DOI: 10.61782/fa.2025.0255

CROSS-DOMAIN TRANSFER LEARNING FOR SEGMENTATION OF LUNG ULTRASOUND IMAGES

M. Muñoz[1,2*] **A. Rubio**[1,2] **G. Cosarinsky**[1]
J. Camacho[1]
[1] Institute for Physical and Information Technologies, Spanish National Research Council, Madrid, Spain
[2] Electronic Department, Universidad de Alcalá, Madrid, Spain

ABSTRACT

Developing deep learning models for lung ultrasound (LUS) segmentation is hindered by limited annotated data. This study demonstrates the effectiveness of cross-domain transfer learning (TL) using ImageNet pre-trained encoders (VGG16, ResNet50, MobileNetV2) within a U-Net architecture. Compared to training from scratch, TL with ResNet50 and VGG16 significantly improved segmentation Dice score (up to 30%). Furthermore, TL allowed reducing the training dataset to 45% while achieving performance comparable to the baseline trained on full data. These findings validate cross-domain TL as a valuable and data-efficient approach for LUS analysis.

Keywords: Artificial Intelligence (AI), Transfer Learning, Fine-tuning, Segmentation, Lung Ultrasound (LUS).

1. INTRODUCTION

Artificial Intelligence (AI) has emerged as a transformative technology with significant potential in medical imaging and diagnostics. In particular, AI-based approaches hold promise for improving the accuracy and efficiency of diagnostic procedures in various medical domains, including pulmonary imaging [1, 2, 3]. Lung ultrasound imaging plays an important role in diagnosing respiratory conditions, offering real-time visualization of pulmonary lesions [4]. However, the interpretation of lung ultrasound images can be challenging, and requires expertise to identify subtle features and abnormalities accurately [5].

AI-based Semantic segmentation, is a powerful approach for assessing LUS interpretation. However, a critical challenge in applying deep learning models for segmentation is the limited availability of large, annotated datasets. The process of acquiring LUS images and obtaining accurate annotations from expert radiologists is both time-consuming and expensive. To mitigate this bottleneck, some research efforts focus on developing automatic labeling tools [6]. While promising, transfer learning offers a complementary strategy by leveraging knowledge from pre-trained models, potentially reducing the amount of labeled data and training time required to achieve high performance.

Transfer learning has the potential to address the challenges associated with acquiring and labeling large datasets for comprehensive model training, making it a promising approach for enhancing the diagnostic utility of AI-based segmentation models in pulmonary ultrasound imaging [7].

The objective of this study is to investigate the application of cross-domain transfer learning techniques for lung ultrasound image segmentation, aiming to harness the power of pre-trained convolutional neural network (CNN) models for diagnostic assistance in pulmonary imaging and, to evaluate the impact of reduced training data on accuracy. Specifically, architectures such as VGG16, ResNet50 and MobileNetV2 are studied as pre-trained encoders within an Attention U-Net architecture.

*Corresponding author: mario.munoz.prieto@csic.es.

2. METHODS

2.1 CNN architectures

For semantic segmentation of LUS images, an Attention U-Net architecture [8] was employed. This network is based on the U-Net [9], which follows an autoencoder design (Figure 1). An autoencoder consists of two primary components: an encoder that progressively reduces the spatial dimensions of the input image while extracting increasingly abstract features, and a decoder that upsamples these features to reconstruct an output of the same size as the input. In U-Net based networks, skip connections between corresponding encoder and decoder layers facilitate the preservation of fine-grained details. The Attention U-Net incorporates spatial and channel attention modules [8] to further enhance feature discrimination.

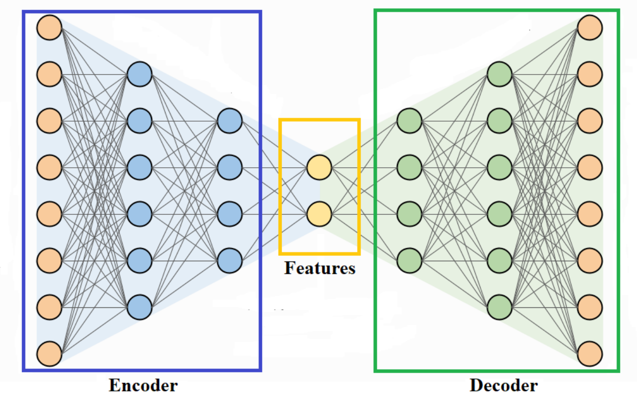

Figure 1. Schematic Autoencoder network architecture

To leverage transfer learning, the encoder portion of the Attention U-Net was replaced with pre-trained networks. Transfer learning is a technique where a model developed for a task is reused as the starting point for a model on a second task. In this case, models pre-trained on the large ImageNet dataset for image classification were utilized as encoder of the new architectures adapting them for LUS image segmentation. This study investigated three such pre-trained models: VGG16 [10], ResNet50 [11], and MobileNetV2 [12]. The classification layers of these models were removed, and feature maps were extracted from a specific layer of each encoder serving as input to the Attention U-Net decoder, which remained consistent across all configurations.

2.2 Dataset

This study utilized a dataset of 5131 lung ultrasound (LUS) images obtained from 30 patients diagnosed with COVID-19. The images were acquired as part of the clinical study described in [13]. The dataset exhibits typical LUS artifacts associated with pneumonia, including A-lines, B-lines, and consolidations.

Image annotation was performed using a semi-automatic labeling tool presented in [6]. This tool employs signal processing algorithms to generate initial segmentation masks. The results of this semiautomatic annotation tool were validated at the video level by a LUS expert physician. The annotations delineate the boundaries of artifacts such as pleura, A-lines, B-lines, and consolidations.

2.3 Methodology

This study evaluated cross-domain transfer learning for LUS image segmentation. An Attention U-Net, trained from scratch, was compared to three transfer learning configurations using pre-trained encoders (VGG16, ResNet50, MobileNetV2). Each configuration was trained four times with different random weight initializations to account for the inherent stochasticity of the training process. Performance was assessed using the mean and standard deviation of the Dice Similarity Coefficient (DSC) between the mean values on a reserved test set.

The dataset was divided into training (60%), validation (20%), and testing (20%) sets using a patient-wise split to provide a more realistic assessment of generalization performance. All models were trained with a learning rate of 3e-5, a batch size of 32, and the binary cross-entropy loss function. A two-stage transfer learning procedure was used: encoder layers were frozen for the first 60 epochs, then unfrozen for fine-tuning during the remaining 60 epochs. The reference model was trained from scratch for 120 epochs. Training was performed using TensorFlow 2.10 on two NVIDIA 2080 Ti GPUs.

The best-performing encoder (based on DSC) was then used in a data reduction experiment. Training set size was reduced 5% by 5% until the model achieved a validation DSC value below the reference model. The training procedure was repeated four times for each reduced dataset aiding consistency to the results. The model is stopped when the DSC is below the baseline model.

3. RESULTS

Table 1 shows the mean and standard deviation of the Dice Similarity Coefficient (DSC) for each model applying

Transfer learning (TL) and fine-tunning (FN) calculated on the test set.

Table 1. Segmentation performance.

Model Att-Unet	Dice (mean±std)
From Scratch	0.544 ± 0.034
VGG16 (TL+FN)	0.691 ± 0.033
MobileNetV2 (TL+FN)	0.516 ± 0.152
ResNet50 (TL+FN)	0.711 ± 0.043

Using ResNet50 as encoder with TL and FN achieved the highest mean DSC (0.71 ± 0.04), a statistically significant improvement over the baseline Attention U-Net (0.54 ± 0.03). With VGG16 as encoder (TL+FN) also significantly outperformed the baseline (0.69 ± 0.03). These results showed no statistically significant difference between Resnet50 and VGG16 encoders. MobileNetV2 (TL+FN) exhibited the lowest DSC (0.52 ± 0.15).
Figure 2 illustrates an example of the differences on segmentation performance between the Attention U-Net base model and the Resnet50-based Attention U-Net.

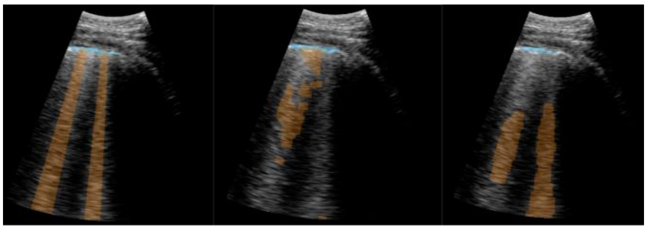

Figure 2. Segmentation performance comparison between the reference model (from scratch) and the Resnet50-Att. U-Net with TL and FN.

Following the initial comparison, the best-performing encoder (ResNet50) was used to investigate the impact of reduced training data. Models were trained on progressively smaller subsets of the training data, ranging from 95%. Figure 3 illustrates the mean DSC (± standard deviation) achieved modifying the percentage of training data used. The results indicate that leveraging transfer learning with the ResNet50 encoder allowed the training dataset size to be reduced to approximately 45% while still achieving a mean DSC comparable to the baseline Attention U-Net model trained from scratch on the full (100%) training dataset.

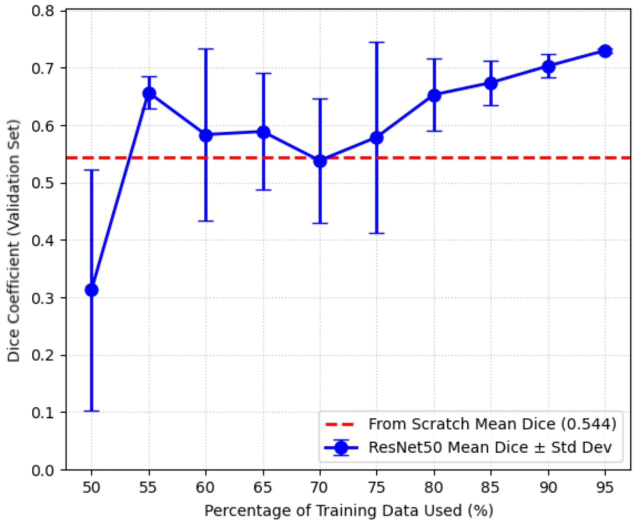

Figure 3. Data reduction results over Resnet50-based Attention U-Net fine-tuned.

4. DISCUSSION

This study investigated the effectiveness of cross-domain transfer learning for LUS image segmentation. The results demonstrate its advantages: pre-trained ResNet50 and VGG16 encoders significantly outperformed the baseline Attention U-Net trained from scratch. This suggests that, despite the significant domain difference between natural images (ImageNet) and medical ultrasound, features learned during pre-training are valuable for extracting relevant patterns in all type of images. While ResNet50 yielded the highest mean Dice score, VGG16's performance was comparable. On the contrary, MobileNetV2 performed worse, potentially because its design, focused on efficiency via parameter reduction, leads to lower representational capacity or features less transferable to this specific cross-domain task.

The data reduction study further highlights the value of transfer learning, with results quantitatively presented in Figure 3. This figure plots the validation Dice score of the ResNet50 transfer learning model against the percentage of training data used. It demonstrates that performance comparable to the baseline model (DSC ≈ 0.54) was maintained even when using only 55% of the original training data. This represents a substantial 45% reduction in the required annotated dataset, underscoring the practical benefit of cross-domain transfer learning in mitigating the need for extensive data annotation. Interestingly, the performance trend was not strictly constant; a noticeable dip towards the baseline performance occurred around the 70% data mark where the mean performance approached or slightly fell below the baseline threshold. However, a sustained drop significantly below the baseline threshold was only observed for data levels below 55%. Furthermore, Figure 3 illustrates increased performance variability (larger standard deviations) at lower data percentages (particularly ≤75%), indicating greater training instability or sensitivity to data sampling when data becomes severely limited. Notably, while the standard deviation observed at the 55% data level appears comparatively low relative to surrounding data percentages in this experiment, this is likely to be due to the stochastic nature of training with only four runs, rather than an indication of particular stability at this specific data level.

It is important to acknowledge certain limitations of this study. Notably, exhaustive hyperparameter optimization was not performed for each encoder, potentially impacting individual model performance. Furthermore, results are based on a specific COVID-19 LUS dataset, which may limit generalizability. Future work should address these limitations by performing systematic hyperparameter tuning and evaluating performance on larger, more diverse LUS datasets encompassing various pathologies. Investigating more recent encoder architectures and comparing cross-domain pre-training with medical domain pre-training could also yield valuable insights.

Despite these limitations, this work provides evidence that cross-domain transfer learning, particularly with ResNet50 or VGG16, is a highly effective and data-efficient approach for LUS segmentation.

5. CONCLUSIONS

This study evaluated the application of cross-domain transfer learning techniques for the semantic segmentation of LUS images. The results conclusively demonstrate that using pre-trained architectures such as VGG16 and ResNet50 as encoders within an Attention U-Net network significantly improves segmentation performance compared to training from scratch. A marked increase in the Dice Similarity Coefficient (DSC) of approximately 30% was observed when employing transfer learning with the most effective encoders.

A key additional finding is the potential of transfer learning to mitigate the need for large annotated datasets. The data reduction analysis indicated that it is possible to decrease, in this specific example, the training dataset size to approximately 45% of its original size while still achieving segmentation performance comparable to the baseline model trained on the full dataset. This data efficiency considerably reduces the burden associated with manual labeling and model training, facilitating the development of AI tools in data-limited settings.

Finally, this work validates the utility of cross-domain transfer learning for medical imaging, and specifically for lung ultrasound. It confirms that features learned from large-scale natural image databases can be effectively transferred and adapted to improve analysis in specialized medical domains such as LUS, despite the inherent differences between the image types. Overall, transfer learning presents itself as a practical strategy for developing accurate and efficient LUS segmentation models.

6. ACKNOWLEDGMENTS

This research was partially supported by the project PID2022-143271OB-I00, funded by MCIN/AEI /10.13039/501100011033/FEDER, UE; and supported by the European Commission + NextGenerationEU, through Momentum CSIC Programme: Develop Your Digital Talent. The funding for these actions/grants and contracts comes from the European Union's Recovery and Resilience Facility-Next Generation, in the framework of the General Invitation of the Spanish Government's public business entity Red.es to participate in talent attraction and retention programmes within Investment 4 of Component 19 of the Recovery, Transformation and Resilience Plan. G. Cosarinsky staff is hired under the Generation D initiative, promoted by Red.es, an organisation attached to the Ministry for Digital Transformation and the Civil Service, for the attraction and retention of talent through grants and training contracts, financed by the Recovery, Transformation and Resilience Plan through the European Union's Next Generation funds.

7. REFERENCES

[1] Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. Future Healthc J. 2021 Jul;8(2): e188-e194. doi: 10.7861/fhj.2021-0095. PMID: 34286183; PMCID: PMC8285156.

[2] Zhao, L., & Lediju Bell, M. A. (2022). A review of deep learning applications in lung ultrasound imaging

of COVID-19 patients. BME Front., 2022, 9780173. https://doi.org/10.34133/2022/9780173X

[3] Mento F, Perrone T, Fiengo A, Smargiassi A, Inchingolo R, Soldati G, Demi L. Deep learning applied to lung ultrasound videos for scoring COVID-19 patients: A multicenter study. J Acoust Soc Am. 2021 May;149(5):3626. doi: 10.1121/10.0004855. PMID: 34241100.

[4] Tung-Chen Y, Martí de Gracia M, Díez-Tascón A, Alonso-González R, Agudo-Fernández S, Parra-Gordo ML, Ossaba-Vélez S, Rodríguez-Fuertes P, Llamas-Fuentes R. Correlation between Chest Computed Tomography and Lung Ultrasonography in Patients with Coronavirus Disease 2019 (COVID-19). Ultrasound Med Biol. 2020 Nov;46(11):2918-2926. doi: 10.1016/j.ultrasmedbio.2020.07.003. Epub 2020 Jul 13. PMID: 32771222; PMCID: PMC7357528.

[5] Herraiz JL, Freijo C, Camacho J, Muñoz M, González R, Alonso-Roca R, Álvarez-Troncoso J, Beltrán-Romero LM, Bernabeu-Wittel M, Blancas R, et al. Inter-Rater Variability in the Evaluation of Lung Ultrasound in Videos Acquired from COVID-19 Patients. Applied Sciences. 2023; 13(3):1321.

[6] Muñoz, M., Rubio, A., Cosarinsky, G., Cruza, J. F., & Camacho, J. (2024). Deep Learning-Based Algorithms for Real-Time Lung Ultrasound Assisted Diagnosis. Applied Sciences, 14(24), 11930. https://doi.org/10.3390/app142411930.

[7] Yu, X., Wang, J., Hong, Q.-Q., Teku, R., Wang, S.-H., & Zhang, Y.-D. (2022). Transfer learning for medical images analyses: A survey. Neurocomputing, 489, 230-254. https://doi.org/10.1016/j.neucom.2021.08.159.

[8] Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; et al. Attention u-net: Learning where to look for the pancreas. arXiv 2018, arXiv:1804.03999.

[9] Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015, Proceedings of the 18th International Conference, Munich, Germany, 5–9 October 2015; Proc., Part III 18; Springer International Publishing: Berlin/Heidelberg, Germany, 2015; pp. 234–241.

[10] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[11] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proc. of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[12] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).

[13] Camacho, J.; Muñoz, M.; Genovés, V.; Herraiz, J.L.; Ortega, I.; Belarra, A.; González, R.; Sánchez, D.; Giacchetta, R.C.; et al. Artificial Intelligence and Democratization of the Use of Lung Ultrasound in COVID-19: On the Feasibility of Automatic Calculation of Lung Ultrasound Score. Int. J. Transl. Med. 2022, 2, 17–25