Mammogram Tumor Segmentation With Preserved Local Resolution: An Explainable AI System

Aya Farrag*[1], Gad Gad*[2], Zubair Md Fadlullah*†[3], and Mostafa M. Fouda‡[4]
*Department of Computer Science, Lakehead University, Thunder Bay, ON, Canada.
†Department of Computer Science, Western University, London, ON, Canada.
‡Department of Electrical and Computer Engineering, Idaho State University, Pocatello, ID, USA.
Emails: [1]afarrag@lakeheadu.ca, [2]ggad@lakeheadu.ca, [3]zubair.fadlullah@lakeheadu.ca, [4]mfouda@ieee.org.

GLOBECOM 2023 - 2023 IEEE Global Communications Conference | 979-8-3503-1090-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/GLOBECOM54140.2023.10436915

***Abstract*—Medical image segmentation is a crucial component of computer-aided diagnosis (CAD) systems, as it aids in identifying important areas in medical images. In order to achieve optimal segmentation results, it is important to preserve the resolution of the input image. The dilated convolution module was introduced to maintain resolution across layers of a deep convolutional neural network by increasing the receptive field exponentially while keeping the parameters increase linearly. However, one drawback of using dilated convolution is that it can result in local spatial resolution loss by increasing the sparsity of the kernel in checkboard patterns. This work proposes a double-dilated convolution module to maintain local spatial resolution in medical image segmentation tasks while having a large receptive field. The module is applied to tumor segmentation in breast cancer mammograms using the state-of-art Deeplabv3+ network. The study also evaluates the developed models with the Gradient weighted Class Activation Map (Grad-CAM) and compares the performance of lesion segmentation networks on mammogram screenings from the INBreast dataset before and after using the proposed dilation module. The results show that the proposed module effectively improves the segmentation performance.**

***Index Terms*—Medical image analysis, tumor segmentation, mammograms, CAD, deep learning, CNN, explainable AI, spatial resolution**

I. Introduction

In recent years, researchers have been increasingly interested in developing Computer-Aided Diagnosis (CAD) systems in a wide range of diagnostic settings using data-driven models, with a heavy focus on Convolutional Neural Networks (CNNs) [1]. These systems are used to help radiologists make more accurate and timely diagnoses by integrating their knowledge with computer output from medical imaging modalities such as X-rays, CT scans, and MRI scans [2]–[4]. Medical image segmentation is a critical task performed in most CAD systems, where the goal is to partition a medical image into multiple regions or segments that correspond to specific anatomical or pathological structures [5]–[8]. Tumor segmentation is a specific application of medical image segmentation used in cancer CAD systems, which involves identifying and delineating the boundaries of a tumor within medical images [9]. This information can be used to quantify the characteristics of a tumor, track changes over time, and evaluate the effectiveness of treatment.

Dense prediction problems require multi-scale contextual reasoning combined with full-resolution output, which is not maintained in a typical CNN architecture with successive pooling layers [10], [11]. Therefore, the authors in [12] introduced dilated convolution module to be used in dense prediction models, which partially addresses the issue of reduced resolution in these models. This was achieved by exponentially increasing the kernel's receptive field, which helped to reduce the need for pooling layers. However, the sparsity of the kernel grows exponentially along with the receptive field, leading to a significant reduction in local spatial resolution [13]. Due to the complexity of medical screening, medical image segmentation requires the preservation of local spatial resolution while processing the images. This paper proposes a modification to the dilated convolution module to address this issue. The proposed module introduces a multi-scale dilation parameter that enables more control over the kernel's resolution while still allowing for an exponentially growing receptive field. Fig. 1 illustrates the various shapes of kernels that can be achieved using our proposed module, including the double-resolution dilation. This demonstrates the module's ability to handle intricate and novel kernel shapes, which helps to increase the receptive field while avoiding the sparsity of kernels at the core of the image with large dilation factors [13].

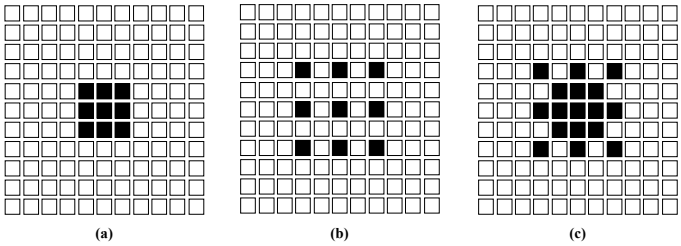

Fig. 1: Our proposed dilation supports receptive field exponential expansion while retaining full resolution at the core of the kernel using multi-dilated convolution. (a) shows a 1-dilated 3x3 kernel. (b) shows a 2-dilated 3x3 kernel. (c) shows a combination of a 1-dilated 3x3 inner kernel and a 2-dilated 3x3 outer kernel. Both (a) and (b) kernels can be realized by traditional dilated convolution methods, whereas only our proposed method can realize the shape in (c).

The interpretability of Deep Learning (DL) models is of utmost importance in medical applications as it is vital for medical professionals and radiologists to comprehend the rationale behind the model's decisions and predictions. While

eXplainable Artificial Intelligence (XAI) models have been successfully implemented in disease classification networks to explain the predicted label of black-box DL models, little attention has been paid to enhancing the interpretability of segmentation networks, even though they also adopt black-box architectures [14], [15]. Additionally, performing post-model explanatory analysis can help identify whether the model is learning relevant features or is merely overfitting to the training images. This enables researchers to adjust the model architecture and hyperparameters to achieve more robust performance that can be generalized to real-life data. Hence, this study aims to tackle the problem of segmentation interpretability by quantitatively evaluating the effectiveness of explainable techniques when applied to medical image segmentation. We introduce explainable AI models in our proposed segmentation network to create transparent systems that can be trusted and integrated into medical practice.

In this study, we focus on the mammogram mass segmentation problem. Breast cancer is the most common cancer affecting women worldwide, and it is the second leading cause of cancer-related deaths [16]. Early detection and diagnosis of breast cancer are crucial in reducing its mortality rate, and the integration of AI methods in breast cancer screening systems can assist radiologists in decreasing mammogram reading time and improving accuracy [17]. To evaluate the effectiveness of our proposed model in medical image segmentation, we use the publicly-available INBreast dataset [18]. Our goal is to develop an efficient and explainable tumor segmentation model by proposing a novel dilated convolution module to maintain local spatial density and adopting interpretability methods in image segmentation networks.

II. Related Work

Identifying the regions of interest in medical images is an essential aspect of the diagnostic process, and numerous CAD solutions have been proposed for this task in the literature. In the past, traditional methods were rule-based and included techniques such as thresholding, boundary-based segmentation, region-based segmentation, and template matching [19]. However, with the advent of deep learning-based techniques, particularly Fully Convolutional Neural Networks (FCNN) [20], significant progress has been made in image semantic segmentation, as compared to classical methods that require handcrafted feature extraction [21]–[23]. This advancement has made it feasible to create large-scale trainable models that have the ability to learn the optimal features necessary for segmentation.

The typical structure of a convolutional neural network (CNN) involves a pyramid shape, with pooling layers following convolutional layers to downsample feature maps generated from each convolution process. This structure has been effective for processing digital images in various high-level computer vision applications, such as recognizing faces, objects, and digits. Nevertheless, using this standard CNN architecture for segmentation tasks leads to a notable reduction in image resolution that can negatively impact segmentation performance, particularly in medical applications, as discussed in [24]. To tackle this issue, previous methods have employed deconvolution networks [25] or multiple scaled versions of the input image with attention to combine outputs [10]. However, it has been argued that excessive downsampling layers can be unnecessary for segmentation tasks and hinder the localization of spatial details [26]. Therefore, dilated convolution was introduced as a solution to maintain spatial resolution while preserving a certain level of abstraction [12].

The dilated convolution introduced a sparse kernel that increases exponentially to cover a larger receptive field, thereby reducing the need for pooling and downsampling layers in several semantic segmentation models [12], [13], [27]. However, in traditional dilated convolution framework, the convolutional kernel is padded with zeros at a fixed dilation rate, resulting in a receptive field that only covers an area with a checkerboard pattern [13]. As a result, only non-zero value locations are sampled, and neighboring information is lost. This problem becomes more severe as the dilation rate increases in higher layers, where the convolutional kernel becomes too sparse to cover any local information since the non-zero weights are too far apart. To address this, researchers in [13] proposed a hybrid dilation convolution that stacked standard dilated convolutions with different rates in a serial manner. This technique was also used in many variants of the Deeplab family, such as Deeplabv2 [28], Deeplabv3 [29], and Deeplabv3+ [30], which achieved state-of-the-art results for the PASCAL VOC benchmark. While this approach improved on traditional dilated convolution, it still had limitations, as dilated kernels were restricted to checkerboard patterns. More recently, semi-dilated convolution [31] proposed a modified version of dilated convolution that can better exploit the geometry of rectangular images by supporting exponential growth of the receptive field in only one dimension of the image. In this work, we follow previous attempts and propose a modification to the dilated convolution to preserve local spatial resolution by separating the dilation factor on the core of the kernel from the dilation factor on its edges while performing medical image segmentation tasks.

As for the explainability aspect, while most explainability research focuses on CNN-based classification networks, few studies have considered this important aspect in developing semantic segmentation networks. For instance, in [32], the authors used the SHAP method to generate colored maps highlighting the image areas that contributed to the model decision for a selected pixel or region, providing understandable explanations for oil slick segmentation models. Similarly, in [33], the second-order derivative of neurons activations at the last encoding layer of their segmentation network was used to provide attention maps that visually explain the underlying network for explainable semantic segmentation in autonomous driving systems. In the area of medical image segmentation, we only found one study that provided explanations for segmenting tumors in liver CT images using an activation maximization-based method [34]. To the best of our knowledge, our work is the first to address the explainability issue in mammogram tumor segmentation, and we qualitatively assess the effectiveness

of the adopted XAI techniques by evaluating their entropy and the pixel-flipping graph, as done in [14], [35].

III. Dataset

We have chosen to analyze the INBreast dataset [18] among the available mammogram datasets for several reasons. Unlike other publicly available datasets that use digitized film-screen mammograms, the INBreast dataset is the only Full-field Digital Mammogram (FFDM) dataset, providing high-resolution images free from inconsistencies that may arise in the digitization process. Additionally, this dataset offers radiologist-drawn pixel-level contours surrounding lesions, instead of just circles around ROIs, which is typical of most databases. This can be critical in diagnosis since the shape of a mass is highly indicative of its malignancy [36]. The INBreast dataset also includes images from both the craniocaudal (CC) view and the mediolateral oblique (MLO) view, allowing for the development of generic CAD systems that can extract information from either view. The dataset contains various types of findings, such as normal, calcification, masses, asymmetries, and architectural distortions. The focus of our study is on segmenting masses, which are three-dimensional structures with convex outward borders, according to BI-RADS guidelines. Among the 410 images, 107 images contained one or more mass lesions.

Since all annotations were saved in XML format, we developed a Matlab script that reads the XML files, extracts annotation information for masses, and uses Matlab's stroke() and imfill() functions to draw contours and generate mask images. We saved the mask images in PNG format for visualization and analysis purposes. An example of the generated mask image and corresponding DICOM image is shown in Fig. 2. We used Matlab's ImageDatastore and PixelLabelDatastore classes to load the raw image and mask image files, respectively, and resized them to a fixed input size of 512x512 pixels. We used a 5-fold validation split to train and validate our models in all our experiments.

IV. Methodology

In order to maintain the local resolution in medical image segmentation networks, we propose a modification to the traditional dilated convolution module that is able to perceive kernels with desnse cores and sparse edges. First, the standard dilated convolution operation is defined as:

$$y[i] = \sum_k x[i+k]w[k] \tag{1}$$

where y is the output feature map, x is the input feature map, and w is the kernel.

Dilated convolution generalizes the standard convolution operation to:

$$y[i] = \sum_k x[i+l.k]w[k] \tag{2}$$

where y is the output feature map, x is the input feature map, w is the kernel, and l is a dilation factor. The dilation operation skips a range in the input defined by the dilated factor while performing the convolution process. Our modified

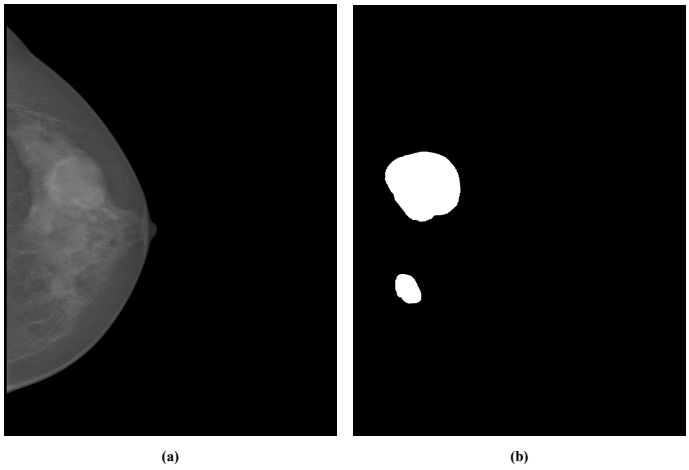

Fig. 2: A mammogram example from the INBreast dataset showing the craniocaudal (CC) view of a left breast image and the corresponding generated mask of existing masses. (a) shows the original image provided in DICOM format. (b) shows the mask generated by extracting masses annotation from the associated XML file using our Matlab script.

dilated convolution can be expressed by the following piecewise function:

$$y[i] = \begin{cases} \sum_k x[i+l_1.k]w[k] & k \leq r_1 \\ \sum_k x[i+l_2.k]w[k] & k > r_1 \end{cases} \tag{3}$$

The traditional dilated convolutional operation employs parameters l and r to determine the dilation factor and the size of the receptive field, respectively, where the receptive field size is given by $(2l+1)^2$. In contrast, our modified version of dilated convolution, which we refer to as *double dilated convolution*, uses two parameters, r_1 and r_2, to define the sizes of two receptive fields, each associated with a distinct kernel. Specifically, the inner (core) kernel has a size of $(2l_1+1)^2$ and a dilation factor of l_1, while the outer (edge) kernel has a size of $(2l_2+1)^2$ and utilizes a dilation factor of l_2. This modification was motivated by the need to address the issue of "gridding" inherent in conventional dilation, as discussed in [13].

In this study, we present a straightforward implementation of our proposed double-dilated kernel that utilizes the most efficient convolution modules available in various deep learning frameworks. The implementation involves applying two distinct kernels with their own dilation rates to the same input and then combining the results of both convolution processes. This is essentially equivalent to applying a single kernel with two different dilation rates to an input. We adopted this approach based on the distributive property of the convolution process, which states that for any three discrete functions $h_1[n]$, $h_2[n]$ and $x[n]$, we can say that:

$$x * h_1 + x * h_2 = x * (h_1 + h_2) \tag{4}$$

Fig. 3 shows how we applied our double-dilated convolution module to the Deeplabv3+ network [30], which was shown

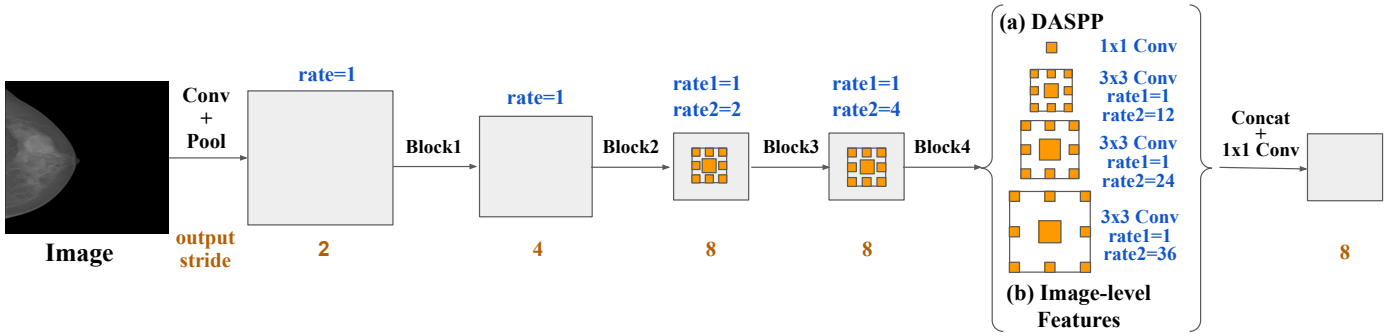

Fig. 3: The modified Deeplabv3+ architecture after plugging the double-dilated convolution module. Each dilated convolution from the original network shown in [29] is replaced with two parallel convolutions: one undialtated convolution (rate1) and one dilated convolution with the same rate used in the original layer (rate2).

to achieve state-of-art results in segmentation tasks [37]. As illustrated in their work [29], both cascaded modules and the spatial pyramid pooling (SPP) module employed in the Deeplab architecture incorporate atrous convolution with different rates to increase the receptive field and include multi-scale context without compromising the image resolution excessively. As a result, the network can produce output maps that are only down-sampled by a factor of 8 instead of the usual 256 in a pyramid-shaped convolution network. We introduced our modification by substituting each dilated convolution layer in the original network with two parallel dilated convolution layers using different dilation rates. One layer maintained a fixed rate of 1 to represent the dense core, while the other preserved the original layer's dilation factor. This modification was applied to both the backbone network and the atrous spatial pyramid pooling (ASPP), which we now call double-dilated ASPP or DASPP. The modified network was trained following a similar protocol as the original training protocol mentioned in [29].

V. Results and Discussion

A. Double-dilated Convolution

To offer a comprehensive evaluation, we assessed the performance of the tumor segmentation task at two levels: the pixel-level and the lesion-level. The Dice coefficient [38] measures the pixel-level similarity between the segmented image and the reference image which can determine the quality of the segmentation model [39]. On the other hand, the miss detection rate and the false-positive rate provide high-level indicators of how well the model predicts a lesion in the image, which can be easily understood by humans [18]. The miss detection rate represents the percentage of reference masses that the algorithm failed to detect, while the false positive rate refers to the percentage of detected masses that are not actual masses. In Table I, we present a comparison between the outcomes of the original Deeplabv3+ model and the modified double-dilated model. The results show that integrating the modified convolution module with double dilation rates was successful in enhancing the network's performance in terms of both the Dice similarity and the Miss Detection rate. The Miss Detection rate refers to the percentage of lesions that are wrongly segmented into the background. This enhancement can be attributed to the increase in the inner kernels' resolution due to the double-dilated module, which encodes multi-scale context information present in the input image while maintaining local information. By minimizing the number of missed detections, early diagnosis can be expedited, and fatality rates can be reduced. However, the proposed model still generated false (unmatched) lesions at the same rate as the single-dilated network, which was relatively high at 20% of the actual number of tumors. This observation is consistent with previous studies that reported relatively high false positive rates in CAD systems [40]–[42]. This highlights the need for more investigation into this behavior in future work.

TABLE I. Results of the 5-fold validation of the original Deeplabv3+ model and the modified model with the double-dilated module. Both lesion-level and pixel-level evaluation metrics are reported.

	Deeplabv3+	Double-Dilated Deeplabv3+
Dice Similarity	0.79	**0.81**
Miss Detection Rate	0.08	**0.04**
False Positive Rate	0.20	0.20

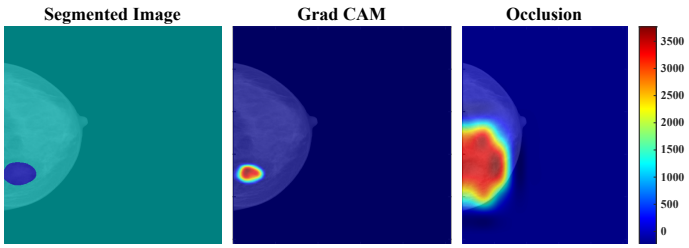

Fig. 4: Examples of segmented mammogram images generated by the modified network shown along with explanations of the output segmentation. Different explanation maps are shown using Activation Visualization, Grad-CAM, and Occulsion Sensitivity.

B. Explainablity via Visualization

To improve the interpretability of our black-box segmentation network, we adopt two explainability methods that provide visual explanations to the segmented output, namely, Grad-CAM and Occlusion Sensitivity. Fig. 4 shows examples of heatmaps generated for one mammogram screening using three explanation techniques. Comparing the results of Occlusion Sensitivity and Grad-CAM, we notice that the former generates explanation maps with larger heated areas by assigning higher weights to non-tumor pixels. This could be attributed to the occlusion technique, which involves removing different regions to measure their contribution to the output decision. In contrast, Grad-CAM generates heatmaps that are more constrained to the mass region because it weights the gradients of the activation maps at the last layer, which are highly correlated to the segmentation output.

TABLE II. IMAGE ENTROPY RESULTS FOR EXPLANATION MAPS GENERATED BY DIFFERENT EXPLAINABLE METHODS WITH THE ORIGINAL AND DOUBLE-DILATED SEGMENTATION NETWORKS. THE TABLE SHOWS THE AVERAGE RESULTS FOR ALL IMAGES IN THE VALIDATION SET.

	Grad-CAM	Occlusion
Original	**0.119**	2.526
Double-Dilated	**0.139**	1.445

To compare the complexity of the two explanation methods, image entropy, which measures the randomness in the image, was calculated for all heat maps generated for mammogram screenings in the validation set. Table II displays the average entropy results for different XAI methods when applied to images segmented by both the original Deeplab V3+ network and the double-dilated network. The Occlusion Sensitivity had a relatively high average entropy, with values distributed over a wide range, while Grad-CAM achieved the lowest complexity with an average value of around 0.12 in both networks, indicating the least randomness in explanation mapping. The proposed network resulted in slightly more complex explanations than the original network when using Grad-CAM, possibly due to the added complexity of the double-dilated kernels reflected in the feature maps employed for explanation mapping in these methods. However, this was not the case with the Occlusion maps, as the algorithm used was not dependent on the activation values.

To quantitatively evaluate the truthfulness of the explanation model, Fig. 5 presents the pixel-flipping graphs for various explanation methods when applied to our proposed segmentation network. The graphs show whether or not removing the pixels identified as the most relevant by the explanation techniques leads to a notable decrease in the network's predictive performance. We plotted the average similarity scores achieved at different rates of pixel flipping across the validation set. The results reveal that the Grad-CAM method showed a higher decay rate than the Occlusion Sensitivity method, indicating its superior truthfulness compared to the latter. The pixel-flipping outcomes align with the entropy results and visual

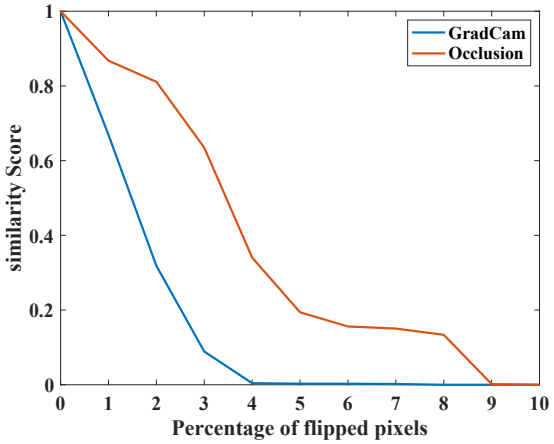

Fig. 5: Plot of pixel flipping graphs for different explanation methods.

analysis, all of which suggest that the Grad-CAM explainability technique provides truthful and understandable explanations for mammogram mass segmentation results. These findings encourage the adoption of this simple yet powerful tool in medical image segmentation networks to enhance their transparency and facilitate their integration into clinical practice.

VI. CONCLUSION

This study addresses two major issues encountered in medical image segmentation systems. Firstly, we tackle the problem of deteriorating local resolutions in medical images, which occurs in existing convolutional neural network-based segmentation architectures, by introducing a double-dilated convolution module that employs complex kernels with denser cores. The module is implemented using the Deeplabv3+ network, and we evaluate our proposed method in performing mass segmentation using mammogram screenings from the publicly available INBreast dataset. Secondly, to overcome the black-box nature of DL networks, we applied and compared Grad-CAM and Occlusion Sensitivity as explainability techniques to provide interpretable segmentation outcomes and compared their performance in terms of complexity and truthfulness.

Our experimental results suggest that the adoption of double-dilated convolution can achieve increased similarity scores and reduced miss-detection rates in CNN-based networks designed for medical image segmentation. Furthermore, our explainable AI investigation demonstrated that Grad-CAM is an effective tool for explaining CAD segmentation outcomes and providing medical professionals with interpretable and truthful information. In future work, we aim to validate the effectiveness of our proposed segmentation and explainability techniques on larger datasets with diverse medical image modalities. Additionally, we plan to extend the concept of double-dilated convolution by developing an N-dilated convolution module that employs a kernel with N sparsity factors on different scales. Finally, we plan to investigate the issue of high false positive rates associated with CAD systems to alleviate the negative psychological impact on patients and reduce the number of unnecessary biopsies.

REFERENCES

[1] H. T. Haridas, M. M. Fouda, Z. M. Fadlullah, M. Mahmoud, B. M. ElHalawany, and M. Guizani, "MED-GPVS: A Deep Learning-Based Joint Biomedical Image Classification and Visual Question Answering System for Precision e-Health," in *ICC 2022 - IEEE International Conference on Communications*, 2022, pp. 3838–3843.

[2] S. Saxena *et al.*, "Role of artificial intelligence in radiogenomics for cancers in the era of precision medicine," *Cancers*, vol. 14, no. 12, article no. 2860, 2022.

[3] H.-P. Chan, L. M. Hadjiiski, and R. K. Samala, "Computer-aided diagnosis in the era of deep learning," *Medical Physics*, vol. 47, no. 5, pp. e218–e227, 2020.

[4] F. A. Mostafa, L. A. Elrefaei, M. M. Fouda, and A. Hossam, "A survey on AI techniques for thoracic diseases diagnosis using medical images," *Diagnostics*, vol. 12, no. 12, article no. 3034, 2022.

[5] J. S. Suri *et al.*, "UNet deep learning architecture for segmentation of vascular and non-vascular images: A microscopic look at unet components buffered with pruning, explainable artificial intelligence, and bias," *IEEE Access*, vol. 11, pp. 595–645, 2023.

[6] J. S. Suri *et al.*, "COVLIAS 1.0Lesion vs. MedSeg: An artificial intelligence framework for automated lesion segmentation in COVID-19 lung computed tomography scans," *Diagnostics*, vol. 12, no. 5, article no. 1283, 2022.

[7] A. K. Dubey *et al.*, "Ensemble deep learning derived from transfer learning for classification of COVID-19 patients on hybrid deep-learning-based lung segmentation: A data augmentation and balancing framework," *Diagnostics*, vol. 13, no. 11, article no. 1954, 2023.

[8] P. K. Jain *et al.*, "Attention-based unet deep learning model for plaque segmentation in carotid ultrasound for stroke risk stratification: An artificial intelligence paradigm," *Journal of Cardiovascular Development and Disease*, vol. 9, no. 10, article no. 326, 2022.

[9] B. Jena *et al.*, "Brain tumor characterization using radiogenomics in artificial intelligence framework," *Cancers*, vol. 14, no. 16, article no. 4052, 2022.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[11] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[12] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv:1511.07122*, Apr. 2016.

[13] P. Wang, P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou, and G. Cottrell, "Understanding convolution for semantic segmentation," in *2018 IEEE winter conference on applications of computer vision (WACV)*, 2018, pp. 1451–1460.

[14] V. Pitroda, M. M. Fouda, and Z. M. Fadlullah, "An explainable AI model for interpretable lung disease classification," in *2021 IEEE International Conference on Internet of Things and Intelligence Systems (IoTaIS)*, 2021, pp. 98–103.

[15] Nillmani *et al.*, "Segmentation-based classification deep learning model embedded with explainable AI for COVID-19 detection in chest X-ray scans," *Diagnostics*, vol. 12, no. 9, article no. 2132, 2022.

[16] "Canadian Cancer Statistics Advisory Committee in collaboration with the Canadian Cancer Society, Statistics Canada and the Public Health Agency of Canada. Canadian Cancer Statistics 2021. Toronto, ON: Canadian Cancer Society; 2021." http://cancer.ca/Canadian-Cancer-Statistics-2021-EN.

[17] J. Mendes, J. Domingues, H. Aidos, N. Garcia, and N. Matela, "AI in breast cancer imaging: A survey of different applications," *Journal of Imaging*, vol. 8, no. 9, p. 228, Sep. 2022.

[18] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "INbreast: Toward a full-field digital mammographic database," *Academic Radiology*, vol. 19, no. 2, pp. 236–248, Feb. 2012.

[19] X. Yu, Q. Zhou, S. Wang, and Y.-D. Zhang, "A systematic survey of deep learning in breast cancer," *International Journal of Intelligent Systems*, vol. 37, no. 1, pp. 152–216, 2022.

[20] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.

[21] X. He, R. S. Zemel, and M. A. Carreira-Perpinán, "Multiscale conditional random fields for image labeling," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, 2004.

[22] J. Yao, S. Fidler, and R. Urtasun, "Describing the scene as a whole: Joint object detection, scene classification and semantic segmentation," in *2012 IEEE conference on computer vision and pattern recognition*, 2012, pp. 702–709.

[23] L. Pei, L. Vidyaratne, M. M. Rahman, and K. M. Iftekharuddin, "Context aware deep learning for brain tumor segmentation, subtype classification, and survival prediction using radiology images," *Scientific Reports*, vol. 10, no. 1, p. 19726, Nov. 2020.

[24] P. Lakhani, "The importance of image resolution in building deep learning models for medical imaging," *Radiology: Artificial Intelligence*, vol. 2, no. 1, article no. e190177, Jan. 2020.

[25] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1520–1528.

[26] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected CRFs," *arXiv:1412.7062*, Jun. 2016.

[27] Z. Wu, C. Shen, and A. van den Hengel, "Bridging category-level and instance-level semantic image segmentation," *arXiv:1605.06885*, 2016.

[28] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLabs: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *arXiv:1606.00915*, 2016.

[29] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking Atrous Convolution for Semantic Image Segmentation," *arXiv:1706.05587*, 2017.

[30] L. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," *arXiv:1802.02611*, 2018.

[31] R. Hussein, S. Lee, R. Ward, and M. J. McKeown, "Semi-dilated convolutional neural networks for epileptic seizure prediction," *Neural Networks*, vol. 139, no. Complete, pp. 212–222, 2021.

[32] P. Dardouillet, A. Benoit, E. Amri, P. Bolon, D. Dubucq, and A. Crédoz, "Explainability of Image Semantic Segmentation Through SHAP Values," in *Proc. 26TH International Conference on Pattern Recognition 2nd Workshop on Explainable and Ethical AI (ICPR-XAIE)*, Montreal, Canada, Aug. 2022.

[33] M. Abukmeil, A. Genovese, V. Piuri, F. Rundo, and F. Scotti, "Towards explainable semantic segmentation for autonomous driving systems by multi-scale variational attention," in *2021 IEEE International Conference on Autonomous Systems (ICAS)*, Aug. 2021.

[34] V. Couteaux, O. Nempont, G. Pizaine, and I. Bloch, "Towards interpretability of segmentation networks by analyzing DeepDreams," in *Interpretability of Machine Intelligence in Medical Image Computing and Multimodal Learning for Clinical Decision Support*, 2019, vol. 11797, pp. 56–63.

[35] R. R. Kontham, A. K. Kondoju, M. M. Fouda, and Z. M. Fadlullah, "An end-to-end explainable AI system for analyzing breast cancer prediction models," in *2022 IEEE International Conference on Internet of Things and Intelligence Systems (IoTaIS)*, 2022, pp. 402–407.

[36] L. W. Bassett, K. Conner, and I. Ms, "The Abnormal Mammogram," *Holland-Frei Cancer Medicine. 6th edition*, 2003.

[37] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes challenge: A retrospective," *International Journal of Computer Vision*, vol. 111, no. 1, pp. 98–136, Jan. 2015.

[38] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[39] Z. Wang, E. Wang, and Y. Zhu, "Image segmentation evaluation: A survey of methods," *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5637–5674, Dec. 2020.

[40] N. Dhungel, G. Carneiro, and A. P. Bradley, "Automated mass detection in mammograms using cascaded deep learning and random forests," in *2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, Nov. 2015.

[41] R. M. Nishikawa, M. Kallergi, and C. G. Orton, "Computer-aided detection, in its present form, is not an effective aid for screening mammography," *Medical Physics*, vol. 33, no. 4, pp. 811–814, 2006.

[42] K. Loizidou, R. Elia, and C. Pitris, "Computer-aided breast cancer detection and classification in mammography: A comprehensive review," *Computers in Biology and Medicine*, vol. 153, article no. 106554, Feb. 2023.