

# Segmentation Model for Detection of Lung Cancer Nodule

Avinash S<sup>1</sup>, Rath S<sup>2</sup>

*Department of Computer Science and Engineering, Government College of Technology, Coimbatore*

*Department of Computer Science and Engineering, Government College of Technology, Coimbatore*

\*\*\*

**Abstract** - Segmentation of Lung nodule is considered as one of the most significant procedures in early diagnosis of Lung Cancer. Due to various difference in shapes and small size of lung nodules the segmentation became a difficult task. In this study we propose a modified U-Net architecture, this integrates spatial-attention, residual structure, multi-scale convolution and depth wise separable convolutions. The architecture supplants the U-Net's convolutional block with a residual unit for faster convergence, improving the global modelling by spatial attention and a bilinear up sampling for addressing the issue of checkboard artifacts caused due to the traditional transpose convolution. Results of the experiments on the LIDC dataset has obtained a dice score 84.11% and IoU 72.78%, which are found to be improved by 1.5% and 1.25% when compared with the U-Net architecture.

**Key Words:** CT, U-Net, IoU, Dice

## 1.INTRODUCTION

Lung cancer is the leading cause of cancer-related deaths worldwide. It is a medical condition where the cell growth is uncontrolled in the lung region and spreading further to other organs, leading to mortality, initially starting by cells lining the wind pipes. Accurate and automated lung nodule segmentation from computed tomography (CT) images are therefore a critical prerequisite for prognosis, treatment planning, and disease monitoring. However, lung nodule segmentation is challenging due to factors such as small nodule size, irregular shapes, low contrast, and similarity to surrounding tissues. Over the years, researchers have proposed a wide range of traditional and deep learning-based approaches to address these challenges. With all the existing medical imaging technologies such as Magnetic Resonance Image (MRI), Computed Tomography (CT), and others, low-dose CTs have an advantage of low radiation and high resolution among other technologies, thus it is used for early screening of lung cancer. Because of the fact that there are a number of lung cancer patients, radiologists are supposed to process a large amount of CT images every day. With the variable shape and smaller sizes of the lung nodules, it becomes harder to diagnose and sometimes leading to misdiagnosis. In the recent

years, due to the advancement in computer-aided detection (CAD) for segmentation serves as a great assistance in reducing the time needed to analyse a patient's condition.

The techniques for lung nodule segmentation are mainly divided into traditional and deep learning based. Techniques such as threshold and region growing, active contour mathematical model based and clustering are under the hood of traditional segmentation. Techniques based on deep learning are divided as 2D and 3D segmentations. The traditional methods work primarily based on the shallow feature information like texture and greyscale of the image.

The 2D segmentation makes use of 2D convolution for extracting features from a 2D images, these techniques are widely used in various image segmentation tasks. One of the noted downsides of this technique is that, it is not capable of perceiving the global context information. Based on this Fully Convolutional Neural Network (FCN) Ronneberger et al. [3] proposed a U-Net architecture specific for medical images segmentation. Liang-Chieh Chen et al. [4] proposed a framework, Atrous Spatial Pyramid Pooling (ASPP) with an intention of improving semantic segmentation accuracy, by classifying every pixel of the image and assigning object labels. The traditional CNN based segmentations were facing difficulties such as loss of spatial resolution because of down sampling and in handling objects of multiple scales. Atrous also known as dilated convolution, instead of using regular contiguous convolutional kernels, this model inserts spaces between the kernel values. This model preserves spatial resolution which is lost by traditional CNN as it reduces image size repeatedly. Along with the multi-scale feature fusion which improves results of segmentation, the attention mechanism plays a vital role in image segmentation, as it suppresses the irrelevant regions by varying the weights of channels to highlight the required regions. The lack of spatial localization information resulting to direct resolution recovery by up-sampling problem was resolved by conducting a series of studies to extract features by CNN, then encoded the features extracted using transformer and then to decode

them for obtaining the image segmentation results, this is achieved by combining CNN and Transformer for medical image segmentation. Cao et al.[5] proposed a U-shaped architecture with the help of Swin-Transformer for recovery of spatial information by fusion of multi-scale features from the encoder with features of the decoder using skip connections.

The 3D segmentation uses 3D convolution for extraction of features from 3D images. Comparing this with the 2D segmentation, it is found that 3D are better in learning spatial spatial relationships among nodes and surrounding tissues. Duo et al.[6] proposed a 3D network for segmentation of the liver and cardiac vessels. Milletari et al.[7] used residual linkage from ResNet and utilized 3D instead of 2D for segmentation of volume of prostate. Even with all upsides the training of 3D networks takes much time, with the values of the evaluation metrics weren't necessarily exceed the 2D networks. This paper proposes a 2D model with multi-scale features and spatial attention mechanism for segmentation of lung nodule.

Here, we propose a modified UNet for segmentation of lung nodule comprising a depth-wise separable convolution, residual unit, spatial attention, adaptive ASPP and hybrid up-sampling. The model replaces original U-Net[8] convolutional blocks with residual unit for handling the gradient explosion because of deepening the number of layers in the network. In contrast from the traditional U-Net, here spatial attention and an adaptive ASPP are added to the middle layer for global context modelling of the model. The features from the encoders are used as the input sequence for the bottleneck, which are then recoded and used as input for up-sampling, the spatial details are recovered by combining them with encoder's features in every layer by skip connections as same as the original U-Net. Hybrid up-sampling is used in place of transposed convolution for recovery of the image resolution.

## 2. Materials and Methods

### A. DATASET

We used LIDC (Lung Image Database Consortium), a public dataset for this research work. It is one of the most commonly used resource for development, training and evaluation of CAD (Computer Assisted Diagnosis) [2] for lung cancer detection and diagnosis. It consists of CT (Computer Tomography) scan images of 1018 subjects collected from seven academic centers and eight medical imaging companies. All these images are labelled by different radiologists and categorized by size of the nodule [1]. The consortium consists a total of 244,527 images

which has both cancerous and non-cancerous images, for the task of segmentation of lung nodule we only need the cancerous images, thus we made us of 15,096 images which are cancerous. The processed dataset is obtained from J. Hou et al.[8] in their work they made use of the LIDC dataset and another dataset obtained from The Department of Medical Imaging of the Fourth Affiliated Hospital of Guangxi Medical University. They first processed the DICOM files are converted to PNG by a software named MicroDicom. Image samples are displayed in Figure 1.

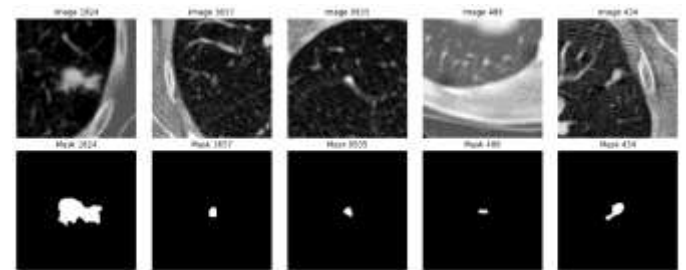


**Figure 1. Lung nodule dataset LIDC dataset.**

The LIDC dataset has a total of 15,096 images with nodules, which is split as 60% for training, 20% for testing and 20% for validation.

### B. ARCHITECTURE

In this work, a framework based on residual unit, spatial attention, hybrid up-sampling and adaptive ASPP is proposed for segmentation of lung nodule. The overall structure is shown in fig. 2, input to the network is a 128x128x1 medical image. The basic structure of the network is a residual unit, consisting of two 3x3 depth-wise separable convolutional units (3 x 3 Conv + BN + ReLu and 3 x 3 Conv + BN) and a 1 x 1 convolutional unit (1 x 1 Conv + BN) for linkage, followed by a 2 x 2 max pooling layer. The high-level semantic features obtained from continuous down sampling are fed to the bottleneck layer with adaptive ASPP and the spatial attention, the output resolution from the bottleneck is then reshaped into the input of the up-sampling. The adaptive ASPP module structure is: a sequence of 3 2D convolutions with padding and dilations 0, 2, and 4. With increase in the number of channels the more frequent the information is extracted and detailed data are reacted, the hybrid up-sampling combines learnable up-sampling, interpolation based up-sampling to get smooth reconstruction, preserved boundaries and stable gradients. It makes use of PixelShuffle, Bilinear Up-sampling, the concatenation from the skip connections and a final 1 x 1 Conv fusion.

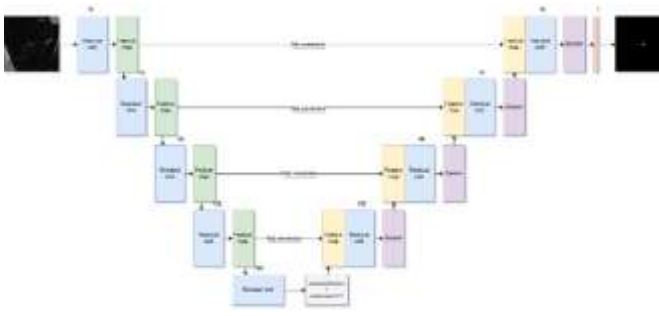


**Figure 2. Overall architecture**

### C. RESIDUAL UNIT

Residual unit's structure is shown in fig.3, the structure was proposed by He et al.[9], which is primarily focused in solving the problem of gradient disappearance with respect to the depth of the network. The motive of structure is to add a skip connection in every two layers of the network. Integrating the residual unit with the U-Net the paths are convolved by  $1 \times 1$  to make the count of channels persistent with that of the residual path.

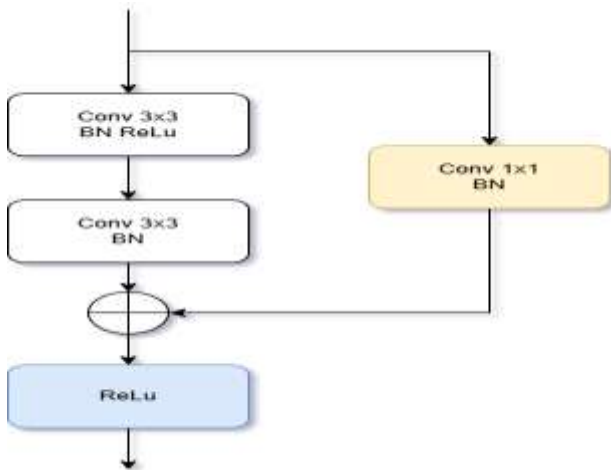


**Figure 3. Residual Unit**

### D. SPATIAL ATTENTION

A lightweight gating attention mechanism for learning where to emphasize features by suppressing irrelevant background responses, this computes a 2D spatial importance mask. It has a channel pooling, a combination of mean pooling and max pooling, which is inherited from Convolutional Block Attention Module. (CBAM) [10], a feature aggregation, spatial-attention estimation and feature recalibration. For a given input feature map:  $X \in \mathbb{R}^{B \times C \times H \times W}$  where B represents batch size, C represents number of channels and H, W are spatial dimensions, the module generates a spatial attention map:  $X \in \mathbb{R}^{B \times 1 \times H \times W}$  which is used for recalibration of the original features.

The input feature map is compressed along the channel dimensions with the help of mean and max pooling

producing two spatial descriptors  $F_{avg}, F_{max} \in \mathbb{R}^{B \times 1 \times H \times W}$ . Average pooling computes the mean response across all channels, whereas the max pooling extracts the strongest activation value across channels. Both are used together, thus allowing the network to generate a rich spatial representation.

The pooled descriptors are concatenated resulting  $F_{cat} \in \mathbb{R}^{B \times 2 \times H \times W}$  in feature aggregation, this contains both global and spatial information.

A convolution operation with a  $7 \times 7$  kernel is applied which computes an attention map  $A = \sigma(f^{7 \times 7}(F_{cat}))$  where  $\sigma$  denotes sigmoid activation function and  $f^{7 \times 7}$  denotes convolution. The large kernel size helps the module capture a border spatial context and neighboring dependencies.

In the end attention map is multiplied element-wise with input feature map to enhance important feature map.

### E. ADAPTIVE ASPP

This module is designed for the task of capturing multi-scale contextual information using parallel dilated convolutions. The traditional convolutions are done at a fixed receptive field, which limits the ability to model fine details and large contextual structures simultaneously [11]. It follows the flow: Parallel multi-scale feature extraction, multi-scale feature fusion and an adaptive feature integration.

The module applies three parallel convolution branches for: local feature extraction, medium-scale context and large-scale context with varied kernel sizes 1, 3 with dilation rate 2 and another 3 with a dilation rate 4 executed simultaneously. The dilation is found to increase receptive field size without increasing the number of parameters significantly and reducing the feature map resolution by pooling. Output of all the branches are concatenated, which contains local features, medium-range context and large-scale contextual information.

The concatenated features are combined together using a  $1 \times 1$  Conv, which is computed as  $Y = \text{Conv}_{1 \times 1}(F_{multi})$ , this operation learns how to combine information from different spatial scales adaptively.

### F. HYBRID UP-SAMPLING

This module is designed for specific tasks that require high resolution predictions such as semantic segmentation, this is designed for feature map up-sampling. The primary purpose is to increase spatial resolution while refining feature representation, by preserving semantic consistency and structural details. For

the given input feature map:  $X \in \mathbb{R}^{B \times C \times H \times W}$ , the module produces an output:  $Y \in \mathbb{R}^{B \times C_{out} \times 2H \times 2W}$ , where the spatial dimensions are doubled. The module consists of a bilinear up-sampling that performs an interpolation-based up-sampling for doubling the spatial dimensions,  $H \times W$  as  $2H \times 2W$ , this is generally used for a smoother gradient-friendly resampling. The bilinear up-sampling focuses on restoring spatial resolution after down-sampling, while preserving coarse structure. After the interpolation, a  $3 \times 3$  Conv is applied for the purpose of fusing the feature channel, recovery of local details and to enable non-linear refinement.

Unlike the naïve transposed convolutions, which has the potential to introduce checkboard artifacts, this approach separates up-sampling and feature learning. By up-sampling followed by refining, the module minimizes feature loss and allow the network to adjust local response dynamically.

### G. LOSS FUNCTION

The loss functions used addresses two common challenges, i) extreme class imbalance, which is common in medical image segmentation as the foreground occupy a small fraction of the image and ii) ambiguous regions, like edges of tumors.

Tversky loss [12] is a generalized Dice-based loss that allows a dedicated control over the relative penalties for false positives (FP) and false negatives (FN). This plays a significant role in medical images as FNs are more critical than FPs. Standard Dice can miss out addressing class imbalance. For a predicted mask  $P \in \{0,1\}^N$  and ground-truth mask  $T \in \{0,1\}^N$  flattened into vectors of length N:

$$TP = \sum_i P_i T_i$$

$$FP = \sum_i P_i (1 - T_i)$$

$$FN = \sum_i (1 - P_i) T_i$$

The Tversky index is:

$TI = \frac{TP + smooth}{TP + \alpha FP + \beta FN + smooth}$  where  $\alpha, \beta$  control the penalty for FP and FN, and smooth prevents division by zero.

The Tversky loss is:

$$\mathbb{L}_{Tversky} = 1 - TI$$

Focal loss is designed for mitigating the dominance of easy examples in standard binary cross-entropy (BCE). Given a BCE loss:

$$BCE_i = [T_i \log P_i + (1 - T_i) \log(1 - P_i)]$$

Focal loss modifies BCE as:  $\mathbb{L}_{focal} = \alpha(1 - p_t)^\gamma BCE$  where  $p_t$  represents the predicted probability of the true class,  $\gamma$  is a focusing parameter used to down weight easy examples and  $\alpha$  balances positive and negative classes.

The combined loss integrates the strength of both loss functions. Where Tversky ensures handling class imbalance and emphasizes sensitivity to false negatives, focal improves optimization on hard pixels, enhancing boundary delineation and small structure detection. The combination yields a stronger training behavior and improves segmentation accuracy for small structure

### 3. RESULTS

To evaluate the proposed network structure introduced in the paper, the ablation experiments were performed on the LIDC dataset with the standard U-Net, SMR-UNet and the proposed model are presented in Table 1 and the segmentation by different models are shown in fig. 4, 5 and 6. The significance of replacing the base convolutional block with a residual block and other modifications to achieve lung nodule segmentation, the Dice and IoU are improved by 1.3% and 2.27% respectively in the results of the test sets, the modified network has an improvement for complex shape of nodule segmentation.

**Table 1. Performance of different models on LIDC**

| Model         | Dice   | IoU    | Precision | Recall | F1     |
|---------------|--------|--------|-----------|--------|--------|
| Original UNet | 0.8359 | 0.7180 | 0.8524    | 0.8199 | 0.8359 |
| SMR-UNet      | 0.8249 | 0.7051 | 0.7775    | 0.8848 | 0.8249 |
| Proposed      | 0.8468 | 0.7343 | 0.8336    | 0.8605 | 0.8468 |

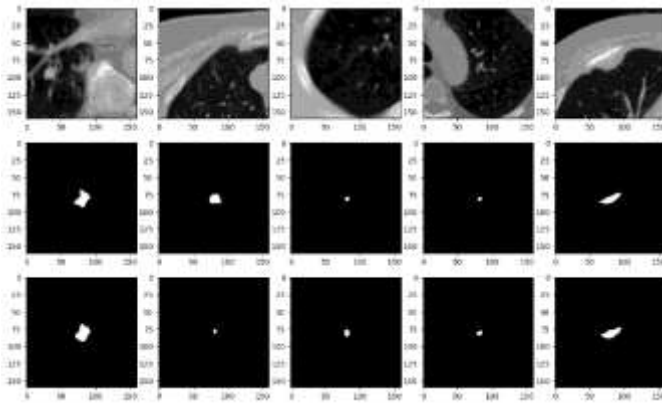


Figure 4. Segmentations of original UNet

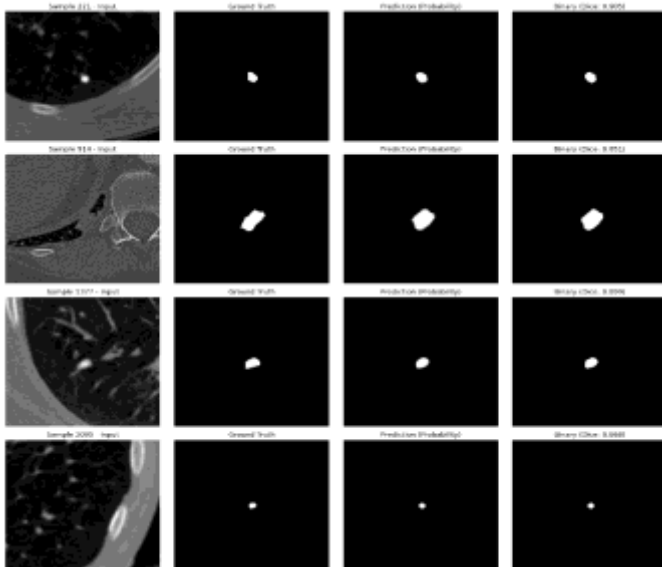


Figure 5. Segmentations of SMRUNet

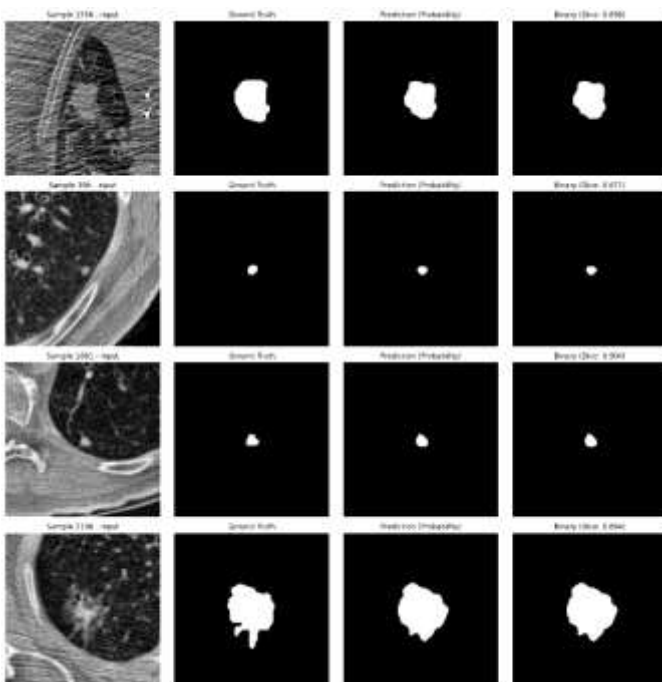


Figure 6. Segmentations of the proposed work

## 4. CONCLUSIONS

In this paper, a 2D segmentation network for lung nodule segmentation is proposed, it uses U-Net as the backbone network, and improves the gradient disappearance and gradient explosion by replacing the convolutional blocks of the original U-Net with residual blocks, the bottleneck with a combination of spatial attention and adaptive ASPP, further the regular up-sampling is replaced with a hybrid up-sampling method, the convolutions in residual unit are done by depth-wise convolution which maps shallow features to deep layers. The attention mechanism is used to reshape the high-level features extracted by the encoder to enhance the global modeling capability of the network. In future work, we will explore new network architectures to improve the robustness and generalization of different CT image datasets.

## REFERENCES

1. McNitt-Gray MF, Armato SG 3rd, Meyer CR, Reeves AP, McLennan G, Pais RC, Freymann J, Brown MS, Engelmann RM, Bland PH, Laderach GE, Piker C, Guo J, Towfic Z, Qing DP, Yankelevitz DF, Aberle DR, van Beek EJ, MacMahon H, Kazerooni EA, Croft BY, Clarke LP. The Lung Image Database Consortium (LIDC) data collection process for nodule detection and annotation. Acad Radiol. 2007 Dec;14(12):1464-74. doi: 10.1016/j.acra.2007.07.021. PMID: 18035276; PMCID: PMC2176079.
2. Armato III, S. G., McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., Zhao, B., Aberle, D. R., Henschke, C. I., Hoffman, E. A., Kazerooni, E. A., MacMahon, H., Van Beek, E. J. R., Yankelevitz, D., Biancardi, A. M., Bland, P. H., Brown, M. S., Engelmann, R. M., Laderach, G. E., Max, D., Pais, R. C., Qing, D. P. Y., Roberts, R. Y., Smith, A. R., Starkey, A., Batra, P., Caligiuri, P., Farooqi, A., Gladish, G. W., Jude, C. M., Munden, R. F., Petkovska, I., Quint, L. E., Schwartz, L. H., Sundaram, B., Dodd, L. E., Fenimore, C., Gur, D., Petrick, N., Freymann, J., Kirby, J., Hughes, B., Castele, A. V., Gupte, S., Sallam, M., Heath, M. D., Kuhn, M. H., Dharaiya, E., Burns, R., Fryd, D. S., Salganicoff, M., Anand, V., Shreter, U., Vastagh, S., Croft, B. Y., Clarke, L. P. (2015). Data From LIDC-IDRI [Data set]. The Cancer Imaging

- Archive.<https://doi.org/10.7937/K9/TCIA.2015.LO9QL9SX>
3. O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent., 2015, pp. 234–241
  4. L. Liu, Q. Dou, H. Chen, J. Qin, and P.-A. Heng, “Multi-task deep model with margin ranking loss for lung nodule analysis,” IEEE Trans. Med. Imag., vol. 39, no. 3, pp. 718–728, Mar. 2020.
  5. F. Shi, B. Chen, Q. Cao, Y. Wei, Q. Zhou, R. Zhang, Y. Zhou, W. Yang, X. Wang, R. Fan, F. Yang, Y. Chen, W. Li, Y. Gao, and D. Shen, “Semi supervised deep transfer learning for benign-malignant diagnosis of pulmonary nodules in chest CT images,” IEEE Trans. Med. Imag., vol. 41, no. 4, pp. 771–781, Apr. 2022.
  6. Q. Dou, L. Yu, H. Chen, Y. Jin, X. Yang, J. Qin, and P.-A. Heng, “3Ddeeply supervised network for automated segmentation of volumetric medical images,” Med. Image Anal., vol. 41, pp. 40–54, Oct. 2017.
  7. F. Milletari, N. Navab, and S.-A. Ahmadi, “V-Net: Fully convolutional neural networks for volumetric medical image segmentation,” in Proc. 4th Int. Conf. 3D Vis. (3DV), Oct. 2016, pp. 565–571.
  8. J. Hou, C. Yan, R. Li, Q. Huang, X. Fan and F. Lin, "Lung Nodule Segmentation Algorithm With SMR-UNet," in IEEE Access, vol. 11, pp. 34319-34331, 2023, doi: 10.1109/ACCESS.2023.3264789.
  9. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770–778.
  10. Sanghyun Woo, Jongchan Park, Joon-Young Lee, In So Kweon “CBAM: Convolutional Block Attention Module”  
<https://doi.org/10.48550/arXiv.1807.06521>.
  11. Zhu, T.; Liu, Q.; Zhang, L. “An Adaptive Atrous Spatial Pyramid Pooling Network for Hyperspectral Classification”. *Electronics* **2023**, *12*, 5013.  
<https://doi.org/10.3390/electronics12245013>
  12. Salehi, Seyed Sadegh & Erdogmus, Deniz & Gholipour, Ali. (2017). Tversky Loss Function for Image Segmentation Using 3D Fully Convolutional Deep Networks. 379-387. 10.1007/978-3-319-67389-9\_44.
  13. G. Litjens et al., “Evaluation of prostate segmentation algorithms for MRI: The PROMISE12 challenge,” Med. Image Anal., vol. 18, no. 2, pp. 359–373, Feb. 2014.
  14. Xizhou Zhu, Dazhi Cheng, Zheng Zhang, Stephen Lin, Jifeng Dai “An Empirical Study of Spatial Attention Mechanisms in Deep Networks”  
<https://doi.org/10.48550/arXiv.1904.05873>
  15. Md Arif Hasan, Krishno Dey “Depthwise Separable Convolutions with Deep Residual Convolutions”  
<https://doi.org/10.48550/arXiv.2411.07544>