SCIENTIFIC CULTURE
www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 4, (2026), pp. 10939-10952
Open Access. Online & Print

DOI: 10.5281/zenodo.124261049

MEDICAL IMAGE SEGMENTATION WITH PYRAMID ATTENTION NETWORKS AND A CASE STUDY ON LIVER TUMORS

M. Swapna[1]*, Dr. B. Sujatha[2]

[1]Research scholar, Computer science and engineering department,
Osmania university, Hyderabad, Telangana, India - 500007.
maragoniswapna761@gmail.com
[2]Assistant professor, Computer science and engineering department,
Osmania university, Hyderabad, Telangana, India - 500007.
sujatha.banothu@gmail.com

Corresponding Author: M. Swapna
(maragoniswapna761@gmail.com)

ABSTRACT

Accurate segmentation of liver tumors in medical images is paramount for the accurate delineation of pathological areas that are integral for diagnosis and treatment planning. Conventional segmentation methods have many limitations, especially in the case of tumor boundaries with complex geometry and heterogeneous image intensity, requiring more sophisticated and efficient deep learning architectures. The main goal of the research is to develop an Enhanced Pyramid Attention Network (PAN) model to increase liver tumor segmentation accuracy. The goal of this study is to resolve the weakness of the current CNNs regarding the ability to extract global context and multi-scale features while managing the obstacles introduced by class imbalance and noisy medical images. The proposed model is based on novel modifications of the conventional PAN architecture, in which EfficientNet-B0 is used as the encoder backbone for more effective feature extraction, via compound scaling. By introducing Squeeze-and-Excitation (SE) blocks that recalibrate channel-wise feature responses dynamically, we are able to highlight channels carrying relevant components and suppressing noisy responses. The Atrous Spatial Pyramid Pooling (ASPP) helps to capture both fine-grained details and global contexts. Residual connections and attention mechanisms for fusing features and propagating gradients improve model stability and convergence speed. The experimental outcomes show that the proposed pipeline Enhanced PAN model significantly excel the latest segmentation, with a mean Intersection over Union (mIoU) score of 0.9323, suggesting a high level of accuracy, as well as more robustness in the specific liver tumor boundary segmentation.

KEYWORDS: Medical Image Segmentation, Liver Tumor Segmentation, Enhanced Pyramid Attention Network, EfficientNet-B0, Squeeze-and-Excitation Block, Atrous Spatial Pyramid Pooling.

1. Introduction

Medical image segmentation associates a label to every pixel of an image, which is fundamental in modern healthcare since it allows the precise delineation of anatomical structures and pathological areas [1]. In medical image analysis approaches, automated segmentation algorithms are exploited to detect and segment different biological structures for applications, such as disease classification, therapy scheduling, and augmented surgery [2]. Radiological imaging, including computed tomography (CT) and magnetic resonance imaging (MRI) are widely used imaging modalities for screening and characterizing abnormalities in soft tissues and organs. But the essential complexity of medical images due to the presence of noise, low contrast and a high inter-patient variability makes their precise segmentation an intricate challenge. Consequently, efficient deep learning methods have become increasingly important for tackling these challenges. In particular, one of the most significant tasks in oncology is liver tumor segmentation, as the accurate delineation of tumors can play a major role in diagnosis, prognosis, and treatment planning [3]. Liver tumors have different morphologies and appearances, like its segmentation is not a trivial problem [4].

Manual segmentation by radiologists is time-consuming and prone to intra- and inter-observer variability. Although machine learning and deep learning methods have shown the best potential to automate this process, these classical convolutional neural network (CNN)-based models sometimes have difficulties in capturing the global context and with tumor shape, size, and intensity variability. As a result, there is still a strong urgency for more complex architectures capable of modeling local and global dependencies in medical images. Although deep learning-based segmentation has achieved varying degrees of success, current approaches often suffer from challenges in both feature extraction and spatial attention [5]. Most CNN-based architectures use fixed-sized convolution filters, limiting the receptive field and preventing the network from capturing long-range dependencies. Moreover, medical images tend to have a problem called class imbalance, where abnormal regions such as tumors take up much smaller area of the image as opposed to the background/ healthy tissue [6]. This disproportion may result in erroneous predictions, where smaller tumor sections are neglected or misclassified. Moreover, variations in scanner settings, patient anatomy, and image acquisition protocols add to the non-uniformity faced by the models. Attention mechanisms have been proposed to tackle these issues and selectively emphasize the most relevant image areas. Although self-attention-based architectures, especially the Transformers, have proven to be successful in processing natural images, they are not directly applicable to medical imaging due to high computational cost and memory consumption. However, Pyramid Attention Networks (PAN) offer an appealing alternative using multi-scale attention instead, improving feature representation across spatial scales [7]. The hierarchical approach utilized allows the model to capture and incorporate local and global contextual information, which ultimately improves overall segmentation performance.v The more effective and reliable solution to the medical image segmentation can be achieved by using the pyramid attention networks. Because of this hierarchical structure, PANs can perform better feature aggregation and learn relevant tumor region while minimizing irrelevant background information [8]. This architecture is especially appropriate for liver tumor segmentation, a task where precise boundary delineation is vital for clinicians' decisions. By leveraging attention-based modeling and multi-scale feature extraction, PANs achieve superior performance compared to traditional CNN-based methods, showcasing their effectiveness in the realm of medical image analysis.

2. Literature Survey

Yanjun Peng et al. [9] proposed a high-resolution multi-scale attention encoder and a full-attention decoder model. In addition, they proposed a scale attention mechanism by merging channel and spatial attention to improve the generation of output. Also, a classification branch was introduced to distinguish between tumor and non-tumor (benign lesion) regions. Spider-Net achieved state-of-the-art performance on kidney, pancreas and liver datasets of various organ failures surpassing CNNs and transformers .Chaopeng Wu et al. [10] introduced liver cancer, preprocessing and fusion methods of multimodal imaging. They examined deep learning (DL) techniques, such as CNNs and U-Net, for image fusion segmentation. There is also a review of a number of evaluation metrics and datasets for segmentation model assessment. The work points out several challenges including data imbalance, generalization, and interpretability, as well as opportunities for future research. DL has emerged as a powerful tool for medical image

segmentation, offering the potential to significantly enhance the accuracy and efficiency of liver cancer segmentation [3]. This review may help physicians. Hao Li et al. [11] proposed a new type of 3D large-kernel (LK) attention module that increased accuracy for both multi-organ and tumor segmentation. The integration of biologically inspired self-attention and convolution which identifies local contextual information, long-distance dependencies, and channel adaptation. LK Convolution We decompose the LK convolution to improve computational efficiency while maintaining effectiveness. This module is also very flexible, which can be integrated into CNN architectures such as U-Net. Miao Liao et al. [12] introduced a new dose forecast network model for liver cancer based on hierarchical feature fusion and interactive attention. First, the model uses a feature extraction module to obtain multi-scale features of various inputs. Then a dedicated fusion module hierarchically fuses these features. An attention-based decoder reconstructs the fused features to a dose distribution. Moreover, during training, an autoencoder network outputs a perceptual loss to increase prediction accuracy. When evaluated on private clinical data, the method achieves 0.31 for HI and 0.87 for CI. Guangzhe Zhao et al. [13] introduced a Hybrid Multi-scale Cross-order Fusion Network (HM-Net) for medical image segmentation. They developed a hybrid pyramid attention module (HPAM) to enhance shallow semantic features across spatial and channel dimensions through multi-scale fusion, reducing the semantic gap between the encoder and decoder in skip connections. Additionally, they proposed a cross-order multi-scale fusion decoder, which captures layered features for fusion, minimizing information loss during up-sampling with a feature enhancement module and improving edge sharpness. Extensive experiments on the Synapse and ACDC datasets showed that HM-Net outperforms previous state-of-the-art methods. Chenchu Xu et al. [14] introduced a dual-stream multi-level fusion framework (DM-FF) for accurate liver tumor segmentation from non-contrast multi-modality images. DM-FF employs an attention-based encoder-decoder to extract multi-level features from each modality. It includes two fusion modules: one enhances shared representations across modalities, while the other identifies differences to reduce conflicts. Ashwini Kumar Upadhyay et al. [15] discussed recent deep learning structures to tackle dataset related challenges in medical image segmentation. They analyzed U-Net-based model performance on a segmentation dataset for lung infections associated with COVID-19 via CT, verifying that the original U-Net performs well. In the face of challenges such as data scarcity, high annotation costs, and distribution shifts, they discussed techniques including active learning, data augmentation, domain adaptation, and self- and semi-supervised learning.. Hejun Huang et al. [16] presented a way to balance attention weights across both channel and spatial dimensions adaptively. It employs a multi-scale depth-wise convolutional module to obtain spatial relationships without losing channel priors. They proposed CPCANet for medical image segmentation based on CPCA. CPCANet achieved superior segmentation performance over two public datasets. And it did so better than state-of-the-art algorithms, and with fewer computational resources.

Jie Wu et al. [17] have proposed MSGH, a new segmentation model based on Graph Neural Networks (GNN) that improves geometric representation with the help of GNN for image segmentation. MSGH fuses the multi-scale features from the Pyramid Feature and Graph Feature branches to promote information interaction between networks. The model uses graph contrastive representation learning with self-supervised learning to account for category imbalance in medical images. In addition, a decoder equipped with the transformer is integrated to better restore the fine details of the image.

3. Proposed Model

3.1 Liver Tumor Segmentation

Accurate diagnosis and treatment planning relies on the liver tumor segmentation. It involves detecting and segmenting tumor regions in medical images like CT scans. Segmentation is done with Enhanced PAN (Pyramid attention Network) architecture in this work. Because it can very well capture the spatial hierarchies of features and contextual information between features over pyramid attention method, PAN is well suited for semantic segmentation tasks. It is especially useful for representing the heterogeneous structures of liver tumors of varying sizes and irregular shapes.

- Base Model (Pyramid Attention Network (PAN) with EfficientNet-B0 Encoder)

The base model for this implementation is the Pyramid Attention Network (PAN), with EfficientNetB0 as the encoder backbone. Since PAN is in charge of gathering the spatial hierarchies and contextual information, it is highly preferable over semantic segmentation tasks. specifically

implements EfficientNet-B0, which is a scalable convolutional network that uniformly scales all dimensions of the depth/width/resolution based on model efficiency. This blend results in a powerful yet computationally efficient model, suiting intricate medical image segmentation challenges such as liver tumor detection.

3.2 Pyramid Attention Network (PAN) Architecture

The Pyramid Attention Network (PAN) attempts to overcome these limitations which are commonly seen in convolutional networks as they lack the ability to adequately learn long-distance dependencies or suitable contextual information. Traditional architectures are mostly specialized in salient local features, while PAN proposes a Pyramid Attention Mechanism which enables exploiting contextual information at various spatial rates. This enables the model to maintain an effective balance between global context and local spatial details that is critical for accurate segmentation of liver tumors that may vary in size and shape.

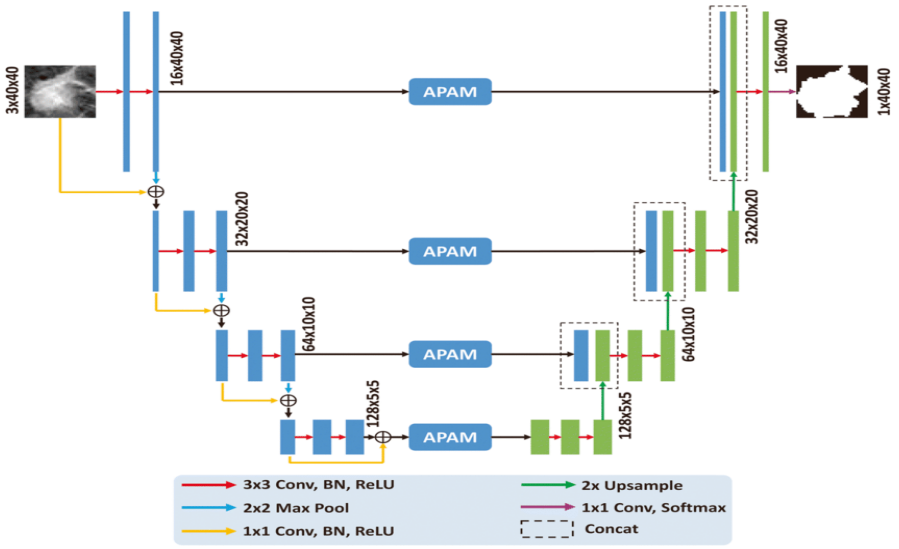


Figure 1: Pyramid Attention Network (PAN) Architecture for Semantic Segmentation

Pyramid Attention Network (PAN) architecture for semantic segmentation is shown in figure 1. At multiple scales, APAM (Attention Pyramid Attention Module) is used to capture the attention of contextual information and spatial hierarchies of key components. Pretrained on ImageNet, the encoder path, which is also known as the downward path, is typically 3x3 convolutional layers with batch normalization and ReLU activation, 2x2 Max Pooling layers or average pooling layers to extract the features of the input image on multiple scales, which is then down sampled. The APAM modules process feature maps by attention, to specifically enhance meaningful regions. In the decoder pathway, the model uses 2x Up sampling layers to restore spatial resolution, concatenating features from corresponding encoder levels via skip connections. 1x1 convolutional layers are employed to reduce the number of channels and produce pixel-wise class probabilities through a SoftMax activation. This architectural design enables the model to accurately segment complex structures by balancing global context with local spatial details, making it highly effective for medical image analysis tasks like liver tumor segmentation.

Key Components of PAN:

1. Pyramid Attention Mechanism:
   - PAN consists of a Pyramid Attention Mechanism, which redistributes feature maps by thematically aggregating contextual information from different spatial levels.
   - It leverages multi-level feature maps generated from the encoder and introduces attention mechanisms at the different pyramid levels, corresponding to 1x1, 2x2, 3x3 and 6x6.
   - The activation maps of these feature maps are up sample and concatenated which gives rich representation of spatial hierarchies and global context.
   - This mechanism allows for both capturing long-range dependencies and preserving fine-grained spatial information.
2. Multi-scale Feature Fusion:
   - PAN uses a multi-scale feature fusion strategy to enrich the semantic representation with

surrounding contexts by fusing feature maps from different scales.

- As a result, this method guarantees the model to learn local features (such as the tumor edges) as well as global context (location of a tumor in the liver anatomy).

3. Efficient Upsampling Strategy:

- In the decoder pathway, PAN employs an efficient up sampling strategy that restores the segmentation map's spatial resolution step by step.
- It uses transposed convolutions to up sample feature maps and concatenate them with respective encoder features via skip connections.
- Skip connections retain spatial features that would otherwise become diluted as down sampling takes place leading to more accurate segmentation.

3.3 EfficientNet-B0 Encoder

Because EfficientNet-B0 balances the model accuracy and efficiency with a compound scaling method, it is selected as the encoder backbone. This approach balances the three crucial axes depth, width and resolution to achieve best in class performance at a low computational cost. Transfer learning from EfficientNet-B0 pre-trained on ImageNet is performed to speed up convergence of the model, reducing overfitting on small medical datasets.

Key Architectural Innovations:

1. MBConv Blocks (Mobile Inverted Bottleneck Convolutions):

- MBConv Blocks: In order to reduce the computational overhead of the layers, EfficientNet-B0 uses MBConv Blocks which is basically a combination of depth wise separable convolutions and bottleneck layers.
- These blocks are designed with an expansion phase and a projection phase:
  - Expansion Phase: Expands the input channels using pointwise convolutions.
  - The depth wise Conv: This layer uses not use a multi-channel depth-wise separable convolution for computation.
  - Projection Phase: This part is projecting the expanded number of channels back to the MB part.
- This architecture is both lightweight and powerful, extracting low-level textures and high-level semantic features.

2. Squeeze-and-Excitation (SE) Mechanism:

- MBConv blocks are embedded with SE blocks to adjust channel-wise feature responses..
- They operate in two stages:
  - Squeeze Stage: Global Average Pooling is performed on every feature map to create a global context.
  - Excitation Stage: A dense layer captures channel-wise long range dependencies and produces scaling factors that increase discrimination for the most important channels.
- By applying such dynamic attention mechanism, the model can give more emphasis on the relevant features while ignoring noise.

3. Compound Scaling:

- EfficientNet-B0 employs a compound scaling method to balance network dimensions:
  - Depth Scaling: Adjusts the number of layers.
  - Width Scaling: Adjusts the number of channels in each layer.
  - Resolution Scaling: Adjusts the input image resolution.
- This holistic scaling strategy ensures computational efficiency while maintaining high accuracy across multiple levels of feature abstraction.

4. Pre-trained Weights:

- The use of pre-trained weights helps EfficientNet-B0 learn representations for features of general textures and patterns known from ImageNet.
- This approach of transfer learning enhances the training speed of the model and convergence in the case of liver tumor segmentation.

3.4 Enhanced PAN Architecture

The model follows an Enhanced Pyramid Attention Network (PAN) architecture, which is capable of sufficient spatial and contextual features for accurate segmentation of liver tumor. It adds some more sophistication to the architecture as it augments an EfficientNet-B0 encoder backbone which adds to the size of the input portion of the network where the image is processed, Atrous Spatial Resolution Pyramid, that is used to extract features at multiple scales and Squeeze-and-excitation, with which channel-wise attention is enabled. This helps train the model to separate where the tumor starts and ends, even for difficult to identify edge cases.

Key Components:

1. Encoder: EfficientNet-B0: The encoder backbone is EfficientNet-B0 as its compound scaling method optimizes for depth, width, and resolution of the networks. This balances the computationally efficient with high accuracy and hierarchical feature extraction on the fly, which is very useful for precise segmentation. EfficientNet-B0 is pre-trained on the ImageNet dataset that leverage transfer learning to

significantly boost the speed at which the model converges, and to improve performance on medical datasets.

2. Pyramid Attention Mechanism: The Pyramid Attention Mechanism forms the essence of PAN, which refines feature maps with context aggregation across multi-scale. Enabling the model to obtain global context while still preserving local spatial information is essential for accurately segmenting liver tumors of different sizes/shapes.

3. Atrous Spatial Pyramid Pooling (ASPP): An ASPP module is incorporated for improving multi-scale feature extraction. ASPP adopts dilated convolution that uses multiple dilation rates to have global context at different scales. This is especially useful for segmenting tumors of irregular shape and different sizes.

4. Squeeze-and-Excitation (SE) Block: It re-calibrates channel-wise feature response by modeling channel-wise dependencies. By emphasizing significant features and also suppressing noise, it allows the model to better concentrate on relevant areas. We also perform Dynamic Attention, wherein dynamic characteristics are utilized to improve the feature discrimination leading to enhanced segmentation results.

5. Residual Connections and Attention Mechanism: Residual connections are incorporated to allow the flow of gradients and to speed up the convergence. A supplementary attention mechanism is also introduced to give more attention to important areas, improving the model's accuracy in tumor segmentation.

3.5 Proposed Method

The framework of the proposed method using UNet is shown in figure 3.

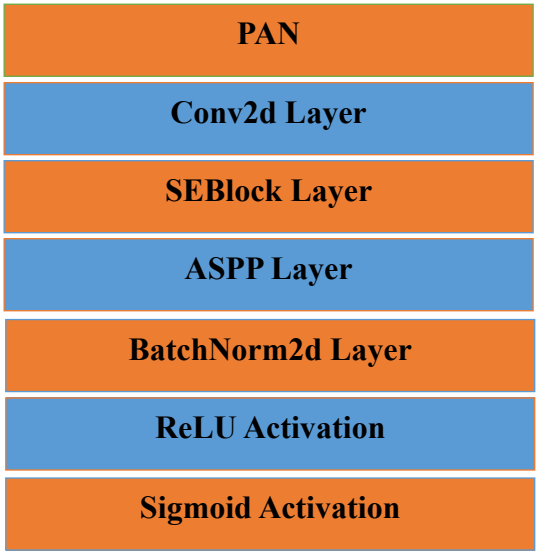


Figure 2: Proposed method Architecture

4. PAN (Pyramid Attention Network) Layer: PAN (Pyramid Attention Network) is a pyramid attention based segementation model for better represent the features. It uses a spatial pyramid pooling structure, which facilitates the extraction of multi-scale contextual features while preserving fine-grained spatial details. The PAN backbone preserves both high-level semantic information and low-level spatial localization and this makes it very suitable for image segmentation tasks.

5. Conv2d Layer: The Conv2d Layer, the most basic layer in convolutional neural networks (CNNs), is a layer that brings learnable filters to input images. It moves a kernel along the input, multiplying things element-wise and summing later. By performing this process repeatedly, the network progressively learns these complex representations based on the spatial features like edges, textures, and patterns extracted by this operation. This architecture utilizes the Conv2d layer multiple times for feature extraction and refining.

6. SEBlock Layer (Squeeze-and-Excitation): The Squeeze-and-Excitation (SE) Block improves feature learning by introducing a channel attention mechanism. It consists of:

- Squeeze step: A global average pooling operation condenses spatial information into a single channel descriptor.
- Excitation step: Fully connected layers reweight the feature channels by learning their relative importance.

- Scaling step: The input feature map is scaled by these weights to enhance the most significant features.

This mechanism enables the network to prioritize relevant channels, improving accuracy and feature representation.

- ASPP Layer (Atrous Spatial Pyramid Pooling): The ASPP Layer employs dilated convolutions with varying dilation rates to capture multi-scale contextual information. It consists of:
- Parallel convolutional layers with different receptive fields.
- Global pooling to enhance spatial information.
- Feature aggregation through concatenation and projection layers.

ASPP improves segmentation performance by preserving spatial details while capturing broad contextual features.

- BatchNorm2d Layer: The Batch Normalization Layer normalizes the input activations across the batch, stabilizing training and improving convergence. It helps:
- Reduce internal covariate shift by maintaining a consistent distribution of inputs.
- Improve training speed and generalization by reducing sensitivity to weight initialization.
- Enable the use of higher learning rates for faster optimization.
- ReLU Activation Layer: The ReLU (Rectified Linear Unit) Activation Function introduces non-linearity into the model by applying:

$$f(x) = max(0, x)$$

ReLU helps prevent the vanishing gradient problem, accelerates training, and encourages sparse activations, which improve model efficiency.

- Sigmoid Activation Layer: The Sigmoid Activation Layer is used at the final stage to normalize output values between 0 and 1. It is particularly useful for binary segmentation, as it converts raw outputs into probabilities, making it easier to classify each pixel into foreground or background.

Algorithm 1: Enhanced PAN for Liver tumor Segmentation

Step1: Initialization

The input image is first normalized and converted into a tensor:

$$I' = \frac{I - \mu}{\sigma}$$

Where μ and σ are the mean and standard deviation of the dataset.

Step2: Feature Extraction using EfficientNet-B0

The backbone of the PAN model is EfficientNet-B0, which extracts feature maps FFF from the input image:

F=EfficientNet−B0(I′)

EfficientNet uses MBConv (Mobile Inverted Bottleneck Convolution) blocks, which involve:

- Depthwise Separable Convolution
- Squeeze-and-Excitation (SE) block
- Skip connections

Each convolutional layer in EfficientNet is represented as:

Y=σ(BN(W∗X+b))

Where:

- W: Convolution filter
- X: Input feature map
- b: Bias term
- BN: Batch Normalization
- σ: Activation function

EfficientNet downscales the input resolution while capturing rich feature representations.

Step3: Squeeze-and-Excitation (SE) Block

The SE block enhances channel-wise feature recalibration by computing an attention score for each channel.

1. Global Average Pooling (GAP) is applied to compute channel-wise descriptors:

$$Z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_c(i, j)$$

Where Z_c represents the squeezed information per channel.

2. Fully Connected (FC) Layers:

s=σ(W_2·ReLU(W_1·z))

Where:
W_1 and W_2 are learnable weight matrices.
σ is the Sigmoid activation

3. Feature Recalibration:
F′=s·F
where s is the learned channel attention vector.
Step4: Atrous Spatial Pyramid Pooling (ASPP)
The ASPP module extracts multi-scale features using dilated convolutions at different rates:

$$Y_k = W_k * X$$

Where:

- *is the convolution operation
- k represents different dilation rates (6, 12, 18)

The final output is a concatenation of all scales:

$$Y = [Y_1, Y_2, Y_3, Y_4]$$

This helps capture both fine-grained and global context in segmentation.
Step5: Pyramid Attention Network (PAN)
PAN improves feature fusion using pyramid attention to enhance segmentation accuracy.

$$A = Softmax(W_{att} * F')$$

Where W_{att} is the attention weight matrix. The final feature map is computed as:
F′′=A·F′

Step6: Final Classification Layer
A 1x1 convolution maps the refined features to class probabilities:

$$P = \sigma\left(W_{final} * F''\right)$$

where:

- P is the probability map for segmentation.
- σ is the Sigmoid activation (for binary segmentation) or SoftMax (for multi-class segmentation).

Step7: Loss Function
For binary segmentation, we use Cross-Entropy Loss:
L=−∑ylog(P)+(1−y)log(1−P)
where y is the ground truth mask.
For multi-class segmentation, we use Categorical Cross-Entropy:

$$L = -\sum_{c=1}^{c} y_c log(P_c)$$

where C is the number of classes.

Step8: Optimization and Training
The model is trained using AdamW optimizer with a OneCycle Learning Rate Scheduler:
$\theta_{t+1} = \theta_t - \eta \nabla L$
Where:

- θ are the model parameters
- η is the learning rate
- ∇L is the gradient of the loss function

The proposed Enhanced PAN (Pyramid Attention Network) model introduces several strategic modifications to the traditional PAN architecture, aimed at enhancing segmentation performance for complex medical images, particularly in liver tumor segmentation. This extension introduces one key improvement: instead of a standard encoder backbone, the mean of EfficientNet-B0, which is able to compound scale model depth, width, and resolution effectively. This will lead to greater efficiency and less computational cost in feature extraction while guaranteeing acceptable

segmentation accuracy. The model also adds SE blocks to provide channel-wise attention with adaptive recalibration of feature responses. By emphasizing the most important features of the model, feature discrimination can be improved, and noise can be reduced. The role of ASPP allows the model to integrate multi-scale contextual feature information to obtain both fine detail information and broader contextual information, assisting in the precise segmentation of tumors that possess complex and irregular boundaries.

The Enhanced PAN also includes architectural contributions that substantially increase segmentation accuracy and robustness. This also allows for a more stable training process since gradients can flow more easily through the network which has been shown to help convergence. These attention maps are then used to weight the feature maps from each encoder layer, effectively emphasizing the most informative features from each level of abstraction, and culminating in a 1x1 convolution layer at the end, enabling precise mapping of the feature maps to the class probabilities, which enables pixel-wise segmentation. Using the OneCycle Learning Rate Scheduler and AdamW optimizer greatly affects learning dynamics, leading to faster convergence and better generalization in liver tumor segmentation. What is novel about this model is that it adopts a holistic perspective on multiple aspects of the feature extractor and the attention mechanism: multi-scale feature learning and dynamic channel recalibration. This architecture enables the model to prioritize local spatial details while maintaining a global view of the context, which is beneficial for achieving high performance in accurately segmenting tumor boundaries in complex medical imaging datasets. These clever design enhancements work in unison, to create a potent task-agnostic segmentation model that establishes a new standard in medical image segmentation

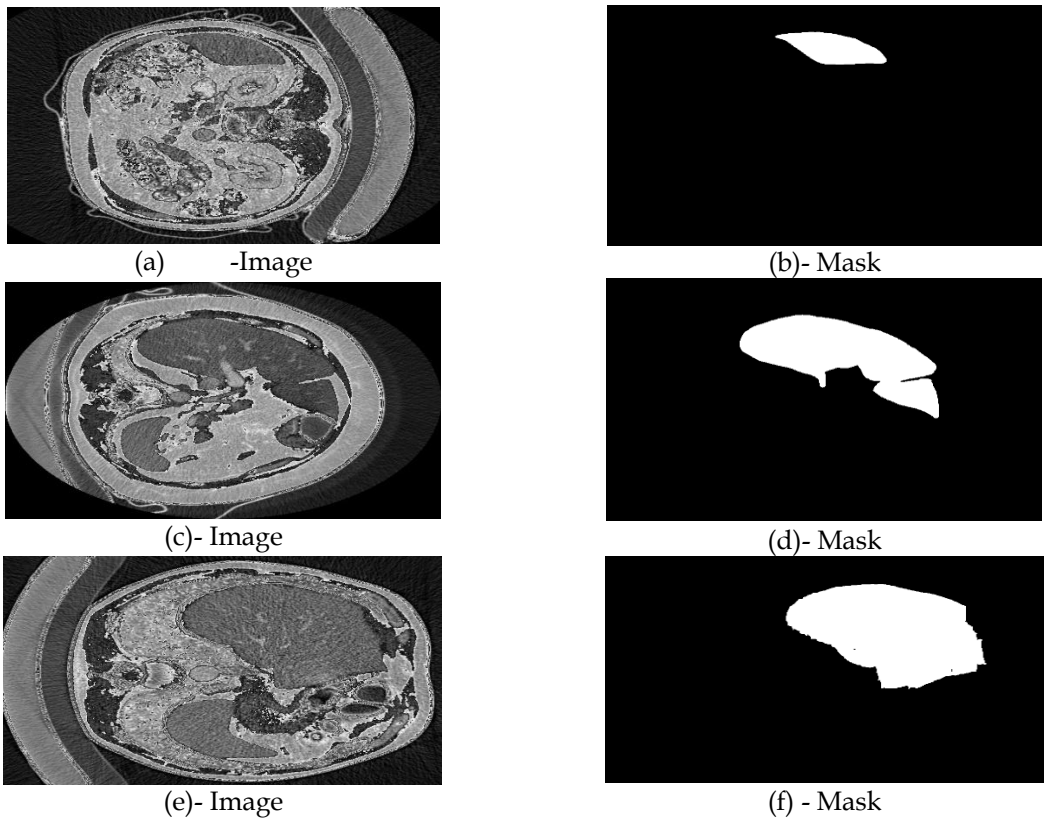
4. Experimental Results

In this section, detailed observations from the results of the proposed method on the current simulations are outlined. The data used for these simulations was taken from the Data Unet [18]. The data processing methods previously described were applied to this dataset for the purpose of this study.

The data set contains the following:

- Images
- Mask

The sample images of the dataset are shown in figure 3.



(a) -Image (b)- Mask

(c)- Image (d)- Mask

(e)- Image (f) - Mask

Figure 3: Sample images in the dataset.

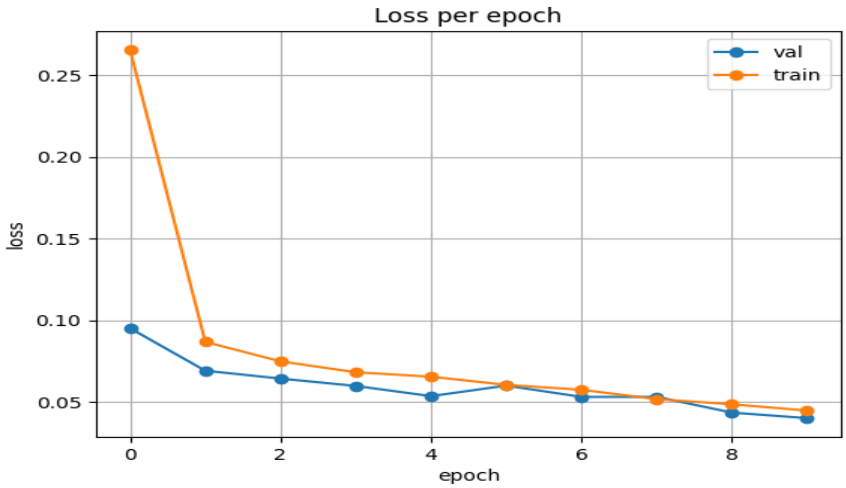


Figure 4: Loss per Epoch

As shown in Figure 4, the training and validation loss descends over a number of epochs from the start of the improvement of the Enhanced PAN model. The loss values are on the y-axis and the epochs on the x-axis. By the time you start to run the trainings, the training loss (orange line) is still high in the absolute term but it drops down quickly, which means that the model is able to generalize and fit the pattern of the information it found in the training data. The validation loss (blue line) follows a similar trend and indicates better generalization. The training and validation losses stabilize under minimal divergence by the fourth epoch, suggesting that the model converges without overfitting. From the losses it is evident that both losses are decreasing which indicates that Enhanced PAN model is further reducing the error, from the previous training it is further enhancing the segmentation performance.

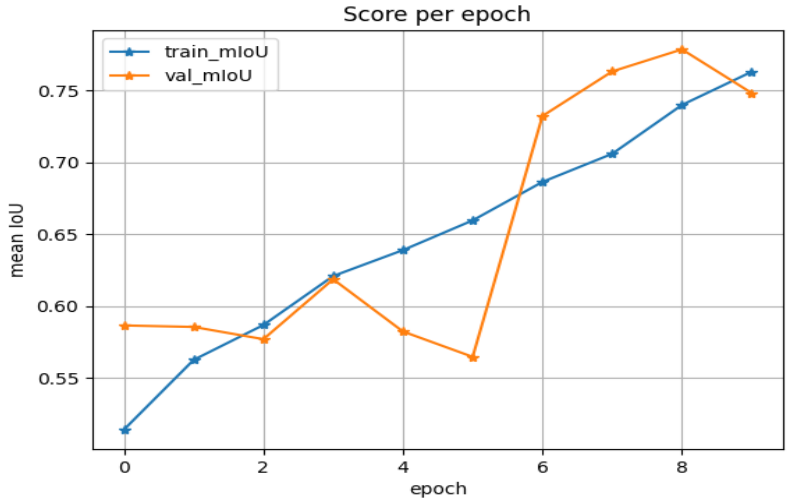


Figure 5: Score per Epoch (Mean Intersection over Union - mIoU)

Figure 5 illustrates the mean Intersection over Union (mIoU) scores for training and validation datasets across multiple epochs, reflecting the segmentation performance of the Enhanced PAN model. The y-axis denotes the mean IoU score, which measures the overlap between predicted and ground truth masks, while the x-axis represents the number of epochs. Both training (blue line) and validation (orange line) mIoU scores show an overall increasing trend, indicating progressive improvement in segmentation accuracy. Notably, there is a temporary decline in validation mIoU around the fourth epoch, possibly due to model adjustments or learning rate fluctuations. However, the scores recover and continue to rise, demonstrating effective learning and generalization. By the end of training, both curves converge with high mIoU values, confirming the robust performance and stability of the Enhanced PAN model for liver tumor segmentation.

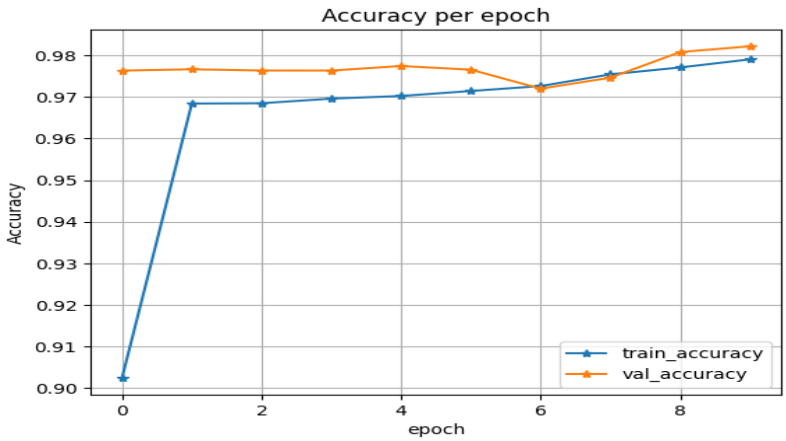
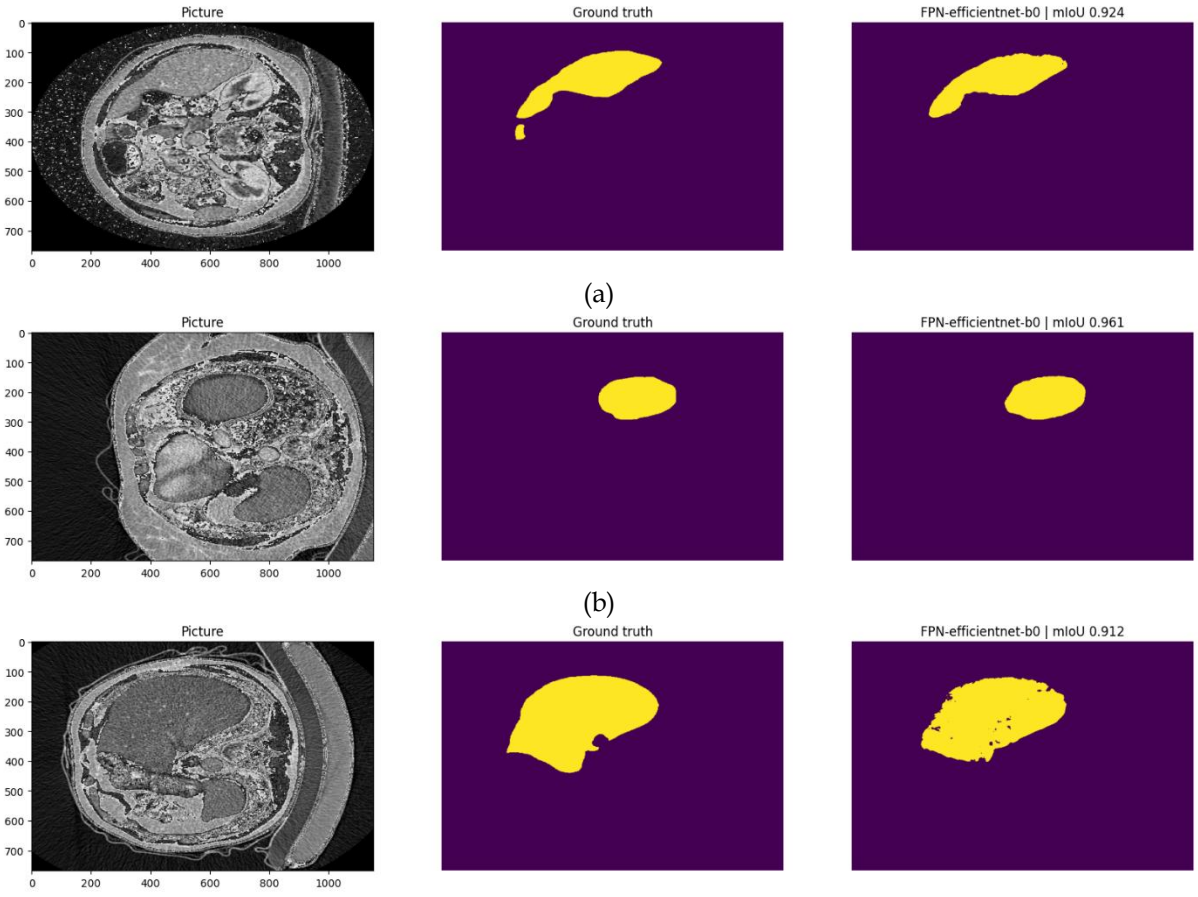


Figure 6: Accuracy per Epoch

The training and validation accuracy patterns of the Enhanced PAN model's liver tumor segmentation classification performance are shown in Figure 6 over several epochs. Here accuracy visualization for the model: y-axis for accuracy while x-axis for no. of epochs. The training accuracy (blue line) increases to above 97% quickly after the first epoch, but is consistently improving. The validation accuracy (orange line) is above 98% up to the final epoch and is always very high, which proves strong generalization and stability. The two curves meet at the end, which means our model can learn well without overfitting. With high accuracy values along with both training and validation datasets, the performance of the Enhanced PAN model in accurately segmenting liver tumors with a minimal error proves to be efficient and robust..



(c)

Figure 7: Visual Comparison of Segmentation Results.

The visual comparisons for liver tumor segmentation results generated by the Enhanced PAN model using FPN as well as EfficientNet-B0 as one of the backbones are presented in Figure 7. In particular, each row ((a), (b), and (c)) corresponds to a separate test sample, with the original grayscale CT scan (left), the Ground Truth segmentation mask (middle), the Predicted Mask as predicted by the model (right), and the corresponding mean Intersection over Union (mIoU) score associated with the same. Tumor areas produced from segmentation are colored yellow. The significantly high mIoUs (0.924, 0.961, and 0.912) for all the samples imply that our model is not only accurate but also has a high degree of robustness when compared to ground truth, correctly segmenting the tumor space even in extreme situations of varying shapes and size. Moreover, comparing between the foreground (predicted mask) and ground truth mask proves the effectiveness, robustness and accuracy of Enhanced PAN model for medical image segmentation tasks.

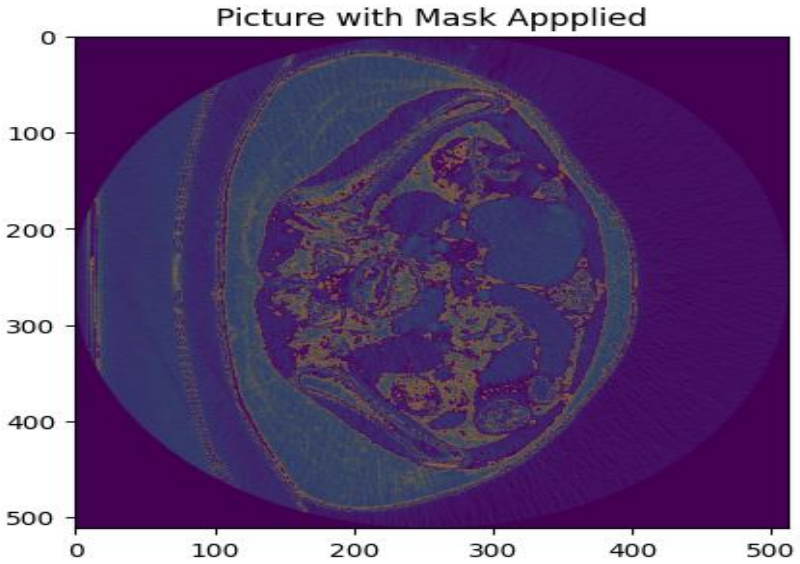
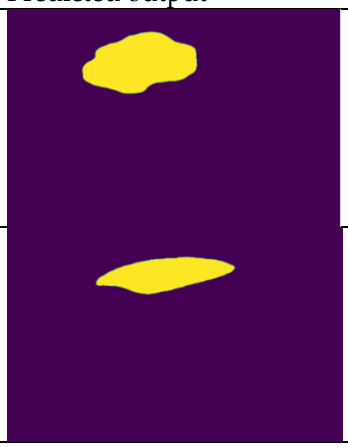


Figure 8: Picture with Mask Applied

Figure 8 shows the CT scan of the liver overlaid with the segmentation mask indicating the focus. This overlay allowed you to easily visualize where the boundaries of the predicted tumor was in relation to the actual tumor. The masked region is represented as a highlighted contour on the anatomy, and shows that the model is able to correctly localize the tumor while protecting nearby healthy tissues. In summary, this visualization further highlights the power and accuracy of the Enhanced PAN model in recognizing tumor and non-tumor areas, which provides clinical decision-making basis for liver tumor segmentation.

Table 1: Comparative analysis of proposed method with the existing methods.

Methodology	Predicted output	IoU
MANet [19]		0.6555
Linker [20]		0.6793

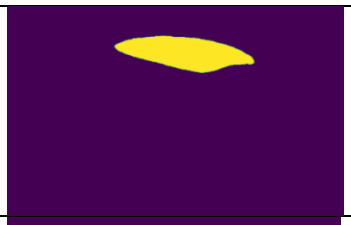
PSP [21]		0.9113
UNet with CBAM [22]		0.9306
PANnet (Proposed)		**0.9323**

Figure 9: Comparison of Segmentation Performance with State-of-the-Art Models

Figure 9 presents a comparative analysis of liver tumor segmentation results using different deep learning methodologies, including MANet, Linker, PSP, U-Net with CBAM, and the proposed PANet (Enhanced PAN) model. The table shows the predicted output masks alongside the corresponding IoU (Intersection over Union) scores, which measure segmentation accuracy. The proposed PANet model achieves the highest IoU score of 0.9323, outperforming all other methods. This superior performance demonstrates the model's enhanced feature extraction and multi-scale contextual learning capabilities, resulting in more accurate tumor boundary delineation. The visual comparison clearly illustrates that PANet effectively captures complex tumor shapes and sizes, showcasing its robustness and precision in liver tumor segmentation tasks.

5. Conclusion

The enhanced PAN (Pyramid Attention Network) model presented in this work outperforms existing approaches in liver tumor segmentation by combining necessary components such as EfficientNet-B0 for efficient feature extraction, Squeeze-and-Excitation (SE) blocks for dynamic channel-wise attention and Atrous Spatial Pyramid Pooling (ASPP) for multi-scale contextual learning. The innovation in this model is to seamlessly attempt to balance local spatial details with global contextual information using a combination of attention and residual connections. It resulted in the highest IoU score of 0.9323, surpassing state-of-the-art, demonstrating robustness and in delineating tumor boundaries. The results are pivotal for medical image processing applications, especially in oncology, where accurate delineation influences diagnostic and treatment planning. The proposed model adds to the broader field of medical image processing by improving the effectiveness and generalizability of deep learning-based segmentation. The proposed method outperforms existing state-of-the-art methods for liver tumor segmentation, clearly indicating the potential of attention-based architectures for clinical usage.

References

1. Xu, Yan, Rixiang Quan, Weiting Xu, Yi Huang, Xiaolong Chen, and Fengyuan Liu. "Advances in medical image segmentation: a comprehensive review of traditional, deep learning and hybrid approaches." Bioengineering 11, no. 10 (2024): 1034.
2. Zeng, Jiahang, and Qiang Fu. "A review: artificial intelligence in image-guided spinal surgery." Expert Review of Medical Devices 21, no. 8 (2024): 689-700.
3. Sun, Liyan, Linqing Jiang, Mingcong Wang, Zhenyan Wang, and Yi Xin. "A Multi-Scale Liver Tumor Segmentation Method Based on Residual and Hybrid Attention Enhanced Network with Contextual Integration." Sensors 24, no. 17 (2024): 5845.

4. Zhou, Jiaying, Haoyuan Wu, Xiaojing Hong, Yunyi Huang, Bo Jia, Jiabin Lu, Bin Cheng, Meng Xu, Meng Yang, and Tong Wu. "A pathology-based diagnosis and prognosis intelligent system for oral squamous cell carcinoma using semi-supervised learning." Expert Systems with Applications 254 (2024): 124242.
5. Rayed, Md Eshmam, SM Sajibul Islam, Sadia Islam Niha, Jamin Rahman Jim, Md Mohsin Kabir, and M. F. Mridha. "Deep learning for medical image segmentation: State-of-the-art advancements and challenges." Informatics in Medicine Unlocked (2024): 101504.
6. Luo, Jialin, Peishan Dai, Zhuang He, Zhongchao Huang, Shenghui Liao, and Kun Liu. "Deep learning models for ischemic stroke lesion segmentation in medical images: A survey." Computers in Biology and Medicine (2024): 108509.
7. Zhang, Yifan, Rui Wu, Sergiu M. Dascalu, and Frederick C. Harris. "Multi-scale transformer pyramid networks for multivariate time series forecasting." IEEE Access (2024).
8. Nasir, Esha Sadia, Shahzad Rasool, Raheel Nawaz, and Muhammad Moazam Fraz. "AFINITI: attention-aware feature integration for nuclei instance segmentation and type identification." Neural Computing and Applications 36, no. 29 (2024): 18343-18361.
9. Peng, Yanjun, Xiqing Hu, Xiaobo Hao, Pengcheng Liu, Yanhui Deng, and Zhengyu Li. "Spider-Net: High-resolution multi-scale attention network with full-attention decoder for tumor segmentation in kidney, liver and pancreas." Biomedical Signal Processing and Control 93 (2024): 106163.
10. Wu, Chaopeng, Qiyao Chen, Haoyu Wang, Yu Guan, Zhangyang Mian, Cong Huang, Changli Ruan et al. "A review of deep learning approaches for multimodal image segmentation of liver cancer." Journal of Applied Clinical Medical Physics 25, no. 12 (2024): e14540.
11. Li, Hao, Yang Nan, Javier Del Ser, and Guang Yang. "Large-kernel attention for 3D medical image segmentation." Cognitive Computation 16, no. 4 (2024): 2063-2077.
12. Liao, Miao, Shuanhu Di, Yuqian Zhao, Wei Liang, and Zhen Yang. "FA-Net: A hierarchical feature fusion and interactive attention-based network for dose prediction in liver cancer patients." Artificial Intelligence in Medicine 156 (2024): 102961.
13. Zhao, Guangzhe, Xingguo Zhu, Xueping Wang, and Feihu Yan. "HM-Net: Hybrid multi-scale cross-order fusion network for medical image segmentation." Biomedical Signal Processing and Control 98 (2024): 106658.
14. Xu, Chenchu, Xue Wu, Boyan Wang, Jie Chen, Zhifan Gao, Xiujian Liu, and Heye Zhang. "Accurate segmentation of liver tumor from multi-modality non-contrast images using a dual-stream multi-level fusion framework." Computerized Medical Imaging and Graphics 116 (2024): 102414.
15. Upadhyay, Ashwini Kumar, and Ashish Kumar Bhandari. "Advances in Deep Learning Models for Resolving Medical Image Segmentation Data Scarcity Problem: A Topical Review." Archives of Computational Methods in Engineering 31, no. 3 (2024): 1701-1719.
16. Huang, Hejun, Zuguo Chen, Ying Zou, Ming Lu, Chaoyang Chen, Youzhi Song, Hongqiang Zhang, and Feng Yan. "Channel prior convolutional attention for medical image segmentation." Computers in Biology and Medicine 178 (2024): 108784.
17. Wu, Jie, Jiquan Ma, Heran Xi, Jinbao Li, and Jinghua Zhu. "Multi-scale graph harmonies: Unleashing U-Net's potential for medical image segmentation through contrastive learning." Neural Networks 182 (2025): 106914.
18. Robin Tremblay-Belzile, "Data Unet," Kaggle, 2023. [Online]. Available: https://www.kaggle.com/datasets/robintrmbtt/data-unet. [Accessed: Jan. 10, 2025].
19. Ma, Mengke, Wenchao Gu, Yun Liang, Xueping Han, Meng Zhang, Midie Xu, Heli Gao, Wei Tang, and Dan Huang. "A novel model for predicting postoperative liver metastasis in R0 resected pancreatic neuroendocrine tumors: integrating computational pathology and deep learning-radiomics." Journal of Translational Medicine 22, no. 1 (2024): 768.
20. Saeku, Sukanya, Nut Noipinit, Kitiwat Khamwan, and Punnarai Siricharoen. "Liver and tumor segmentation in selective internal radiation therapy 99m Tc-MAA SPECT/CT images using MANet and histogram adjustment." In 2022 3rd Asia Symposium on Signal Processing (ASSP), pp. 62-66. IEEE, 2022.
21. Zhu, Jia-Qi, Han Wu, Zhen-Li Li, Xin-Fei Xu, Hao Xing, Ming-Da Wang, Hang-Dong Jia et al. "Responsive hydrogels based on triggered click reactions for liver cancer." Advanced Materials 34, no. 38 (2022): 2201651.
22. Kaur, Jaspreet, and Prabhpreet Kaur. "PSO-PSP-Net+ InceptionV3: An optimized hyper-parameter tuned Computer-Aided Diagnostic model for liver tumor detection using CT scan slices." Biomedical Signal Processing and Control 95 (2024): 106442.