021; Received in revised form 22 November 2021; Accepted 22 November 2021

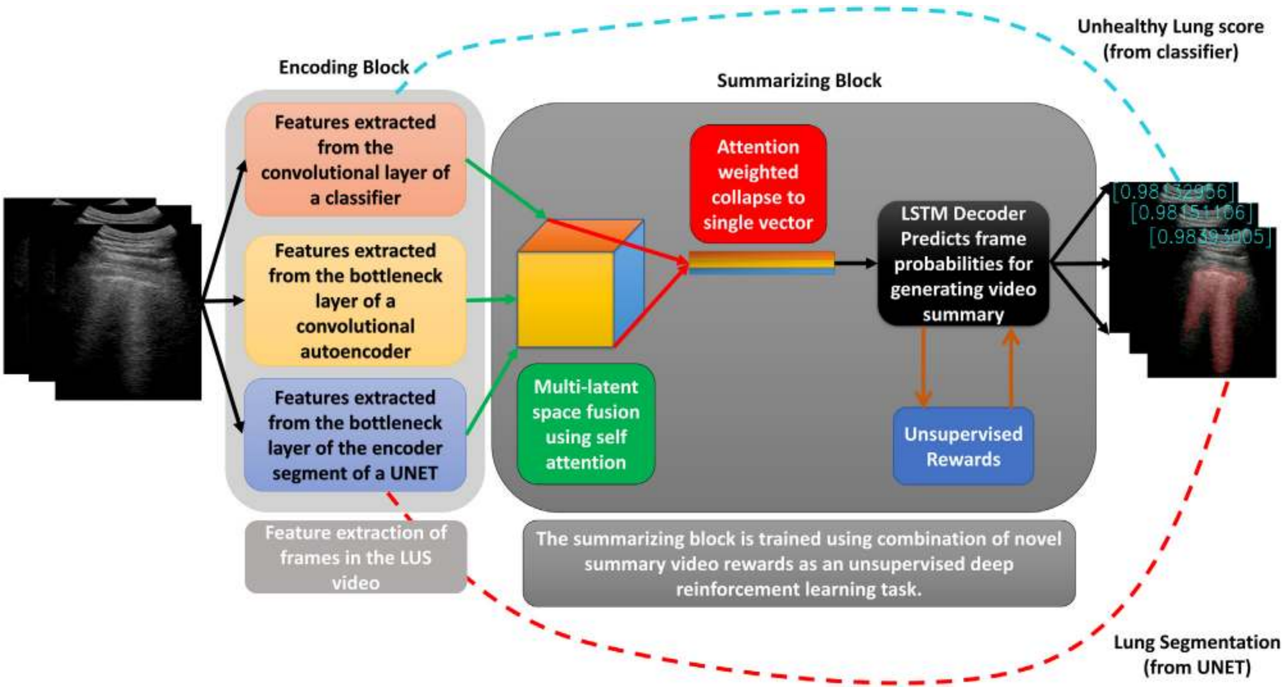

Encoding Block
Features extracted from the convolutional layer of a classifier
Features extracted from the bottleneck layer of a convolutional autoencoder
Features extracted from the bottleneck layer of the encoder segment of a UNET
Feature extraction of frames in the LUS video
Summarizing Block
Multi-latent space fusion using self attention
Attention weighted collapse to single vector
LSTM Decoder Predicts frame probabilities for generating video summary
Unsupervised Rewards
The summarizing block is trained using combination of novel summary video rewards as an unsupervised deep reinforcement learning task.
Unhealthy Lung score (from classifier)
[0.98132966]
[0.98151106]
[0.98393005]
Lung Segmentation (from UNET)

Fig. 1. Graphical abstract of the video summarization algorithm. The input video frames are encoded using standard encoder architectures and then summarized using an LSTM decoder. The summarized videos contain pathologically relevant key frames and are overlaid with classification scores and segmentation maps for easy diagnosis by clinicians.

we have developed a fast and reliable methodology to interface the gap between machine data and human experts. A PyTorch based multi-latent space reinforcement learning driven video summarization tool is presented which will help to extract the relevant key frames from a given ultrasound video. The key frames within the video convey diagnostically relevant information by overlaying machine classification score (healthy–unhealthy lung) and highlighting pathologically relevant features like B-lines, Pleura, and A-lines making it easier for the clinician to make the final decision. The tool will aid the human expert by providing key frames with overlaid pathological segmentations that reduce the amount of manual interventions as well as lower the storage and bandwidth requirements in telemedicine.

2. Description

To perform the task of video summarization and provide machine scores with pathological markings, an unsupervised multi-latent space based reinforcement learning (RL) methodology is employed. The algorithm is designed to provide robust and superior video summarization capability by carefully analyzing the components in the summarization pipeline like the encoder–decoder pair and the formation of the reward function. In this paper, the summarization pipeline is touched upon with a graphical abstract in Fig. 1. and the complete implementation of the algorithm and its details can be found in [4] (the link to the paper is provided in the GitHub repo [5]). In addition to the paper, the code and trained models are also readily accessible from the GitHub repo which includes a sample set of 4 ultrasound videos provided to demonstrate the working of the summarization algorithm.

2.1. Encoder–decoder

The formation of latent vectors from multiple encoders (Classification, Segmentation and Auto-Encoding) provide the summarization network different aspects of the lung image making it a robust methodology for ultrasound video summarization. Unlike natural images, ultrasound images are characterized by important pathological features which are crucial for diagnosis. For e.g., a classifier trained to distinguish between healthy and unhealthy lungs might focus on the presence of B-lines, Subpleural consolidations among other features to categorize while a segmentation network would focus on the texture and intensity to segment the lung into anatomical structures like Pleura, A and B-lines whereas an autoencoder forms the most compact representation of the image frame and hence the three encoders are responsible for encoding different aspects of the lung image. By employing a fusion of the three (i.e the classification network, segmentation network and the auto-encoding network) we are able to obtain a representation of the image that is robust for summarization. The fusion of these multi-latent space features are performed using an attention mechanism to ensure relevant features are fused in order to obtain the best summarization possible. Since videos are frames in a sequence the summarization is performed using the LSTM architecture [6] which intuitively handles sequences and incorporates temporal information between frames. The LSTM classifies the frames in the videos as key frames (assigns probabilities to frames indicative of its importance to be included in the summary) which is then used to form a summary by selecting the top frame-probability scores.

2.2. Rewards

The summarization algorithm is trained following an unsupervised RL methodology [7,8]. The ground truth clinical annotation of keyframes in videos becomes impractical due to the clinician needing to comb through thousands of frames among the videos to annotate key frames. In ultrasound videos, keyframes are often few and far between which makes it even harder for clinicians to reliably label these videos. Using RL, training can be done by forming suitable reward functions to encourage the algorithm to select diagnostically relevant key frames based on pathological features without the need for manual labeling by a radiologist. The selection of such key frames are enforced by novel components of the reward such as (1) *clsf*: a score based on health state of the lung which helps preferentially pick out frames that are unhealthy (obtained from the classifier network), (2) *ssim*: a structural similarity index score to promote the selection of dissimilar frames and (3) *rep* + *div*: a representative and diversity score [7] to obtain a balanced summarization between representative and diverse frames of the video.

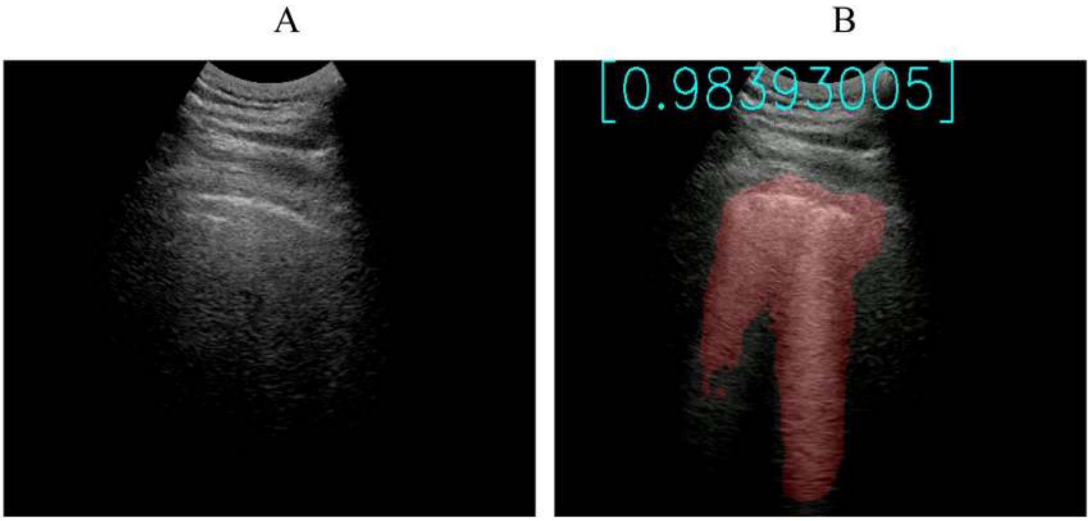


Fig. 2. Preferential (diagnostically important) summarization aspect of the algorithm. (A) Presence of ambiguous regions in the ultrasound scan and (B) the preferential selection of pathologically relevant structures like the B-lines for the summary.

Table 1
Quantitative Metrics for the video summarization methodology.

Approach	Encoder - LSTM Decoder	Precision	Recall	F_1	ReF
General	Classification	46.38	16.11	23.91	78.62
	Segmentation	61.58	24.87	35.43	75.08
	Auto-Encoding	59.46	22.84	33.00	78.11
Our	**Attention Encoder Fusion**	**80.24**	**30.37**	**44.06**	**77.29**

Table 2
Quantitative Metrics for ablation studies involving rewards in parts.

Approach	Rewards	Encoder-LSTM Decoder	Precision	Recall	F_1
General	rep + div	Classification	46.93	16.26	24.15
		Segmentation	60.78	24.20	34.62
		Auto-Encoding	58.19	22.30	32.24
Our	**clsf + ssim**	**Attention Encoder Fusion**	**77.33**	**29.30**	**42.50**

2.3. Results

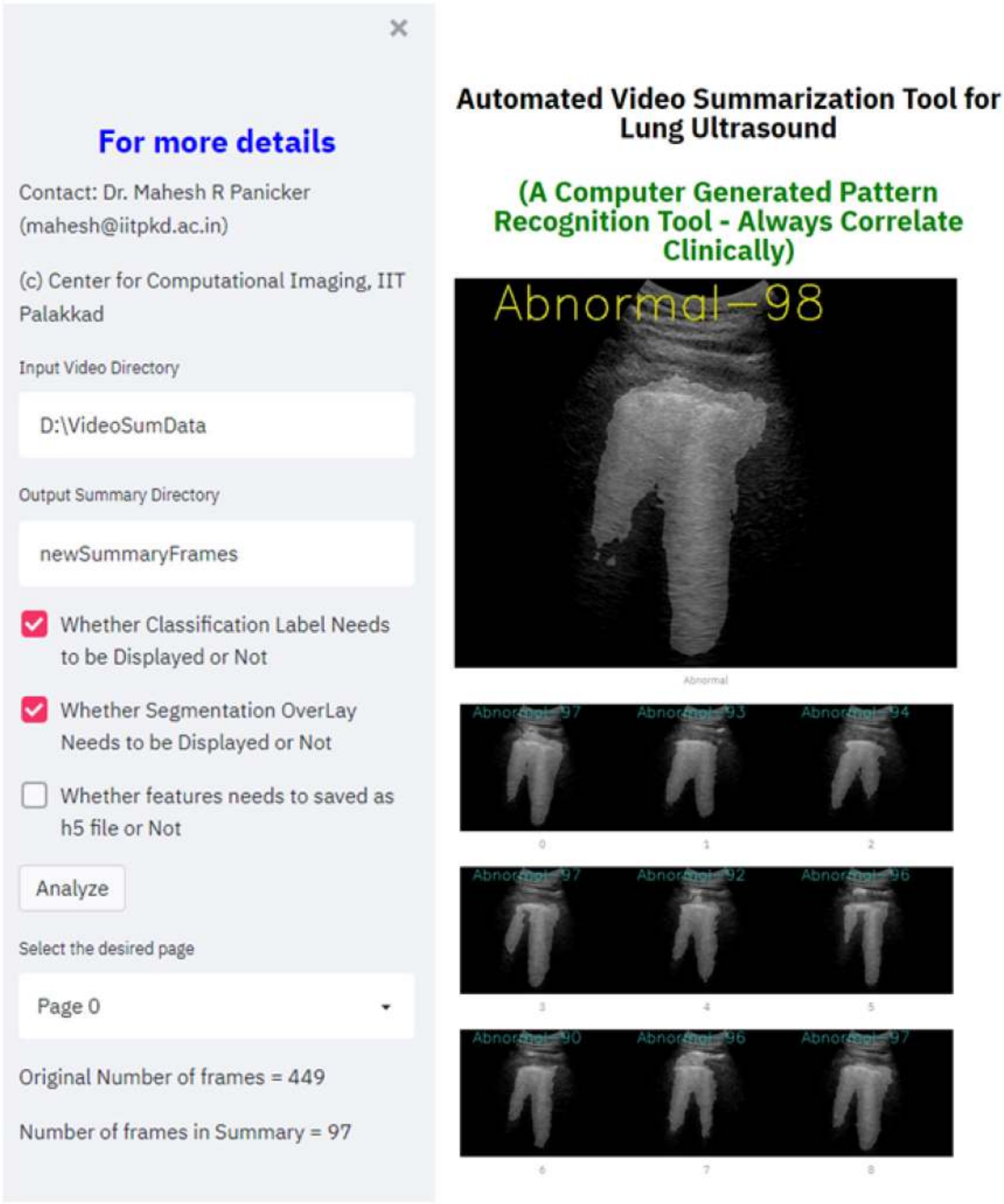
The summarization network was trained using 100 lung ultrasound videos obtained from 3 countries during the pandemic (2020) and tested using an independent set of 26 ultrasound videos from 2021 which were annotated by expert clinicians from 3 geographies to form the ground truth to benchmark our system. The unsupervised summarization algorithm is trained using all the rewards mentioned in Section 2.2 and the results are presented in Table 1. The high Precision (80%), F_1-Score (44%) and Reduction (ReF) (77%) displays the superior and robust summarization capability of our approach of using attention based fusion of latent vectors for unsupervised schemes. Since we are limited to only selecting a small subset (less than one-fourth the length of the whole video for the summary as noted by the reduction factor (ReF) %) of equally likely summary frames (in ultrasound videos multiple frames convey similar diagnostic information), a lower value of recall is expected. It is immediately clear when comparing our results to non latent space fusion based approaches that the fusion of the encoders is essential to produce an accurate and robust video summary. A summarization result for the attention encoder fusion is presented in Fig. 2. to illustrate the summarization and overlaying of machine scores for lung health and segmentations of pathologically relevant features. In addition, we also validate the efficacy of our novel rewards that are introduced (*clsf* + *ssim*) which are tuned to the application of lung ultrasound video summarization. The results from Table 2. are supportive of the use of focused rewards that help the algorithm in preferential summarization of diagnostically relevant frames over general rewards.

3. Software impact

In the wake of the ongoing and post COVID-19 years, the necessity and application of telemedicine are well understood as a means of remote consultation [9]. One such is the point of care ultrasound (POCUS) which has been one of the central themes in telemedicine [10–12]. POCUS eliminates the need for the physical presence of an expert radiologist by offsetting it with large scan data obtained by sparingly trained technicians to encompass essential diagnostic information. Hence surplus data is to be sent to expert radiologists for referral via telemedicine. This poses a two fold problem (1) telemedicine is severely limited throughout the world by the available bandwidth and storage, (2) the time availability of expert radiologists. Thus developing software to make the obtained data suitable for telemedicine transmission and aid clinicians in making decisions quickly are of the utmost necessity and the work presented here is a step in that direction aimed at developing focused methods to make video summarization (in this case lung ultrasound) robust by using concepts from artificial intelligence.

The software developed herein has great capability in speeding up imaging and diagnosis/ prognostication of pulmonary cases that is particularly helpful during the COVID-19 pandemic when the medical facilities and practitioners are stretched to the limits. As caseloads increase exponentially, the proposed video summarization methodology helps in monitoring disease progression on a periodical basis and in evaluating the lung involvement for many patients using a simple set up of a portable ultrasound machine. From a societal perspective, this software innovation can transform emergency departments (ED), community scanning centers or patients at home with a mere ultrasound facility or POCUS into a good diagnostic location able to mimic radiology services of a full fledged hospital with sophisticated imaging modalities like CT that are beyond the reach of millions due to patient load, financial outgo and availability of radiologists.

With the introduction of this video summarization methodology two main challenges have been overcome. First, the need for an expert's physical presence, time and concentration for performing the study as identifying and interpreting anomalies had been considerably relaxed. This is achieved through software segregation of diagnostically relevant information from long ultrasound scans that can be performed by technicians or junior doctors with minimal experience. The algorithm summarizes the long videos into key frames that are diagnostically



For more details

Contact: Dr. Mahesh R Panicker (mahesh@iitpkd.ac.in)

(c) Center for Computational Imaging, IIT Palakkad

Input Video Directory

D:\VideoSumData

Output Summary Directory

newSummaryFrames

☑ Whether Classification Label Needs to be Displayed or Not

☑ Whether Segmentation OverLay Needs to be Displayed or Not

☐ Whether features needs to saved as h5 file or Not

Analyze

Select the desired page

Page 0

Original Number of frames = 449

Number of frames in Summary = 97

Automated Video Summarization Tool for Lung Ultrasound

(A Computer Generated Pattern Recognition Tool - Always Correlate Clinically)

Fig. 3. Web-application deployment. The tool has the option to select the classification labels and the segmentation maps to be overlaid on the frames. Provision is also given to save the frame features. The tool will show the summary video, a collage of random 9 frames from the summary video and the frame numbers.

relevant and provide highlights of machine scores (labels that predict whether the lung is healthy or unhealthy) as well as overlays pathological segmentations (A-lines, B-lines and Pleura). This highlighted summary can then be sent to expert radiologists via telemedicine for final diagnosis, making it easier for the radiologist at a remote location to judge the progression of the disease quickly. Second, the use of a summarization algorithm makes it possible to summarize large volumes of data into smaller sizes. The algorithm is capable of providing a robust summary with high precision by using just one-fourth the original video length. This reduces the storage and transmission bandwidth requirement by a quadruple factor thereby enabling better communication over low bandwidth internet connections that are typical in many remote locations. The final output achieved is a web-application software for video summarization that summarizes the given video succinctly and provides information related to diagnosing the case with machine overlaid scores and segmentations of pathologically relevant features as shown in Fig. 3.

Finally, unlike natural videos, medical videos are critical as missing information in the summary is deleterious. Medical videos across modalities are diverse — the different presets of the machines used, the skill of the radiologist involved in obtaining the scan, geographical differences etc. making it harder to obtain a robust summary. To overcome this challenge we propose the approach of latent vector fusion to increase the reliability by using a multi-feature map from the video to summarize it. This paves way for further research into standardizing procedures for scans, video and feature preprocessing to compensate for different presets in order to make summarization more reliable and robust.

4. Limitations and future development

The work presented here is a proof of concept towards a robust ultrasound video summarization software. At present the system is trained and validated with ultrasound scans from various countries and different ultrasound machines by separate clinicians. Future work would include analyzing the proposed system with better pruned US scans i.e. data from distinct machine vendors, standardized ultrasound scans with specific presets which are expected to further increase the performance of the proposed methodology. Also, the methodology described above is not limited to lung ultrasound and can easily be extrapolated to other ultrasound videos like wrist, elbow or liver ultrasound, as well as to other imaging modalities.

5. Conclusion

We have developed a robust ultrasound video summarization tool for lung ultrasound to aid clinicians by preferentially selecting diagnostically important frames for summarizing and overlaying them with machine lung health score as well as highlighting pathologically relevant features. This will make it easier and quicker for the clinicians to reach a diagnosis for the progression of the disease. This will also result in shorter videos making it suitable for storage and transmission in case of POCUS as well as telemedicine which is arguably the future trend in ultrasound radiology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to acknowledge the funding from the Department of Science and Technology - Science and Engineering Research Board, India (DSTSERB (CVD/2020/000221)) for the CRG COVID19 funding. The authors are grateful to Compute Canada and the CDAC-NVIDIA-PARAM SIDDHI AI system for providing the computing resources for the project. The authors wholeheartedly thank the team of expert clinicians, Dr. Yale Tung Chen, Dr. Jacob L Jaremko, Dr. Brian Buchanan, Dr. Kiran Vishnu Narayan, Dr. Kesavadas C and Dr. Greeta Mathews for their best efforts in data collection, expert advice and medical annotations despite the busy times during the pandemic that helped benchmark our system performance.

Illustrative examples

An illustrative video example of the working of the proposed algorithm can be found at https://youtu.be/Th-XGQWRvpo titled ***Demo for Video Summarization***.

References

[1] Igor Barjaktarevic, Jon-Émile S. Kenny, David Berlin, Maxime Cannesson, The evolution of ultrasound in critical care: From procedural guidance to hemodynamic monitor, J. Ultrasound Med. Off. J. Am. Inst. Ultrasound Med. 40 (2) (2021) 401.

[2] Jennifer Oluku, Attila Stagl, Kamalpreet S. Cheema, Karmen El-Raheb, Richard Beese, The role of point of care ultrasound (PoCUS) in orthopaedic emergency diagnostics, Cureus 13 (1) (2021) e13046.

[3] Giovanni Volpicelli, Luna Gargani, Stefano Perlini, Stefano Spinelli, Greta Barbieri, Antonella Lanotte, Gonzalo García Casasola, Ramon Nogué-Bou, Alessandro Lamorte, Tomas Villén, Lung ultrasound for the early diagnosis of COVID-19 pneumonia: An international multicenter study, Intensive Care Med. 47 (4) (2021) 444–454.

[4] Roshan P Mathews, Mahesh Raveendranatha Panicker, Abhilash R Hareendranathan, Yale Tung Chen, Jacob L Jaremko, Brian Buchanan, Kiran Vishnu Narayan, Greeta Mathews, Unsupervised multi-latent space reinforcement learning framework for video summarization in ultrasound imaging, 2021, arXiv preprint arXiv:2109.01309.

[5] Roshan P Mathews, Mahesh Raveendranatha Panicker, Unsupervised multi-latent space reinforcement learning framework for video summarization in ultrasound imaging, 2021, GitHub repository https://github.com/rpm1412/LUS_Video_Summarization.

[6] Felix A. Gers, Jürgen Schmidhuber, Fred Cummins, Learning to forget: Continual prediction with LSTM, Neural Comput. 12 (10) (2000) 2451–2471.

[7] Kaiyang Zhou, Yu Qiao, Tao Xiang, Deep reinforcement learning for unsupervised video summarization with diversity-representativeness reward, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32, 1, 2018.

[8] Tianrui Liu, Qingjie Meng, Athanasios Vlontzos, Jeremy Tan, Daniel Rueckert, Bernhard Kainz, Ultrasound video summarization using deep reinforcement learning, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2020, pp. 483–492.

[9] Rashid Bashshur, Charles R. Doarn, Julio M. Frenk, Joseph C. Kvedar, James O. Woolliscroft, Telemedicine and the COVID-19 pandemic, lessons for the future, Telemedicine E Health 26 (5) (2020) 571–573.

[10] Marina Carbone, Vincenzo Ferrari, Michele Marconi, Roberta Piazza, Andrea Del Corso, Daniele Adami, Quintilia Lucchesi, Valeria Pagni, Raffaella Berchiolli, A tele-ultrasonographic platform to collect specialist second opinion in less specialized hospitals, Updates Surg. 70 (3) (2018) 407–413.

[11] Christina Herrero, Yhan Colon, Akash Nagapurkar, Pablo Castañeda, Point-of-care ultrasound reduces visit time and cost of care for infants with developmental dysplasia of the hip, Indian J. Orthop. (2021) 1–6.

[12] Timothy T. Tran, Maung Hlaing, Martin Krause, Point-of-care ultrasound: Applications in low-and middle-income countries, Curr. Anesthesiol. Rep. (2021) 1–7.