



ConnectedUNets++: Mass Segmentation from Whole Mammographic Images

Prithul Sarker^(✉), Sushmita Sarker, George Bebis, and Alireza Tavakkoli

Department of Computer Science and Engineering, University of Nevada, Reno, USA
{prithulsarker,sushmita}@nevada.unr.edu

Abstract. Deep learning has made a breakthrough in medical image segmentation in recent years due to its ability to extract high-level features without the need for prior knowledge. In this context, UNet is one of the most advanced medical image segmentation models, with promising results in mammography. Despite its excellent overall performance in segmenting multimodal medical images, the traditional U-Net structure appears to be inadequate in various ways. There are certain U-Net design modifications, such as MultiResUNet, Connected-UNets and AU-Net, that have improved overall performance in areas where the conventional U-Net architecture appears to be deficient. Following the success of UNet and its variants, we have presented two enhanced versions of the Connected-UNets architecture: ConnectedUNets+ and ConnectedUNets++. In ConnectedUNets+, we have replaced the simple skip connections of Connected-UNets architecture with residual skip connections, while in ConnectedUNets++, we have modified the encoder decoder structure along with employing residual skip connections. We have evaluated our proposed architectures on two publicly available datasets, the Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM) and INbreast.

Keywords: Convolutional Neural Network · Mammogram · Semantic segmentation · U-Net · ConnectedU-Nets · MultiResUNet

1 Introduction

Breast cancer is the most frequent type of cancer that causes death in women, with 44,130 instances reported in the United States in 2021 [1]. The need for frequent mammography screening has been stressed in many studies in order to reduce mortality rates by finding breast malignancies before they spread to other normal tissues and healthy organs. A mammogram is an X-ray image of the breast to record changes in the tissue. The disease is typically identified by the presence of abnormal masses and microcalcifications in mammograms [2, 3]. Radiologists examine a high number of mammograms on a daily basis looking for abnormal lesions and assessing the location, shape, and type of any suspicious

Prithul Sarker and Sushmita Sarker have equal contribution and are co-first authors.

area in the breast. This is an important procedure which requires high precision and accuracy, however, it is still costly and prone to errors since detecting these regions is challenging as their pixel intensities often coincide with normal tissue.

Deep learning advances [4], especially Convolutional Neural Networks (CNN) [5], have shown a lot of promise in addressing these issues. Despite being a game-changer in computer vision, CNN architectures have a key drawback: they require an enormous amount of training data. In order to solve this problem, U-Net [6] is introduced which is built on a simple encoder-decoder network with multiple sets of CNN. Even with a limited quantity of labeled training data, U-Net has demonstrated tremendous promise in segmenting breast masses, to the point where it has become the de-facto standard in medical image segmentation [7]. In light of the success of U-Net, various U-Net versions, such as Connected-UNets [8] and AU-Net [9], have been proposed. These variations have demonstrated promising results but appear to be inefficient in terms of fully recovering the region of interest in a given image.

In this work, we have proposed and experimented with two enhanced versions of the Connected-UNets architecture. Although the proposed networks share an architectural similarity, they are designed for different use cases which are crucial in real-world scenario. The proposed architectures take the entire mammogram image as input and perform mass segmentation along with mass boundary extraction. The main contributions of our work include:

1. We have proposed ConnectedUNets+ and ConnectedUNets++, two novel and improved versions of the Connected-UNets, by utilizing residual skip connections and enhanced encoder-decoder in order to achieve better convergence.
2. We have assessed the proposed architectures using full mammogram images in contrast to the baseline model which operates on cropped images of correctly detected and classified masses by an object detection model.
3. We have experimented using all the images from two publicly available datasets, the Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM) [10] and INbreast [11] for segmenting the region of interest (ROI) of breast mass tumors.

To the best of our knowledge, our paper is the first to address the shortcomings of other papers' methodologies and to conduct an unbiased comparison. We applied the same loss function, optimizer, and image size to all architectures to maintain objectivity. Additionally, to ensure a fair and accurate comparison, we used full mammograms as input for all the models and adopted a comparable preprocessing approach.

2 Related Works

U-Net [6], a deep learning network having an encoder-decoder architecture, is among the most prominent deep neural networks commonly employed in medical image segmentation. The network has a symmetric architecture, with an encoder

which extracts spatial information from the image and a decoder which constructs the segmentation map from the encoded data. The encoder and decoder are linked by a series of skip connections which are the most innovative component of the U-Net architecture since they enable the network to recover spatial data that has been lost due to pooling procedures. Abdelhafiz et al. [12] used a vanilla U-Net model to segment mass lesions in whole mammograms. To segment suspicious regions in mammograms, Ravitha Rajalakshmi et al. [13] presented a deeply supervised U-Net model (DS U-Net) combined with a dense Conditional Random Field (CRF). Li et al. [14] proposed a Conditional Residual U-Net, named CRUNet, to improve the performance of the basic U-Net for breast mass segmentation.

Though U-Net is among the most popular and successful deep learning models for biomedical image segmentation, several improvements are still possible. Specifically, the concatenation of encoder and decoder features reveals a significant semantic gap despite the preservation of dispersed spatial features, which is a shortcoming of the simple skip connections. To deal with this issue, Ibtihaz et al. [15] proposed the MultiResUNet architecture by incorporating some convolutional layers along with shortcut connections in U-Net. Instead of simply concatenating the feature maps from the encoder stage to the decoder stage, they first pass them through a chain of convolutional layers and then concatenate them with the decoder features, which makes learning substantially easier. This idea is inspired from the image-to-image conversion using convolutional neural networks [16], where pooling layers are not favorable for the loss of information. MultiResUNet has shown excellent results on different biomedical images, however, the authors did not experiment with mammograms.

Based on the U-Net architecture, Baccouche et al. [8] proposed an improved architecture that connects two simple U-Nets, called Connected-UNets. In addition to the original idea of the U-Net architecture, which includes skip connections between the encoder and decoder networks, it cascades a second U-Net and adds skip connections between the decoder of the first U-Net and the encoder of the second U-Net. The key idea was to recovering fine-grained characteristics lost in U-Net's encoding process. However, the authors first used YOLO [17] to detect the location of masses in mammograms, and then applied their method to segment only correctly localized masses. Such an approach is not optimum in practical settings where it is desirable to simultaneously localize and segment masses in whole mammograms rather than processing cropped mammograms.

Several modifications of the U-Net architecture have also been proposed by incorporating an attention mechanism, which has shown to be extremely effective in medical image segmentation. Oktay et al. [18] proposed a new attention U-Net by adding an attention gate into the conventional U-Net. This enhanced the accuracy of the predictions. However, they didn't evaluate their model for breast mass segmentation. Similarly, Li et al. [19] built an attention dense U-Net for breast mass segmentation, which was compared to U-Net [6], Attention U-Net [18], and DenseNet [20]. In another study by Sun et al. [9], an attention-guided dense upsampling network, called AUNet, was built for breast mass seg-

mentation in full mammograms. The major drawback of the papers mentioned above is they did not use all of the images available in the CBIS-DDSM dataset for experimentation (i.e., they only used a portion of the images in the training and test sets). As a result, higher scores were reported in their studies.

In this paper, we propose ConnectedUNets+ and ConnectedUNets++, two enhanced versions of the Connected-UNets and Connected-ResUNets architectures by focusing on the limitations of the aforementioned architectures. An important contribution of our work is that we compared the proposed architectures with previously reported works under identical conditions. Moreover, we did not employ any object detection models for mass localization; instead, we conducted all of our experiments using whole mammograms.

3 Methodology

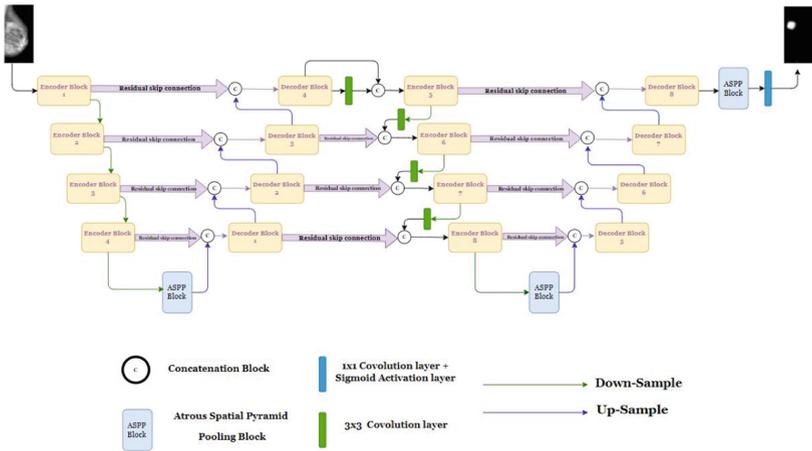
3.1 Architecture

We have used Connected-UNets [8] as our baseline model because of its architectural elegance and performance. Even though, at first glance, ConnectedUNets+ and ConnectedUNets++ may merely seem a logical extension of Connected-UNets, the introduction of residual skip connections between the encoder and decoder is essential for successful segmentation. This not only improves the metric scores but also bridges the semantic barrier between the encoder-decoder features. The most crucial distinction is that ConnectedUNets+ and ConnectedUNets++ have been designed to enable mass segmentation from full mammograms rather than cropped mammograms. Figure 1a shows an overview of our proposed ConnectedUNets++ architecture. Please take note that we have not included a separate illustration for the ConnectedUNets+ since both the architectures are identical with the exception of the encoder-decoder block, which has been maintained standard like the baseline model. For both the models, we have replaced the simple skip connections of Connected-UNets with more optimal residual skip connections between encoder and decoder and between the UNets as well. However, for the ConnectedUNets++ architecture, we have also modified the encoder-decoder block by including three 3×3 convolutions and one residual connection followed by an activation layer ReLU (Rectified Linear Unit) and a batch normalization (BN) layer as shown in Fig. 1b. Specifically, the residual skip connections consist of four 3×3 convolutions where each of them is accompanied by one 1×1 convolution. The architecture of the residual skip connection is shown in Fig. 1c. The number of convolution blocks decreases in the deeper layer of the network as the semantic gap between encoder and decoder decreases due to getting closer to the bottleneck. Table 1 shows the number of filters used in each of the residual skip connections for different layers.

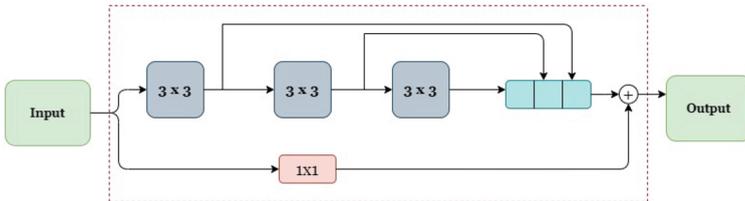
As mentioned by Ibtehaz et al. [15] by adding these residual path connections, the proposed architectures are more immune to perturbations, and outliers. They also help to obtain better results in less time and fewer epochs. Additionally, we have used Atrous Spatial Pyramid Pooling (ASPP) blocks to preserve the same

bottleneck structure for both architectures as Connected-UNets. The architectural details of the encoder, decoder and ASPP blocks are described in Table 2.

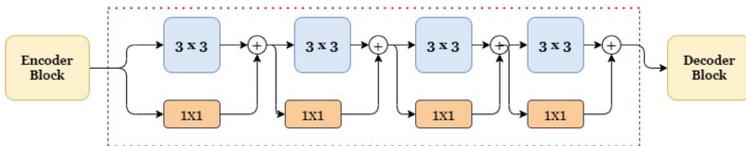
On the encoder side, each encoder block's output is subjected to a maximum pooling operation before the features are forwarded to the next encoder, and the output of the last encoder passes through an ASPP block before being fetched to the first decoder. Each decoder block is made up of a 2×2 transposed convolution unit that up-samples the preceding block's features before concatenating them with the encoder features received by the residual skip connection; and these features are then fetched to the decoder above. A second U-Net is connected via



(a) Proposed architecture of ConnectedUNets++



(b) Architecture of encoder and decoder blocks



(c) Architecture of residual skip connections

Fig. 1. Detailed ConnectedUNets++ architecture

Table 1. ConnectedUNets++ and ConnectedUNets+ architecture details in terms of residual skip connections. These specifics apply to both architectures.

Residual skip connection	No. of conv. layer	Conv. layer kernel size	No of filters
Residual skip connection 01/08	4	3×3	32
	4	1×1	32
Residual skip connection 02/05/09	3	3×3	64
	3	1×1	64
Residual skip connection 03/06/10	2	3×3	128
	2	1×1	128
Residual skip connection 04/07/11	1	3×3	256
	1	1×1	256

Table 2. ConnectedUNets++ architecture details in terms of encoder-decoder and ASPP block. Dilation rate of convolution layers in ASPP block is shown in corresponding braces. ASPP block details also apply to ConnectedUNets+.

Block	Layer	Filters	Block	Layer	Filters
Encoder (01/05) Decoder (04/08)	3×3 Conv.	8	Encoder (02/06) Decoder (03/07)	3×3 Conv.	17
	3×3 Conv.	17		3×3 Conv.	35
	3×3 Conv.	26		3×3 Conv.	53
	1×1 Conv.	51		1×1 Conv.	105
Encoder (03/07) Decoder (02/06)	3×3 Conv.	35	Encoder (04/08) Decoder (01/05)	3×3 Conv.	71
	3×3 Conv.	71		3×3 Conv.	142
	3×3 Conv.	106		3×3 Conv.	213
	1×1 Conv.	212		1×1 Conv.	426
ASPP Block (Bottleneck)	3×3 Conv.	512	ASPP Block (Output Layer)	3×3 Conv.	32
	3×3 Conv.(6)	512		3×3 Conv.(6)	32
	3×3 Conv.(8)	512		3×3 Conv.(8)	32
	3×3 Conv.(12)	512		3×3 Conv.(12)	32
	1×1 Conv.	512		1×1 Conv.	32

a new set of residual skip connections, which are utilized to transfer information from the previous U-Net. The output of the final decoder block of the first U-Net is fed into a 3×3 convolution layer before it gets concatenated with itself again, followed by an activation ReLU and a BN layer. This acts as the first encoder block's input to the second U-Net. The output of the max pooling operations of each of the three encoder blocks is fed into a 3×3 convolution layer and then concatenated with the output of the preceding decoder block of the first U-Net.

The ASPP block receives the output features of the second U-Net’s last encoder block; the remaining blocks are the same as discussed in the first U-Net. Finally, the predicted segmentation mask is generated by passing the output of the last decoder to another ASPP block followed by a 1×1 convolution layer and a sigmoid activation layer. In contrast to the work presented in the Connected-UNets paper [8], our work considers a full-fledged mammographic image as input instead of only a ROI since it typically fails to detect micro masses present in the image. The ROI extracted segmentation does not help in the real life scenario because the detection and localization of the mask has to be done using a different neural network or manually.

3.2 Dataset Preprocessing and Experimental Setup

We evaluated the proposed architecture on two publicly available datasets, CBIS-DDSM [10] and INbreast [11]. CBIS-DDSM contains 2478 mammography images from 1249 women and included both craniocaudal (CC) and mediolateral oblique (MLO) views for most of the exams of which 1231 cases contain single or multiple breast masses. This dataset includes real-world mammograms with background artifacts, poor contrast, corners, borders, and different orientations. To address these issues, we have applied several preprocessing steps which include border cropping to tackle the white border or corner problem, normalizing pixel values to the interval 0 to 1, eliminating background artifacts, and finally, applying CLAHE for contrast enhancement. This improves the mammogram’s fine details, textures, and features that would otherwise be challenging for the model to learn. To preprocess the ground truth masks, we discarded the same amount of the borders to get rid of any artifacts and used appropriate padding to make the masks square. In the case of the CBIS-DDSM dataset, after preprocessing and fusing multiple masks of the same image, we split the 1231 images in the training set using an 85:15 ratio for training and validation (i.e., 1046 and 185 images, respectively). The test dataset had 359 images.

INbreast dataset was built with full-field digital mammograms and has a total of 115 cases which include both masses and calcifications. In total, the dataset has only 107 images of breasts with masses. We have used 69 images for training, 17 images for validation, and 21 images for testing. For preprocessing, we have solely applied CLAHE both on the mammograms and the ROIs.

In all of the experiments, adam [21] optimizer is used with an initial learning rate of 0.0001. Batch size of 16 is used during training and testing. We have experimented with input size of 224×224 and 256×256 . The best score for all the architectures was obtained with the 224×224 input size, which is reported in the results section. Most of the papers mentioned in the related work section use a mixture of Dice and IOU loss. The primary motivation for directly using this loss is to maximize those metrics. However, this gives no information about convergence. So, to remove confusion regarding the convergence, the loss function used here is binary crossentropy which is the standard loss function for image segmentation task.

3.3 Evaluation Metrics

In semantic segmentation, the region of interest typically occupies a small area of the entire image. Therefore, metrics like precision and recall are inadequate and often lead to a false sense of superiority, inflated by the perfection of detecting the background. To evaluate our approach, we have considered four metrics in our experiments: Dice score (F1 score), Jaccard Index (IoU Score), accuracy and Hausdorff distance (H). Even though the Dice score (Eq. 1) and Jaccard index (Eq. 2) are two widely used metrics for semantic segmentation, they are biased towards large masses. The Hausdorff distance (Eq. 3) is an unbiased metric that treats all objects equally independently of their size. It measures the maximum deviation along the boundary between the ground truth and predictions.

$$Dice\ score(A, B) = \frac{2 \times Area\ of\ Intersection(A, B)}{Area\ of(A) + Area\ of(B)} = \frac{2 \times (A \cap B)}{A + B} \quad (1)$$

$$IoU\ score(A, B) = \frac{Area\ of\ Intersection(A, B)}{Area\ of\ Union(A, B)} = \frac{A \cap B}{A \cup B} \quad (2)$$

$$H = max(h(GT, pred), h(pred, GT)) \quad (3)$$

4 Experimental Results and Discussion

To assess the performance, all models have been run for 300 and 400 epochs for the CBIS-DDSM and INbreast datasets, respectively with early stopping and ReduceLROnPlateau callback function. Table 3 shows the comparison of the proposed architectures' results with some state-of-the-art methods on the CBIS-DDSM dataset. As it can be observed, ConnectedUNets++ consistently outperforms the baseline model Connected-UNets and Connected-ResUNets along with other models used for mass segmentation. ConnctedUNets+ performed better than Connected-UNets on the test dataset even though the only difference between them is the residual skip connections.

As seen in Table 3, the number of parameters of our proposed methods is higher than the number of parameters of the baseline architectures since we used residual skip connections between encoders and decoders as well as between the two U-Nets. Furthermore, all of the architectures performed better on the training and validation sets, but performed poorly on the test set. In particular, the architectures fail to detect any mass from some of the images in the test set, thus giving an output of no ROI and reducing the metric values. We speculate that this is due to the inferior and substandard quality of the scanned images. In Table 4, we compared our proposed architecture with the baseline model by using individual cases from the CBIS-DDSM test dataset considering various thresholds for different metrics. As shown, the proposed architecture was able to predict more cases for each threshold with respect to Connected-ResUNets; however, the average score was better for the baseline architecture. Here we argue that ConnectedUNets++ has been able to correctly segment smaller masses more accurately than Connected-ResUNets.

Table 3. Comparison of the proposed architectures and state-of-the-art methods on the CBIS-DDSM dataset. Here, DS: Dice score, JI: Jaccard index, Acc.: Accuracy, Param.: No of parameters(in million).

Model name	Param.	Training			Validation			Test		
		DS	JI	Acc.	DS	JI	Acc.	DS	JI	Acc.
U-Net	7.8	0.73	0.57	99.85	0.73	0.58	99.87	0.41	0.27	99.69
MultiResUNet	7.3	0.74	0.59	99.89	0.76	0.61	99.88	0.40	0.26	99.7
AUNet	11.01	0.89	0.81	99.94	0.90	0.82	99.94	0.46	0.31	99.69
Connected-UNets	20.1	0.81	0.68	99.90	0.81	0.68	99.91	0.40	0.27	99.69
ConnectedUNets+ (ours)	23.5	0.78	0.64	99.88	0.77	0.63	99.89	0.44	0.30	99.7
Connected-ResUNets	20.7	0.84	0.73	99.91	0.84	0.72	99.92	0.47	0.32	99.69
ConnectedUNets++ (ours)	28.15	0.88	0.79	99.94	0.88	0.79	99.94	0.48	0.33	99.7

Table 4. Comparison of prediction on the CBIS-DDSM test dataset and correctly predicted number of cases over multiple thresholds of Dice score (DS), Jaccard index (JI) and Hausdorff distance (HD) metric between the baseline architecture and our proposed architecture. Here, NC and AS represent the number of cases and average score, respectively.

Architecture name	DS \geq 0.45		DS \geq 0.65		JI \geq 0.35		JI \geq 0.55		HD \leq 2.75	
	NC	AS								
Connected-ResUNets	160	0.78	134	0.82	154	0.66	119	0.73	174	1.85
ConnectedUNets++ (ours)	181	0.76	143	0.81	172	0.64	123	0.71	191	1.92

Table 5. Comparison of the proposed architectures and state-of-the-art methods on INbreast dataset.

Model name	Param.	Training			Validation			Test		
		DS	JI	Acc.	DS	JI	Acc.	DS	JI	Acc.
U-Net	7.8	0.87	0.77	99.91	0.91	0.83	99.91	0.91	0.84	99.91
AUNet	11.01	0.94	0.89	99.94	0.96	0.90	99.93	0.94	0.89	99.94
Connected-UNets	22.4	0.94	0.89	99.99	0.97	0.95	99.95	0.97	0.94	99.99
ConnectedUNets+ (ours)	23.5	0.97	0.94	99.97	0.98	0.96	99.97	0.98	0.95	99.97
Connected-ResUNets	20.7	0.94	0.88	99.99	0.96	0.93	99.98	0.97	0.94	99.99
ConnectedUNets++ (ours)	28.15	0.97	0.95	99.99	0.98	0.97	99.98	0.99	0.97	99.99

In Table 5, we also compared the proposed architectures against the baseline and other models on the INbreast dataset. As it can be observed, our model performs noticeably better on the INbreast dataset. We hypothesize that this is due to the two datasets' disparate image quality. The mammograms in the CBIS-DDSM dataset have been scanned, hence the images are of poor quality. The images in the INbreast dataset, however, have been digitally enhanced, and their quality is outstanding. Due to page restrictions, we had to omit comparison of correctly predicted cases like Table 4 for the INbreast dataset.

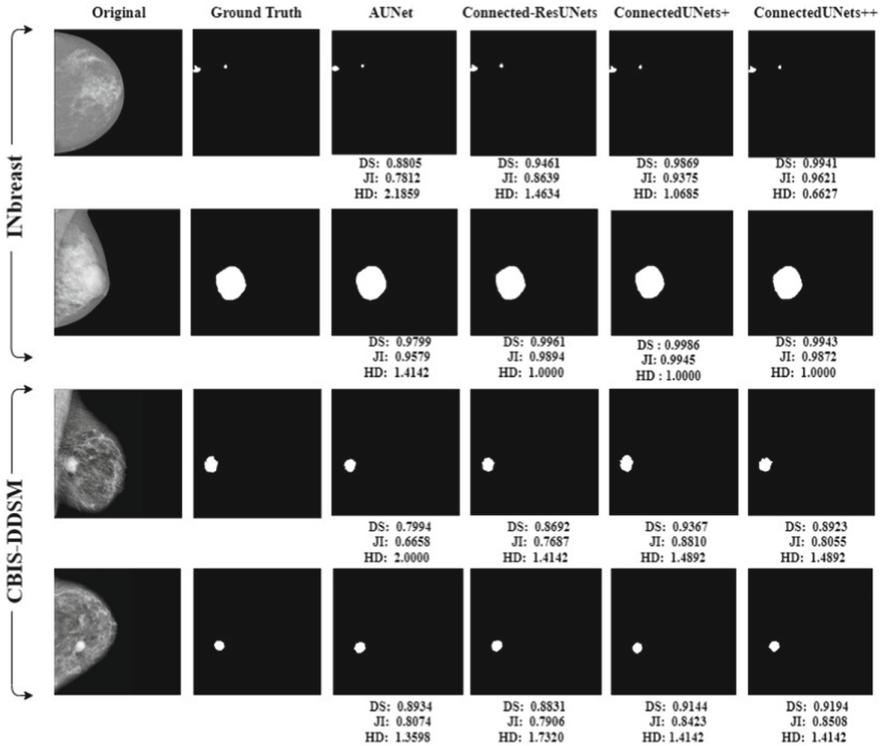


Fig. 2. Segmentation results of different networks. From left to right, the columns correspond to the input images, the ground truth labels, the segmentation results of AUNet, Connected-ResUNets, ConnectedUNets+ and ConnectedUNets++, respectively on the INbreast and CBIS-DDSM dataset.

The introduction of residual skip connections between the encoder and decoder has had a major impact on the segmentation task. The obscure and vague boundaries which other architectures fail to correctly segment (either under-segment or over-segment), are properly segmented by ConnectedUNets+ and ConnectedUNets++. Additionally, in some complex cases, because of the quality and nature of the ROIs, it becomes challenging to segment homogeneous ROIs. Even in those cases, ConnectedUNets++ exceeds other architectures in terms of Dice score, Jaccard index, and Hausdorff distance. As Hausdorff distance is highly recommended for cases with complex boundaries, our results show that the proposed architecture can predict mass boundaries more accurately. Since the boundary shape of a mass is a strong indicator of benign and malignant cases [22], the proposed architecture is more suited for mass prediction and segmentation in real-world scenarios.

Segmentation examples both for the INbreast and CBIS-DDSM datasets are shown in Fig. 2. We have compared the segmentation result of ConnectedUNets+ and ConnectedUNets++ with two best-performed architectures: AUNet

and Connected-ResUNets. In all the cases, the proposed architectures outperformed all other architectures and achieved almost perfect scoring regardless of the ROI's shape. Figure 2 showcases how well the models worked not just for the larger ROI (Case 2) but also for those with smaller ROI (Case 1) on INbreast dataset. The models yield similar results for CBIS-DDSM dataset as well.

5 Conclusion

We started this work by thoroughly examining the Connected-UNets architecture in order to identify potential areas for improvement. In this context, we identified certain inconsistencies between the encoder's and decoder's features. Inspired by MultiResUNet, we added some additional processing between them to make them more homogeneous. Furthermore, to give Connected-UNets the capacity to perform multi-resolutional analysis, we introduced residual blocks into the encoder and decoder architecture which resulted in ConnectedUNets++, a novel architecture that incorporates these changes. Unlike architectures previously discussed in the literature, we have used all of the available images for training and testing. For future work, we plan to conduct additional experiments to determine the ideal choice of nodes, layers, and hyperparameters. Furthermore, we would also like to assess the efficacy of our model on datasets of medical images from various modalities.

Acknowledgements. Portions of this material is based upon work supported by the Office of the Under Secretary of Defense for Research and Engineering under award number FA9550-21-1-0207.

References

1. American Chemical Society: Breast cancer facts & figures 2019–2020. *Am. Cancer Soc.* 1–44 (2019)
2. Elter, M., Horsch, A.: CADx of mammographic masses and clustered microcalcifications: a review. *Med. Phys.* **36**(6Part1), 2052–2068 (2009)
3. Jiang, Y., Nishikawa, R.M., Schmidt, R.A., Metz, C.E., Giger, M.L., Doi, K.: Improving breast cancer diagnosis with computer-aided diagnosis. *Acad. Radiol.* **6**(1), 22–33 (1999)
4. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**(7553), 436–444 (2015)
5. Zaheer, R., Shaziya, H.: GPU-based empirical evaluation of activation functions in convolutional neural networks. In: 2018 2nd International Conference on Inventive Systems and Control (ICISC), pp. 769–773. IEEE (2018)
6. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
7. Litjens, G., et al.: A survey on deep learning in medical image analysis. *Med. Image Anal.* **42**, 60–88 (2017)

8. Baccouche, A., Garcia-Zapirain, B., Castillo Olea, C., Elmaghraby, A.S.: Connected-UNets: a deep learning architecture for breast mass segmentation. *NPJ Breast Cancer* **7**(1), 1–12 (2021)
9. Sun, H., et al.: AUNet: attention-guided dense-upsampling networks for breast mass segmentation in whole mammograms. *Phys. Med. Biol.* **65**(5), 055005 (2020)
10. Lee, R.S., Gimenez, F., Hoogi, A., Miyake, K.K., Gorovoy, M., Rubin, D.L.: A curated mammography data set for use in computer-aided detection and diagnosis research. *Sci. Data* **4**(1), 1–9 (2017)
11. Moreira, I.C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M.J., Cardoso, J.S.: INbreast: toward a full-field digital mammographic database. *Acad. Radiol.* **19**(2), 236–248 (2012)
12. Abdelhafiz, D., Bi, J., Ammar, R., Yang, C., Nabavi, S.: Convolutional neural network for automated mass segmentation in mammography. *BMC Bioinform.* **21**(1), 1–19 (2020)
13. Ravitha Rajalakshmi, N., Vidhyapriya, R., Elango, N., Ramesh, N.: Deeply supervised U-Net for mass segmentation in digital mammograms. *Int. J. Imaging Syst. Technol.* **31**(1), 59–71 (2021)
14. Li, H., Chen, D., Nailon, W.H., Davies, M.E., Laurenson, D.: Improved breast mass segmentation in mammograms with conditional residual U-Net. In: Stoyanov, D., et al. (eds.) *RAMBO/BIA/TIA 2018*. LNCS, vol. 11040, pp. 81–89. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00946-5_9
15. Ibtehaz, N., Rahman, M.S.: MultiResUNet: rethinking the U-Net architecture for multimodal biomedical image segmentation. *Neural Netw.* **121**, 74–87 (2020)
16. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778 (2016)
17. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779–788 (2016)
18. Oktay, O., et al.: Attention U-Net: learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999* (2018)
19. Li, S., Dong, M., Du, G., Mu, X.: Attention Dense-U-Net for automatic breast mass segmentation in digital mammogram. *IEEE Access* **7**, 59037–59047 (2019)
20. Hai, J., et al.: Fully convolutional DenseNet with multiscale context for automated breast tumor segmentation. *J. Healthcare Eng.* **2019** (2019)
21. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014)
22. Mahmood, T., Li, J., Pei, Y., Akhtar, F., Rehman, M.U., Wasti, S.H.: Breast lesions classifications of mammographic images using a deep convolutional neural network-based approach. *PLoS ONE* **17**(1), e0263126 (2022)