2023 IEEE 24th International Conference on Information Reuse and Integration for Data Science (IRI) | 979-8-3503-3458-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/IRI58017.2023.00052

A Fully Connected Reproducible SE-UResNet for Multiorgan Chest Radiographs Segmentation

Debojyoti Pal
Artificial Intelligence and Data Science, Jio Institute
Navi Mumbai, India
debojyoti.pal@jioinstitute.edu.in

Tanushree Meena
Artificial Intelligence and Data Science, Jio Institute
Navi Mumbai, India
tanushree.meena@jioinstitute.edu.in

Sudipta Roy*
Artificial Intelligence and Data Science, Jio Institute
Navi Mumbai, India
sudipta1.roy@jioinstitute.edu.in

***Abstract*—Deep learning (DL) models are a popular choice for resolving intricate issues in medical imaging, such as the classification of diseases, detection of anomalies, and segmentation of tissues in real-world scenarios. To be useful in these contexts, the models must be able to provide accurate results for new, previously untrained data. Existing methods not only fail to consider the intrinsic features of small target lesions but are also not evaluated on separate datasets. To solve these problems we propose a novel architecture, SE-UResNet, capable of segmenting multiple organs having different size and shapes from Chest X-Ray (CXR) images. The proposed architecture introduces a residual module in between the encoding and decoding modules of an attention U-Net architecture for better feature representation of high-level features. The architecture also replaces the attention gates in the decoder module of attention U-Net with Squeeze and Excite (S&E) modules. SE-UResNet is experimented on benchmark CXR datasets such as NIH CXR for lungs, heart, trachea and collarbone segmentation as well as VinDr-RibCXR for ribs segmentation tasks with respect to other state-of-the-art segmentation models. The proposed model achieves an average DSC of 95.9%, 76.8%, 78.7%, 78.8%, and 86.0% for lungs, trachea, heart, collarbone and ribs segmentation for the aforementioned datasets. Furthermore, the proposed model has only been tested on two benchmark CXR datasets: Shenzen and JSRT to establish the reproducibility and robustness of the model. The performance of SE-UResNet on several benchmark CXR datasets demonstrates the model's ability to generalize, making it a reliable baseline for medical image segmentation. Furthermore, it can also be used for assessing the reproducibility of DL models based on their performance on different datasets.**

***Index Terms*—Medical imaging, Chest X-Ray, Deep learning, Segmentation, Reproducible model**

I. Introduction

Automatic segmentation of medical images has become an important area of research in medical image analysis due to the increasing volume of medical data and the need for accurate and efficient diagnosis and treatment [1]. Chest X-rays (CXRs) [2], in particular, have been used extensively in the diagnosis and management of respiratory diseases such as COVID-19 [3], which has led to a surge in the volume of medical data. This process not only saves time and reduces the burden on medical experts in annotating large datasets, but it also helps to ensure consistency and accuracy in the annotations. In addition to disease detection and diagnosis, automatic annotation can also be useful in medical research for identifying patterns and trends in large datasets. This can lead to new insights into the underlying causes of diseases and improve the effectiveness of treatments.

The deep learning models [4] have made considerable progress in interpreting and annotating medical images, especially in the regions of interest with a large area. The ability to extract features by stacking convolutional layers has led to the development of fully convolutional layers (FCNs) [5] for medical image segmentation. Nevertheless, their application is limited due to their reliance on a large number of annotated data and their requirement of a considerable time frame to train the model. The development of U-Net [6] as a segmentation model solved the issue of requiring large datasets for accurate prediction of segmentation masks. The usage of skip connections [7] to reintroduce the features lost during encoding of feature vectors proved effective in generating segmentation masks. However, U-Net failed in providing accurate segmentation masks for small lesions. Attention is a key mechanism in deep learning that enables neural networks to selectively focus on informative features or parts of the input data while ignoring the irrelevant ones. Attention gates have been integrated into various U-Net architectures [8], [9]. Various Convolutional Neural Networks (CNNs) [10], [11], [12] and Transformer models [13] have been integrated as backbone to the U-Net architecture to improve its accuracy for small lesion segmentations. However, they fail to account for the anatomical details relevant for proper segmentation masks and require a large training dataset. A brief overview of previous studies is illustrated in Table I To address the aforementioned limitations, we present a novel segmentation model for CXR images called SE-UResNet. This generalized model is designed to produce precise segmentation masks for multiple organs using a small database for training. Moreover, SE-UResNet can generate accurate segmentation masks automatically for untrained datasets based on its prior training. The key contributions of this work are outlined below:

- We introduce a residual block which acts as a bridge between the encoding and decoding modules. The shortcut connections present in the residual module forces the model to better learn the difference feature maps

979-8-3503-3458-6/23/$31.00 ©2023 IEEE
DOI 10.1109/IRI58017.2023.00052

TABLE I
A BRIEF SUMMARY OF PREVIOUS STUDIES

DL Techniques	Pre-processing Techniques	Datasets	Advantages	Disadvantages	Evaluation Metrices
ResNet [4]	Data Standardisation, Data normalisation, pixel-wise scaling	BraTS, LiTS, DIARETDB1, MICCAI Head and Neck auto segmentaion Challenge 2015	A content-based image retrival network for multi-modal image segmentation and classification.	The model uses pre-trained network, which will hinder chances of reproducibility for newer complex datasets.	Mean Average Precision, Accuracy.
S3+C2 (FCN) [5]	Data normalisation, zero-centring followed by pixel-wise scaling	CVC-EndiSceneStill, Kvasir-SEG	DL model combining FCN with atrous convolution for automatic polyp segmentation.	The model has high trainable parameters, use pre-trained weights and requires large number of images to train.	Mean Intersection over Union, Dice Score Coefficient
U-Net [6]	Overlap-tile strategy, followed by extrapolating missing context by mirroring technique	ISBI cell tracking challenge 2015	Introduces a novel, efficient u-shaped encoder decoder architecturecapable of generating accurate segmentation masks with less training data	The model poses information loss in the bottleneck, which connects the encoder with decoder. This results inaccurate segmentation masks for small lesion segmentation tasks	Warping Error, Rand Error and Pixel Error.
U-Net ++ [7]	Data Normalisation, data resizing followed by pixel-wise rescaling	Data Science Bowl, ASU-Mayo, MICCAI LiTS Challenge 2018, LIDC-IDRI	Introduces novel arrangement of skip connections in U-Net architecture for better feature representation, making the model efficient and generalisable.	The skip connection arrangement for individual layers is unnecessary and often leads to generation of extra pixels in segmentation mask	Intersection over Union
Attn U-Net [8]	Data Normalisation	TCIA Pancreas CT-82, Abdominal CT-150	Introduces novel attention gates which enhances model performancefor small lesion segmentation tasks	The model is prone to over-fitting resulting in additional pixels while generating output masks.	Dice Similarity Coefficient, Precision, Recall, and surface to surface distance
Attn UW-Net [9]	Data Normalisation, resizing the image to 128x128	NIH CXR dataset	Introduces novel intermediate block making the model capable of automatically segmenting multiple organs from CXR using less training data	Generating False positive pixels in mask generation. The model is unable to segment ribs in CXR images	Specificity, Precision, Recall, and F1-score
MResU-Net [10]	Data Normalisation, applying contrast limited adaptive histogram equalization followed by pixel-scaling.	DRIVE, STARE	Modified U-Net model achieving benchmark performances for multiple datasets.	The model uses large number of training images and involves high training parameters.	Sensitivity, Specificity, F1-score, and G-mean
URes-Net [11]	Data Normalization, cropping of images followed by translation of image slice to center the target image.	OASIS-3	Lightweight DL model for registration and segmentation of T1 weight brain MRI images.	The model doesn't take into account properties such as inverse and symmetric constraints for image registration.	Dice Similarity Coefficient
Ens4B-UNet [12]	DICOM to RGB conversion followed by resizing to 512x512	SIIM-ACR Pneumothorax Segmentation Challenge	Ensemble of best performing U-Net architectures having CNN backbones.	Using pre-trained weights and post-processing techniques to attain high accuracies.	Intersection over Union, Accuracy, Precision, Recall, and F1 score
Swin U-Net [13]	Image resized to 224x224.	Synapse multi-organ segmentation dataset, Automated cardiac diagnosis challenge	Integrating swin transformers in the baseline U-Net architecture, for multiorgan segmentation.	High trainable parameters, pre-trained weights were assigned whichleads to suboptimal performance for multimodal image segmentation.	Dice Score Coefficient and Hausdorff Distance

between the encoder and decoder modules. This makes the model robust and sensitive to smaller, resulting in higher accuracy on relevant tasks.

- The proposed model replaces the attention gates with S&E modules. These modules make the model selectively emphasize informative features and suppress less relevant ones, thereby making the model generalized.
- SE-UResNet has been evaluated on two benchmark CXR datasets: NIH CXR dataset [14] for lung, heart, trachea and collarbone segmentation followed by VinDr-RibCXR dataset [15] for rib segmentation. The proposed model has been compared with SOTA segmentation models with respect to metrics such as Dice Score Similarity (DSC).
- To demonstrate the versatility of SE-UResNet, we evaluated its performance on Shenzen [16] and JSRT [17] CXR datasets, utilizing the weights learned from the NIH dataset. Our findings indicate that the model is effective in automatic segmentation, further highlighting its versatility and reproducibility.

II. METHODOLOGY

A. Dataset Description

The proposed model is evaluated on four benchmark datasets namely, NIH CXR [14], VinDr-RibCXR [15], Shenzen [16] and JSRT [17] as is mentioned in Table II. Among four datasets, only NIH CXR doesn't have any ground truth data. As a result the CXR images are individually annotated for the regions of lungs, heart, trachea and collarbone for training and evaluation purpose. The annotations are verified by in-house radiologists to generate precise ground truth masks. A subsection of 200 CXR images out of the 112,120 images of the NIH Chest X-Ray [14] dataset are annotated. The VinDr-RibCXR [15] contains annotations of ribs and is considered a benchmark dataset for segmentation of individual ribs from CXR scans. These two datasets were split into a ratio of 60:20:20 for training, validation and testing. The results of the proposed model on the test data is presented in Table II and a comparative study with SOTA models is presented in Table IV. The Shenzen [16] and JSRT [17] contain annotations of lungs verified by expert radiologists. The SE-UResNet is tested entirely on these datasets without training to evaluate the reproducibility of the model. The performance of the model on the Shenzen [16] and JSRT [17] datasets are reported in Table V.

TABLE II
SUMMARISED TABLE OF SEGMENTATION DATASETS USED IN EXPERIMENTS.

Dataset	No. of Images	Image Dimensions	Organs
NIH CXR [14]	200	1024x1024	Lungs, ,Trachea, Heart, and Collarbone
VinDr-RibCXR [15]	265	2048x2500	Ribs
Shenzen [16]	566	3000x2919	Lungs
JSRT [17]	247	256x256	Lungs

B. Method

The proposed SE-UResNet consists of an encoding module followed by a residual module and a decoding module as shown in Fig 1. The residual module acts as a bridge between the encoding and decoding module is responsible for learning the anatomically intrinsic features essential for segmentation of small lesions. To further enhance the model's performance and robustness we have incorporated squeeze-and-excite (S&E) module in the encoding, decoding, and residual module.

1) Encoding Module: This module is responsible for the extraction of higher-level features from the input image, therefore leading to a more abstract representation of the input as

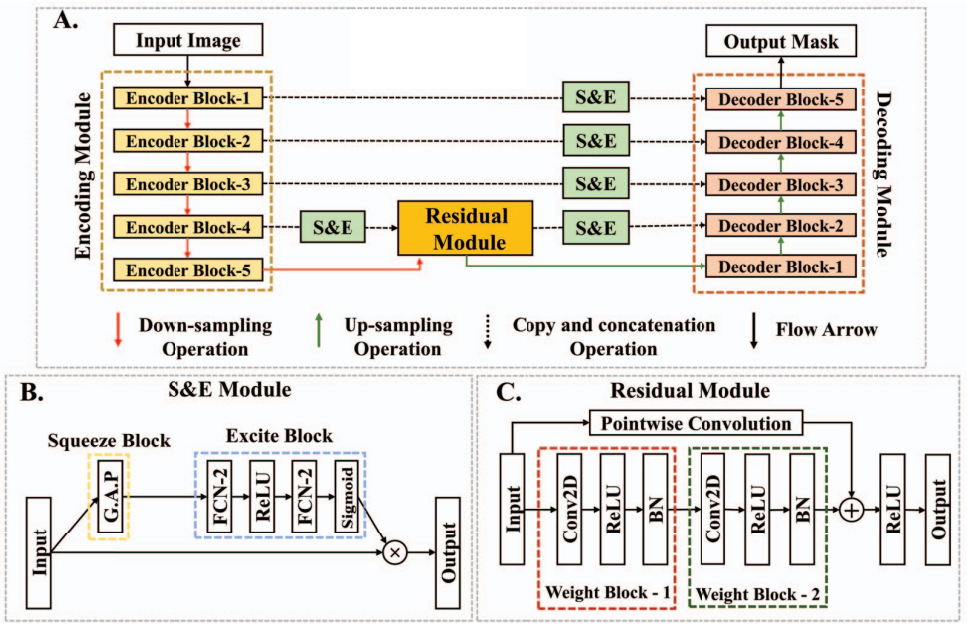

Fig. 1. The schematic representation of SE-UResNet. A. Overview of the SE-UResNet B. Detailed architecture of S&E module and C. Detailed architecture of Residual module.

we move through the layers. The encoding module comprises of five encoder blocks. Each encoder block comprises of two weight blocks connected with each other using a dropout layer. A weight block has a convolutional layer with a kernel size of 3×3 followed by a non-linear activation function (ReLU) and a batch-normalisation (BN) layer as represented in equation 1.

$$W(x) = max(0, Conv_{3\times3}(x)) + \beta \tag{1}$$

where W(x) represents the weight block for an input vector x, Conv 3×3, and β represents the convolutional operation for a filter size of 3×3 and the shift parameter learnt during BN operation. Individual encoder blocks are connected to each other via a down-sampling layer. The number of channels in the feature maps is doubled after each down-sampling operation to allow for the capture of more complex patterns and features. This process continues until the residual module is reached.

2) Residual Module: The final block of encoding module, mentioned as 'Encoder Block-5' in Fig. 1, is connected to the residual module using an up-sampling operation. The residual module R, comprises of two weight blocks, Weight block-1 and Weight block-2 as shown in Fig 1.C. A shortcut connection is also introduced wherein the input vector is sampled and passed through a pointwise convolutional layer. The resultant vector is then combined with the output of the weight block, W through an element-wise addition operation followed by a ReLU activation:

$$R(x) = max(0, Conv_{1\times1}(W(x) + x)) \tag{2}$$

where Conv 1×1 represents the pointwise convolutional operation and + represents the element-wise addition operation. The intuition behind this design is that the original input will act as an "identity mapping" that helps the network to better learn and approximate the underlying function. This shortcut connection ensures that even if the output of the operations in residual module is close to zero, the original input x, will be preserved in the final output. This in turn prevents the vanishing gradients problem.

3) Decoding Module: The output of the residual module is passed onto the decoder path. The decoder path comprises of five decoder module which is responsible for expanding the feature map back to the original size of the input image. The decoder path starts with an up-sampling layer, which increases the spatial resolution of the feature map. This is followed by a concatenation operation that merges the feature map with the corresponding feature map from the encoder path. However, we have introduced a S&E module for every corresponding feature map before the concatenation operation. This helps to preserve the fine-grained details of the image that may have been lost during down sampling. Furthermore, the S&E module helps the network to focus on the most important channels of the feature maps by adaptively weighting the channel-wise feature responses. The resulting feature map is then passed through a series of weight blocks to further process the features and extract more complex information. This process is repeated for each level of the decoder path, with each attention gate taking as input the up sampled feature map and the corresponding feature map from the encoder path at that level. Finally, the output of the last layer in the decoder path is passed through a sigmoid activation function to produce the final output of the model.

4) S&E Module: The S&E module are introduced to enhance the representation power of the weight blocks. Individual S&E module has two main blocks: squeeze block and excite block as shown in Fig 1.B. The squeeze block compresses the global information of the input feature vector using a global average pooling (G.A.P) layer. This results in a channel descriptor, which contains information about the feature map's important characteristics. The resultant vector is passed onto the excitation block which computes the attention for each channel. These weights are learned through a series of fully connected layers, FCN-1 and FCN-2 which are used to scale the original feature map. The resultant feature map is then combined with the input vector using an element-wise multiplication operation. S&E module enables SE-UResNet to identify the edges and contours of objects in the image without increasing the model complexity. As a result, every encoder block is connected to its corresponding decoder block via an S&E module. This enables the proposed model to effectively segment smaller regions such as trachea and collarbone. The inclusion of S&E module in the residual module reduces the number of false positive pixels. This results in the generation of accurate segmentation masks for complex segmentation tasks such as rib segmentation. The proposed arrangement of the module help reduce overfitting and improve the generalization performance of the network as is evident in Section -IV.

C. Implementation details

The proposed model was developed on TensorFlow v2.8.0 and trained on an Intel i9 processor with a RTX A4000 chip. Adam optimizer was used with a learning rate of 0.001. The model was trained for 150 epochs with a batch size of 8. Binary cross-entropy was used as a loss function with mean IoU as an accuracy function to keep track of epoch wise accuracy.

TABLE III
PERFORMANCE OF SE-URESNET ON MULTIORGAN SEGMENTATION TASK.

	DSC (in%)	IoU (in%)	Precision (in%)	Recall (in%)
Ribs	86.04 ± 3.9	75.62 ± 5.6	84.66 ± 4.6	87.35 ± 6.7
Trachea	76.88 ± 6.5	64.89 ± 8.0	78.64 ± 9.9	74.03 ± 10.1
Heart	78.79 ± 7.4	69.95 ± 9.8	78.39 ± 9.7	86.45 ± 7.5
Lungs	95.93 ± 1.5	92.11 ± 2.8	97.94 ± 1.6	93.96 ± 2.8
Collar bone	78.83 ± 6.4	65.52 ± 8.4	84.23 ± 8.3	84.28 ± 7.6

*Individual cell values are in the format of Mean ± SD

III. RESULTS AND DISCUSSION

The proposed SE-UResNet is producing the best effect on the mentioned experimental setup over the NIH CXR dataset [14] and VinDr Ribs dataset [15]. The proposed SE-UResNet has been evaluated on performance metrices such as DSC, mean IoU, precision and recall. Table III reports the performance of the proposed model with individual cells having a format of Mean ± Standard Deviation (SD). The proposed method attains the highest DSC of 95.93%, 81.86%, 76.88%, 78.83% and 86.04% for lung, heart, trachea, collarbone and ribs segmentation, respectively. The proposed model is compared with the ResUnet [10], Attention Unet [8], Swin Unet [13], Unet++ [7], Linknet-b7 [18], DC U-Net [19], UResNet [11], and BiSeNet-v2 [20]. DSC is chosen as performance metric to compare with the aforementioned models. The complete comparative analysis is illustrated in Table III whereas the qualitative analysis between the proposed model and other segmentation models family is shown in Fig.2.

The proposed SE-UResNet urpasses all other segmentation models and generates a segmented mask that almost perfectly matches the ground truth mask as is observed in Fig 2. Among the other segmented models, the Unet++, UResNet and attention U-Net are the best performing models for small lesion segmentation like collarbone. However, these models fail to predict the exact dimensions as well as the intrinsic details of the right collarbone. The Linknet-b7, Swin U-Net, and ResUnet not only fails to maintain the size to length ratio of the collarbone segmentation but also miss the intrinsic details of the segmented organ. The BiSeNet-v2, Linknet-b7, DC U-Net and the Swin U-Net fails to precisely detect the shape of the trachea. The BiSeNet-v2, Linknet-b7 in particular fails to give a proper segmentation mask displaying the left and right bronchi which sub-divides from the trachea. UResNet and Attention U-Net on the other hand generates extra pixels in their trachea segmentation outputs. In the case of lungs and heart, the proposed model not only predicts a proper shape of the organs but also takes into account the finer details. The attention U-Net, BiSeNet-v2, Linknet-b7, ResUnet, Swin Unet model fail to showcase the finer details and predicts extra region.

The ribs are complex skeletal structures that protect the thoracic cavity and respiratory system. They have a complex morphology with varying thicknesses, curvatures, and orientations. Due to their intricate structure, the ribs often overlap with other structures in the chest, such as the lungs and soft tissues, making it difficult to accurately distinguish the ribs from the surrounding structures. The proposed model is capable of segmenting the ribs very precisely which is a very complex task.While ResUnet, Unet++, and UResNet have the ability to accurately identify the exact locations of individual ribs, they still exhibit limitations in their generated segmentation masks as they tend to overlook certain areas of the ribs. Additionally, these models struggle with identifying the precise location of the lateral ribs, leading to ambiguity in their segmentation results. Swin Unet, Linknet-b7, DC U-Net and BiSeNet-v2 model fails to segment ribs properly and give extra region. The qualitative and quantitative analysis with other SOTA segmentation models provided in Fig. 2 and Table IV showcases the robustness of the model. SE-UResNet has less to almost no extra pixels in the predicted segmentation masks. The proposed SE-UResNet is able to maintain an average DSC of atleast 76% for multiorgan segmentation in two different datasets. The low SD of around 6% for small lesions (heart, trachea and collarbone) and 1% for lung

TABLE IV
COMPARATIVE ANALYSIS FOR DIFFERENT MODELS ON NIH CXR [14] AND VINDR-RIBCXR [15] DATASETS WITH RESPECT TO DSC (IN %).

Model	ResUnet [10]	Attention Unet [8]	Swin Unet [13]	Unet++ [7]	Linknet-b7 [18]	DC U-Net [19]	BiSeNet-v2 [20]	UResNet [11]	SE-UResNet (Proposed model)
Heart	74.7 ± 10.4	74.6 ±10.8	77.7 ± 8.2	75.38 ± 7.57	76.5 ± 9.8	73.0 ± 7.5	69.9 ± 12.1	77.7 ± 7.1	**81.86 ± 7.17**
Lungs	94.6 ± 10.8	93.4 ± 6.4	93.7 ± 3.2	91.45 ± 1.18	93.0 ± 4.7	94.7 ± 2.9	91.58 ± 4.3	94.1 ± 1.9	**95.93 ± 1.54**
Trachea	70.2 ± 9.2	75.8 ± 6.8	71.9 ± 7.3	70.15 ± 11.17	69.7 ± 6.8	75.9 ± 6.7	51.2 ± 14.5	71.5 ± 14.6	**76.88 ± 6.58**
Collar Bone	56.2 ± 16.1	72.2 ± 7.4	58.5 ± 12.0	74.63 ± 1.08	65.9 ± 8.9	60.6 ± 18.6	50.62 ± 18.3	77.01 ± 6.6	**78.83 ± 6.38**
Ribs	83.2 ± 4.1	83.91 ± 7.53	75.5 ± 4.1	83.6 ± 4.3	80.5 ± 3.8	84.5 ± 3.4	78.9 ± 4.4	85.7 ± 4.2	**86.04 ± 3.93**

*Individual cell values are in the format of Mean ± SD

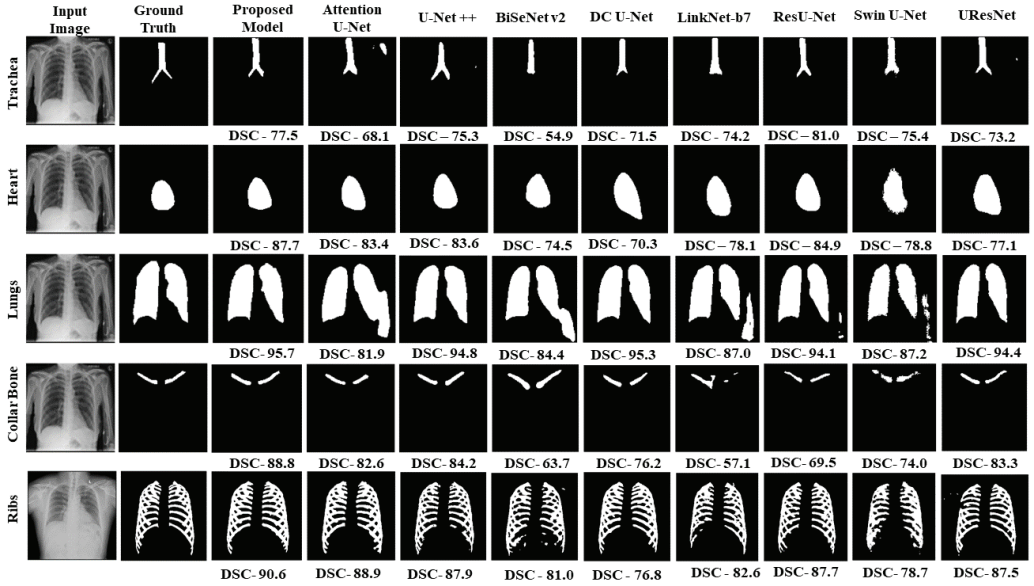

Fig. 2. Qualitative analysis of the SE-UResNet with SOTA segmentation models on different organs: Trachea, Heart, Lungs, Collarbone and Ribs (from top to bottom).

segmentation indicates the consistency of the proposed model.

These results shows that our model perform well on these two datasets. However, to test the reproducibility of our model, we evaluate SE-UResNet on two benchmark datasets. The detailed analysis is discussed in the next section.

IV. REPRODUCIBILITY

DL models have a tendency to excel at performing well on a training dataset but fail when presented with new, previously untrained datasets. This occurs because the models struggle to generalize the patterns they have learned from the training dataset and apply them to new data, leading to suboptimal performance on new datasets. In this section, the performance of the proposed model is evaluated on two established benchmark CXR datasets for lung segmentation, namely Shenzen and JSRT. The lack of ground truth for heart, trachea, collarbone and ribs for these datasets limited our experiment to only lungs segmentation. Using the pre-trained weights obtained from the NIH dataset, the model is tested on the Shenzen and JSRT datasets, with the results presented in Table III. The table reports performance metrics such as Dice Score, mean IoU, Precision, and Recall, where

TABLE V
PERFORMANCE OF THE SE-URESNET ON SHENZEN AND JSRT CXR DATASETS

DATASET	DSC (in%)	IoU (in%)	Precision (in%)	Recall (in%)
Shenzen [16]	94. 39 ± 2.6	90.01 ± 4.4	94.08 ± 2.1	94.84 ± 5.03
	97.55 - 82.17	95.21 – 71.40	99.35 – 85.6	99.79 - 73.26
JSRT [17]	91. 18 ± 5.3	85.59 ± 7.5	96.6 ± 1.3	94.84 ± 8.9
	97.41 - 70.34	94.21 – 61.40	99.30 – 93.3	98.90 - 59.83

*The dark shaded row represents (Mean ± SD) and light shaded row represents (Max - Min)

the Mean ± SD is presented in dark shaded rows, followed by Maximum – Minimum values in the subsequent light shaded rows. The high DSC of 94.39% and 91.18% along with a low SD of 2.6% and 5.3% for Shenzen and JSRT datasets respectively establishes the reproducibility of the model. The residual module in addition with the S&E modules enables the proposed model to learn generalized and robust features. Dropout layers present in every encoder block also improves the reproducibility of the SE-UResNet (as shown in Fig. 3A and 3B) by preventing the model from overfitting.

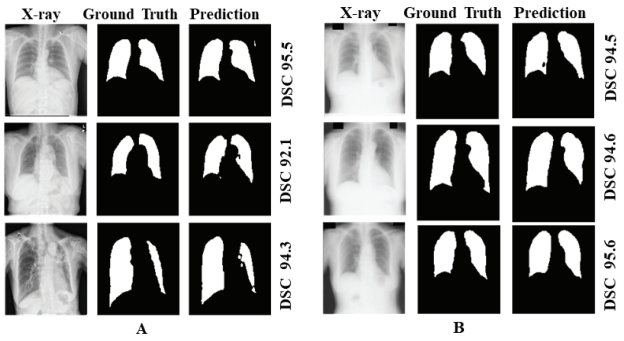

Fig. 3. Visualisation of segmented lung mask generated by the proposed model on A. Shenzen and B. JSRT Dataset.

V. Conclusion

In this paper, we have proposed a novel, robust, and generalisable segmentation network SE-UResNet. The SE-UResNet introduces a residual module which improves the quality of the generated segmentation masks. S&E modules are also added in the decoding and residual modules which allows the model to focus on small scale features. The proposed SE-UResNet solves the issue of automatic segmentation of small lesions from a small training dataset. The performance of SE-UResNet for heart, lungs, trachea, collar bone and ribs in terms of DSC are 81.86%, 95.93%, 76.88%, 78.83% and, 86.04%. The experimental results on NIH and VinDr-RibCXR datasets prove that the SE-UResNet is the best for automatic segmentation of organs of various shapes and sizes from a chest radiograph input. Further experiments on Shenzen and JSRT datasets establish that the model is reproducible and can generate automatic segmentations on untrained chest radiographic datasets. Moreover, the introduction of residual module makes the model flexible. In future, we plan to integrate different convolutional network blocks in place of the residual block to make segmentation models for other imaging modalities such as CT scans and MRI.

Acknowledgement

This research work was supported by the RFIER-Jio Institute "CVMI-Computer Vision in Medical Imaging" project under the "AI for ALL" research center.

Source code availability

The source code is available at: https://github.com/Dynamo13/SE-UResNet.

References

[1] S. Roy, T. Meena, and S.-J. Lim, "Demystifying supervised learning in healthcare 4.0: A new reality of transforming diagnostic medicine," *Diagnostics*, vol. 12, no. 10, p. 2549, 2022.

[2] T. Meena and S. Roy, "Bone fracture detection using deep supervised learning from radiological images: A paradigm shift," *Diagnostics*, vol. 12, no. 10, 2022.

[3] P. Tarighi, S. Eftekhari, M. Chizari, M. Sabernavaei, D. Jafari, and P. Mirzabeigi, "A review of potential suggested drugs for coronavirus disease (covid-19) treatment," *European Journal of Pharmacology*, vol. 895, p. 173890, 2021.

[4] S. Ayyachamy, V. Alex, M. Khened, and G. Krishnamurthi, "Medical image retrieval using resnet-18," in *Medical imaging 2019: imaging informatics for healthcare, research, and applications*, vol. 10954. SPIE, 2019, pp. 233–241.

[5] Y. Wen, L. Zhang, X. Meng, and X. Ye, "Rethinking the transfer learning for fcn based polyp segmentation in colonoscopy," *IEEE Access*, vol. 11, pp. 16 183–16 193, 2023.

[6] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

[7] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*. Springer, 2018, pp. 3–11.

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

[9] D. Pal, P. B. Reddy, and S. Roy, "Attention uw-net: A fully connected model for automatic segmentation and annotation of chest x-ray," *Computers in Biology and Medicine*, vol. 150, p. 106083, 2022.

[10] D. Li, D. A. Dharmawan, B. P. Ng, and S. Rahardja, "Residual u-net for retinal vessel segmentation," in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 1425–1429.

[11] T. Estienne, M. Vakalopoulou, S. Christodoulidis, E. Battistela, M. Lerousseau, A. Carre, G. Klausner, R. Sun, C. Robert, S. Mougiakakou *et al.*, "U-resnet: Ultimate coupling of registration and segmentation with deep nets," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part III 22*. Springer, 2019, pp. 310–319.

[12] A. Abedalla, M. Abdullah, M. Al-Ayyoub, and E. Benkhelifa, "Chest x-ray pneumothorax segmentation using u-net with efficientnet and resnet architectures," *PeerJ Computer Science*, vol. 7, p. e607, 2021.

[13] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-unet: Unet-like pure transformer for medical image segmentation," in *Computer Vision–ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part III*. Springer, 2023, pp. 205–218.

[14] D. Chicco and G. Jurman, "The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation," *BMC genomics*, vol. 21, pp. 1–13, 2020.

[15] H. C. Nguyen, T. T. Le, H. H. Pham, and H. Q. Nguyen, "Vindr-ribcxr: A benchmark dataset for automatic segmentation and labeling of individual ribs on chest x-rays," *arXiv preprint arXiv:2107.01327*, 2021.

[16] S. Candemir, S. Jaeger, K. Palaniappan, J. P. Musco, R. K. Singh, Z. Xue, A. Karargyris, S. Antani, G. Thoma, and C. J. McDonald, "Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration," *IEEE transactions on medical imaging*, vol. 33, no. 2, pp. 577–590, 2013.

[17] J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K.-i. Komatsu, M. Matsui, H. Fujita, Y. Kodera, and K. Doi, "Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules," *American Journal of Roentgenology*, vol. 174, no. 1, pp. 71–74, 2000.

[18] C. Akyel and N. Arıcı, "Linknet-b7: Noise removal and lesion segmentation in images of skin cancer," *Mathematics*, vol. 10, no. 5, p. 736, 2022.

[19] A. Lou, S. Guan, and M. Loew, "Dc-unet: rethinking the u-net architecture with dual channel efficient cnn for medical image segmentation," in *Medical Imaging 2021: Image Processing*, vol. 11596. SPIE, 2021, pp. 758–768.

[20] C. Yu, C. Gao, J. Wang, G. Yu, C. Shen, and N. Sang, "Bisenet v2: Bilateral network with guided aggregation for real-time semantic segmentation," *International Journal of Computer Vision*, vol. 129, pp. 3051–3068, 2021.