Journal of Intelligent Systems and Internet of Things
Vol. 14, No. 01, PP. 77-89, 2025
DOI: https://doi.org/10.54216/JISIoT.140106

Intelligent Segmentor: Self Supervised Deep Learning based Multi Organ and Tumor Segmentation with Pseudo Lables Generation from CT Images

P. Savitha[1,*], Laxmi Raja[1], R. Santhosh[1]

[1]Department of Computer Science and Engineering, Faculty of Engineering, Karpagam Academy of Higher Education Coimbatore, Tamil Nadu, India

Emails: saviinfo@gmail.com; santhoshrd@gmail.com; laxmirajaphd@gmail.com

Abstract

Multi Organ and tumor segmentation is the challenging task in medical imaging and surgical planning scenarios due to its diverse applications includes lesions and organs measurements and disease diagnosis respectively. Although collecting and examining labels for all classes pose severe challenges. Furthermore, Graphical Processing Unit (GPU) optimization emerge as another critical factor for multi organ and tumor segmentation. To address the mentioned conventional challenge, we designed a deep learning-based model named "Intelligent Segmentor" which performs automated segmentation in end-to-end fashion with novel semi supervised training approach. Initially, the obtained multi organ CT images is then subjected to pre-processing in terms of geometric standardization, noise removal, and intensity normalization respectively. The pre-processed image is then further provided to dual view training for effective Pseudolabel generation. The labelled data along with generated pseudolabels are provided to train the model for amplifying the model performance. After that, there are two inputs are provided to the designed segmentation model which includes dual encoders such as GoogleNet and VGG-16 for contextual and spatial information extraction in five stages, Tweaked Feature Pyramidal Network (TFPN) for dimensionality reduction and side features extraction, and Gated Fusion Module (GFM) for fusing the side features to form unified feature map. Finally, the unified feature map is the examined through convolution layers for multi organ and tumor output. We adopted FLARE 2023 dataset for validating the proposed work with existing works on 13 various organs and tumor segmentation tasks. From the results, the proposed research achieves better Dice Similarity Coefficient (DSC) and Normalized Surface Dice (NSD) through online validation and final testing than the existing works.

Received: February 05, 2024 Revised: April 14, 2024 Accepted: July 02, 2024

Keywords: Multi Organ and Tumor Segmentation; Computed Tomography (CT); Deep Learning (DL); Pseudo Label; Self Supervised

1. Introduction

Medical imaging has really enhanced the healthcare by supporting with the disease diagnosis, treatment planning and disease monitoring by offering the non-invasive visualization of the body's inner structures. Computed tomography (CT) scans are the one of the modalities which is routinely implemented due to they can generate inclusive cross-sectional of body images. It is necessary for identifying the problems in different organs when the explanation and investigation of these images can be difficult [1]. This is the part of rising interest in automated medical image segmentation, particularly with the Deep Learning (DL) approaches. Identifying and crucial different anatomical entities within a single scan is called as Multi organ segmentation from CT images. This procedure is convoluted due to the organ forms, locations, sizes and appearance may diverge between patients and even within the same patient over time [2-3]. DL suggests an attractive replacement for traditional image processing approaches that routinely establishes in adequate in handling these fluctuations. CNN is a subset of DL that has altered the regulation by providing cutting edge performance in image analysis purpose including medical image segmentation [4]. DL approaches like U Net and its difference have established surprising efficiency while it comes to segmenting the medical images. Usually, these approaches are collected of an encoder-decoder framework which amenities the recreation of complex segmentation maps and the removal of hierarchical features [5-7]. The encoder stores the contextual information at various scales when the decoder recreates the accurate spatial

features necessary for perfect segmentation. In the province of the medical image investigation, both local information and global background are essential and the framework is particularly well-suited [8-10]. The shortage of explained databases is one of the main difficulties in multi organ segmentation. The labor exhaustive and dedicated nature of manual explanation constructs gathering huge, high quality databases demanding. This constraint has prompted the study into transfer learning, data augmentation and semi supervised and unsupervised learning techniques [11-13]. The variability in CT image quality and resolution that can be conveys about the differences in imaging equipment and approaches, suggests alternative issues. They must be opposing to these changes for DL approaches to be valuable in real clinical situations. Techniques like multi scale feature extraction, hybrid designs which merge CNN with other types of neural networks to enhance the model effectiveness and flexibility such as RNN and the tradition of the attention procedures have all been examined [14-16]. The metrics are routinely employed to estimate the precision and dependability of segmentation models that weight the importance of evaluating segmentation performance including Dice coefficient, IoU and Hausdroff distance [17]. Moreover, to assurance that these approaches may assure the real necessities of medical diagnosis and treatment preparation, clinical confirmation over the relationship with the healthcare practitioners is essential. Multi organ segmentation is still not easy even though remarkable advances in image processing and Machine Learning (ML) due to an amount of intrinsic difficulties [18-19].

The challenges including Organ size and shape diversity, differences in the quality of imaging modalities, the immediacy and overlap of organs, a shortage of annotated information and computational complexity [20]. Therefore, the improvement of a precise multi organ segmentation models is directly required. To address these issues, we have presented a new multi organ and tumor segmentation model known as Intelligent Segmentor. The presented approach initially passes over a number of preprocessing steps. Next to the preprocessing steps, the unlabelled information is employed to create pseudo labels for the images. The input images and these offered pseudo labels merge to construct a refined region of Interest (RoI) image. We also provided the input in two outlooks while our model achieves both multi organ and tumor segmentation such as RoI images and compressed RoI images. The compressed images are feed into the VGG-16 backbone model when the multi organ input is nourishing to GoogleNet backbone model. Additionally, in five stages the two backbone models take out the spatial and contextual data. And then, a Tweaked Feature Pyramid Network (TFPN) is implemented to perform superior grained feature extraction on the recovered features during the first five stages. Fine grained feature maps are offered by this approach for the multi organ and tumor segmentation. Then, a Gated Fusion Module (GFM) is employed to fuse these offered feature maps onto a single feature map. Additionally, we have applied Convolutional layers methods to the combined feature map to get outputs which are an accurate for both multi organ and tumor segmentation. This new approach seeks to produce a more accurate and dependable solution. The major contribution of this work are provided as below,

- For ensuring better segmentation precision and reduced complexity, we perform three pre-processing techniques such as geometric standardization, noise removal, and intensity normalization. Whereas many of the mentioned state of the art works performs only limited pre-processing steps which hinders the segmentation efficacy.
- We have generated better Pseudolabel by performing dual view training on the designed Segmentor model in semi supervised way to enhance the model performance. The state of the art works lacks in training process thereby affecting the model performance.
- We design a novel multi organ and tumor Segmentor model named Intelligent Segmentor which performs automated segmentation by adopting two image inputs (i.e. RoI enabled and Compressed RoI images) for multi organ and tumor segmentation respectively with two encoders and fusion modules respectively.

2. Related works

In this paper [21] authors proposed a novel AdaptNet framework for Abdomen Multi-Organ and Tumor Segmentation using partial labeled data. This proposed method comprises several key components including segmentation network, pseudo-label generation network and adaptive controller which were useful for generating dynamic weight. Authors in this paper [22] proposes a Conditional Dynamic Attention Network for multi-tumor and organ segmentation. This method incorporates with a multisource parameter generator which fuses the conditional and multiscale data to determine within various tasks. Then, a dynamic attention module has been used for increases focus on task related aspects. This paper authors [23] develops an approach by utilizes pseudo-labeling in a semi-supervised two-stage model to improve segmentation accuracy. The model was trained on pseudo-labeling to reduce GPU memory usage by implementing a two-stage strategy. At first training a abdomen location model to recognize the abdominal area and then restricting segmentation to this localized area. In this paper authors [24] present a method by employs a diffusion model to facilitate label-efficient segmentation. This denoising diffusion method was trained on a unlabeled CT data to create 2D images. Then the model was capable for segmentation which requires minimal labelled data. Moreover, a linear classification and fine-tuning decoder has been utilized to improve segmentation performance. Authors in this paper [25] proposes an Attention mechanism-based deep supervision network for multi-organ segmentation. This framework utilizes a U-Net architecture based on a lightweight VGG network through attention mechanisms and enables quick model convergence without pre-training. Furthermore,

multi-stage deep supervision improves feature extraction for segmenting tumors and organs. In this paper author [26] introduces a two-stage method by utilizing 2D techniques for 3D segmentation task. This method leverages state-of-the-art video object segmentation method and semi-supervised training to use massive unlabeled data. In addition, we develop a method to generate both plausible and consistent pseudo-labeled data for retraining by utilizing an uncertainty estimation. This paper authors [27] presents a DoDNet framework for multi-organ and tumors segmentation through partially labelled datasets. It contains a shared encoder-decoder architecture, task encoding module, controller for dynamic filter generation, dynamic segmentation head. In this paper authors [28] develops a novel framework for multi-organ and tumor segmentation. This framework incorporates with minimal computational overhead for continuous organ and tumor segmentation. In addition, it replaces the conventional output layer with lightweight, class-specific heads, enabling independent predictions and adapting new classes without disrupting existing ones. Authors in this paper [29] proposes a Transformer-based dynamic on-demand a network for multi-organ and tumor segmentation through labeled data. This method comprises a hybrid backbone of CNN and Transformer through dynamic head that adaptively creates kernels by utilizing a transformer self-attention mechanism for long-range dependencies and decode organ embeddings. Authors in this paper [30] presents a Re-parameterizing Mixture-of-Diverse-Experts for abdominal organ and tumor segmentation. This method contains Mode block in RepMode which learns generalized parameters for all tasks and integrates into a U-shaped network through dynamic heads addressing multi-scale challenges.

This paper authors [31] proposes a network which utilize an adaptive attention mechanism across layers to improve 3D segmentation from CT images. This method uses a cross-layer connections to reuse semantic and spatial data which includes pre-activation residual blocks and squeezing and excitation blocks. The paper authors [32] proposes an incremental learning approach designed for multi-organ segmentation. This method develops the model by new organ annotations though losing access to past ones which rely on the present model to hold previous data. Researchers believe that various data distributions generate "catastrophic forgetting" in IL techniques and confirms ILs suitability for medical imaging. This paper authors [33] proposes a method by leveraging a federated learning to jointly segment multiple organs and tumors from partially labeled data. This method uses a multi-task segmentation technique by utilizing FL to leverage independent datasets through various annotations. In this paper authors [34] develops a multi-task learning with iterative training for multi-organ and tumor segmentation. Furthermore, base models were trained to generate initial pseudo-labels in a fully supervised setting. Moreover, by incorporating multi-task learning a model has been trained to segment organs and tumors. Then using a sliding window strategy and a simplified test-time augmentation to balance segmentation. Authors in [35] paper proposes a Task-Guided Network Architecture for multi-organ and tumor segmentation through labelled datasets. This model uses a residual block and attention module which fuse images features through task encoding constraints, highlighting task-specific one and suppressing unrelated features. Authors in this [36] paper develops a MS-KD framework for multi-organ segmentation through Multiple Binary-labeled datasets. This framework uses a region-based supervision with each provides logits-wise and feature-wise supervision in their organ region and background region. This paper authors in [37] proposed a nnU-net for multi-organ segmentation using labelled datasets. This proposed method uses the nnU-Net as a backbone and uses a strategy which feeds auxiliary data for organ identification and spatial data recovery. Moreover, Deep supervision and uses both dice and focal loss optimize the model. The authors in [38] paper develops novel Cross Pseudo-based Semi-Supervised Learning for abdominal multi-organ segmentation. It incorporates small amount of labelled CT images and huge amount of unlabeled data. Furthermore, it utilizes an auxiliary branch with organ shape constraints and redesigned loss function to enhance segmentation accuracy. In this paper authors [39] presents a method which combines pseudo labels with a two-stage segmentation to improve the accuracy for abdominal multi-organ segmentation. Additionally, three nnU-Nets has been used first one generates high-quality pseudo labels, second one coarse segmentation to guide cropping and lastly for achieving effective segmentation. This paper authors [40] develops a PWS-Seg method for abdomen organ and tumor segmentation. This PWS-Seg technique uses a progressive method through partly annotated image and class-wise label fusion for reliable pseudo labels in order to achieve higher segmentation accuracy.

3. Methodology

Figure 1 represents the overall architecture of the proposed Intelligent Segmentor model. The designed Intelligent Segmentor composed of two different encoders for processing the multi organ and tumor images respectively. Furthermore, the Tweaked Feature Pyramidal Network (TFPN) is utilized for feature dimensionality reducing and side feature extraction. The extracted side features from both side of encoders are fused using Gated Fusion Module (GFM) to form unified feature map. The unified feature is then subjected to process in convolution layers to get tumor and multi organ output.

A. Pre-Processing

(a) Geometric Standardization

Geometric standardization (GS) is utilized for ensuring the attained multi organ CT images with similar spatial resolution and voxel size respectively. The GS composed of two supplementary methods such as resampling, and cropping & padding respectively.

(i) Resampling: The major objective of resampling processes to guaranteeing the all the CT multi organ images have similar resolution and voxel size. Interpolation method is utilized for firmly adjusting the image voxel grid. Initially, for all the dissimilar CT images we determine the new voxel size ($\mathrm{i.e.}\,\mathrm{Vox}_{\mathrm{new}} = 1.5\mathrm{mm} \times 1.5\mathrm{mm} \times 1.5\mathrm{mm}$). Once determined, the scaling factor is computed for every dimension as,

$$\mathrm{Sc}^{\mathrm{x}} = \frac{\mathrm{Vox}^{\mathrm{x}},\mathrm{original}}{\mathrm{Vox}^{\mathrm{x}},\mathrm{new}};\ \mathrm{Sc}^{\mathrm{y}} = \frac{\mathrm{Vox}^{\mathrm{y}},\mathrm{original}}{\mathrm{Vox}^{\mathrm{y}},\mathrm{new}};\ \mathrm{Sc}^{\mathrm{z}} = \frac{\mathrm{Vox}^{\mathrm{z}},\mathrm{original}}{\mathrm{Vox}^{\mathrm{z}},\mathrm{new}} \quad (1)$$

Where, $\mathrm{Sc}^{\mathrm{x}}, \mathrm{Sc}^{\mathrm{y}}, \mathrm{and}\ \mathrm{Sc}^{\mathrm{z}}$ denotes the scaling factor of CT image of every dimension. After computing the scaling factors, resampling is done by applying interpolation method as,

$$\mathrm{Img}_{\mathrm{RS}}(\mathrm{x}', \mathrm{y}', \mathrm{z}') = \mathrm{Img}\left(\frac{\mathrm{x}'}{\mathrm{Sc}^{\mathrm{x}}}, \frac{\mathrm{y}'}{\mathrm{Sc}^{\mathrm{y}}}, \frac{\mathrm{z}'}{\mathrm{Sc}^{\mathrm{z}}}\right) \quad (2)$$

Where, img denotes the original CT multi organ images from FLARE 22 dataset, and $\mathrm{Img}_{\mathrm{RS}}$ denotes the resampled images with co-ordinates $(\mathrm{x}', \mathrm{y}', \mathrm{z}')$.

(ii) Cropping and Padding: The cropping and padding is mainly utilized for focusing the specific Region of Interest (RoI) for guaranteeing the reliable image dimensions. To be clearer, the cropping the image induce only the desired regions and pad them to zero. For cropping & padding, initially the relevant bounding box of the organs anatomy can be determined. After that, the image can be cropped to the bounding box as,

$$\mathrm{Img}_{\mathrm{crop}} = \mathrm{Img}[\mathrm{x}^{\mathrm{mini}}:\mathrm{x}^{\mathrm{maxi}}, \mathrm{y}^{\mathrm{mini}}:\mathrm{y}^{\mathrm{maxi}}, \mathrm{z}^{\mathrm{mini}}:\mathrm{z}^{\mathrm{maxi}}] \quad (3)$$

Finally, the cropped image can be padded to standardized size as,

$$\mathrm{Img}_{\mathrm{padd}} = \mathrm{Pad}(\mathrm{Img}_{\mathrm{crop}}, \mathrm{tar}_{\mathrm{size}}) \quad (4)$$

Where, $\mathrm{Img}_{\mathrm{padd}}$ is the padded image after cropping, and $\mathrm{tar}_{\mathrm{size}}$ denotes the desired size needed to be padded. To be more distinctive, if the desired RoI is spleen which is located among CT image slices 40 to 90, and columns 150 to 250 the $\mathrm{Img}_{\mathrm{crop}}$ can be computed as,

$$\mathrm{Img}_{\mathrm{crop}} = \mathrm{Img}[40:90, 150:250, :] \quad (5)$$

For padding, if the size of $\mathrm{tar}_{\mathrm{size}}$ is set by $512 \times 512 \times 512$, and the size of cropped image is $40 \times 150 \times 512$ then the CT image can be pad to fit in the target size.

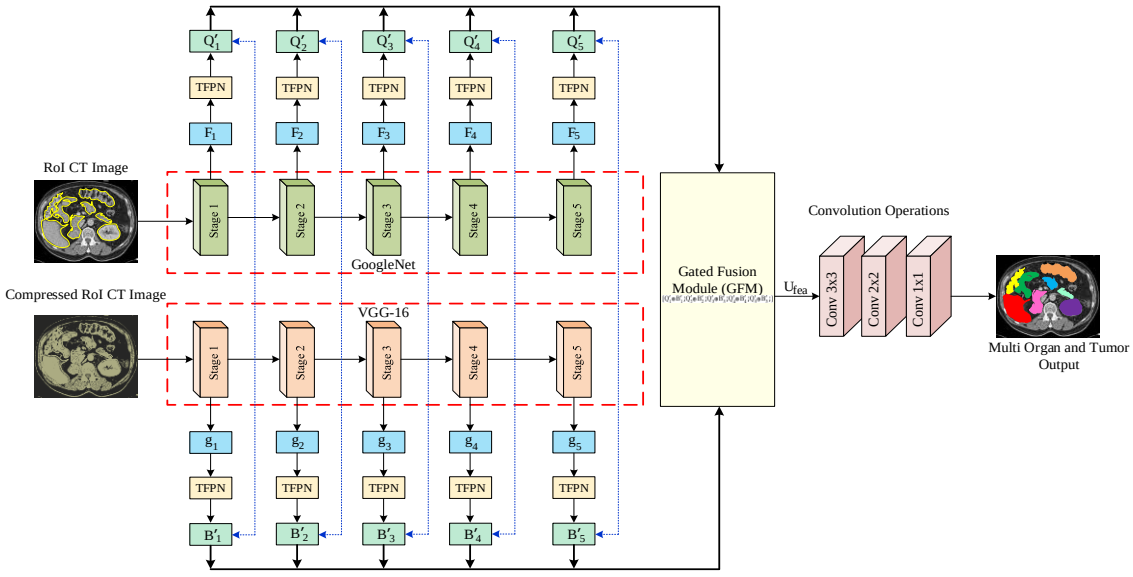


Figure1. Illustration of Proposed Intelligent Segmentor Model

(b) Noise Removal

The standardized images are then fed for noise removal where the salt and pepper noise in the CT image can be removed to preserve the image edges. For noise removal, we adopt medial filtering technique in which every pixel value can be replaced with the median intensity value.

Initially the neighbourhood pixel location can be defined in which the pixel location of the CT multi organ image Img is defined by (j, i). The square kernel window with size $k \times k$ is positioned around the pixel. To be clearer, the size of window k is set odd $(i.e.\ 5 \times 5, 7 \times 7)$ to guaranteeing that the pixel is in center. Let $Wi(j, i)$ denotes the pixel intensity set on a window that centered at (j, i) as,

$$Win(j,i) = \left\{Img(x,y) | j - \frac{k-1}{2} \leq x \leq j + \frac{k-1}{2}, i - \frac{k-1}{2} \leq y \leq i + \frac{k-1}{2}\right\} \quad (6)$$

After that the window pixel intensities $Win(j, i)$ can be sorted in ascending order as,

$$Sort\big(Win(j,i)\big) = \{Img_1, Img_2, \dots, Img_{k\times k}\}, \quad Img_1 \leq Img_2 \leq \cdots \leq Img_{k\times k} \quad (7)$$

From the sorted windows, the median value is obtained. Since the window size is odd, the formulation of median is defined at the position $\frac{k\times k+1}{2}$ as,

$$Med\big(Win(j,i)\big) = Img_{\frac{k\times k+1}{2}} \quad (8)$$

Finally, the actual pixel intensity $Img(j, i)$ is replaced by the obtained median pixel and that can be formulated as,

$$Img'(j,i) = Med(Win(j,i)) \quad (9)$$

Where, Img' denotes the salt and pepper removed image.

(c) Intensity Normalization

Intensity normalization is utilized for ensuring dependable intensity values on multi organ CT noise removed image. For intensity normalization, we utilize Z-score normalization which helps the noise removed images to be eradicate intensity ranges and image contrasts. For the noise removed image Img' with mean and standard deviation of the intensity values can be formulated as,

$$\aleph = \frac{1}{M}\sum_{j=1}^{M} Img_j \quad (10)$$

$$\forall = \sqrt{\frac{1}{M}\sum_{j=1}^{M}(Img_j - Img_j)^2} \quad (11)$$

From the above equation, the pixels numbers in the Img' is denoted by M, and Img_j is the intensity value of the j-th pixel. Finally, the intensity values can be normalized as,

$$Img_{Norm}(j) = \frac{Img_j - \aleph}{\forall} \quad (12)$$

From the above equation, $Img_{Norm}(j)$ denotes the normalized intensity of the j-th image pixel.

B. Dual View Training & Pseudolabel Generation

Once pre-processed, we perform dual view training on multi organ CT images for ensuring learning consistency during segmentation. Note that the dual view training is semi supervised technique that influences augmentation on similar input for generating efficient pseudo labels. The dual view training is initialized by splitting the adopted dataset into unlabelled (Da_{UL}) and labelled (Da_{La}) set respectively. After that the model (Seg_{Mo}) for brain tumor segmentation is initialized for training with Da_{La} from supervised learning method.

For the attained unlabelled images $y \in Da_{UL}$, we generate double augmented forms as y_1 & y_2 respectively. The augmented forms mentioned such as elastic deformation, flips, rotations, etc..,. The augmented image views y_1 & y_2 are passed to the Seg_{Mo} to get segmented predictions $\hat{x}_1$ & $\hat{x}_2$ respectively. By combining $\hat{x}_1 \oplus \hat{x}_2$ using majority voting technique, Pseudolabel can be generated as $\hat{x}$. In order to enhance the quality of datasets, we perform consistency regularization which ensures reliability among $\hat{x}_1$ & $\hat{x}_2$ by reducing the reliability loss Lo_{re}. The Lo_{re} is computed by calculating Mean Square Error (MSE) among the model predictions and can be formulated as,

$$Lo_{re} = MSE(\hat{x}_1, \hat{x}_2) \quad (13)$$

The loss function guarantees model to outputs same outputs even though for varied augmentation on similar image. Finally, we perform training on pseudo labels by combining the reliability loss with Da_{UL}, and labelled loss with Da_{La} as,

$$Lo = Lo_{La} + \varphi Lo_{re} \quad (14)$$

From the above equation, φ is the balanced hyper parameter that subsidizes reliability during model training. The training process can be performed iteratively for robust model training. For providing input to the trained $\mathrm{Seg_{Mo}}$ with pseudo labels, we provide two RoI views (i.e. Normal image with RoI $\mathrm{N_{img}}$ and compressed RoI image $\mathrm{C_{img}}$)

C. Intelligent Segmentor Model

The $\mathrm{N_{img}}$ and $\mathrm{C_{img}}$ is passed to the designed Intelligent Segmentor model that consist of double independent convolution networks for extracting multi-level contextual and spatial features. There are two different encoders are utilized named GoogleNet and VGG-16 were utilized with five stages are backbone encoder respectively.

The GoogleNet is utilized for processing the $\mathrm{N_{img}}$ for multi organ segmentation which generates five features maps with increasing channel dimensions and decreasing spatial sizes and ensures dimensionality reduction, robust generalization, flexibility and scalability respectively. Furthermore, the VGG-16 is utilized for processing the $\mathrm{C_{img}}$ for tumor segmentation that also generates five feature maps by ensuring effective receptive field, better depth information, and uniformity respectively.

Initially, the $\mathrm{N_{img}}$ is provided as an input to the GoogleNet with dimensions $\mathrm{H \times W \times C}$ in which the input is passed to five stages of feature map extraction as $\{\mathrm{F_1, F_2, F_3, F_4, F_5}\}$. In similar manner, the $\mathrm{C_{img}}$ also passed to the VGG-16 with compressed dimensions $\mathrm{Com(H \times W \times C)}$ that also produces five stages of feature maps as $\{\mathrm{g_1, g_2, g_3, g_4, g_5}\}$. For GoogleNet, the first stage has decreased in spatial size and increased in channels due to stride 3 convolution and max pooling whereas the consecutive stages (2,3,4, and 5) have increasing number of channels with reduced spatial size due to inception module. For VGG-16, the spatial sizes are decreased in all stages with increase in channels (64,128,256,512,512). To be clearer, the feature maps are decreased in spatial sizes with the multiples of $3\{{}^{1}/_{3}, {}^{1}/_{6}, {}^{1}/_{9}, {}^{1}/_{11}, \text{and } {}^{1}/_{15}\}$.

The obtained $\{\mathrm{F_1, F_2, F_3, F_4, F_5}\}$ and $\{\mathrm{g_1, g_2, g_3, g_4, g_5}\}$ are then provided to the Tweaked Feature Pyramidal Network (TFPN) for extracting compressed side features for reducing the computation cost. Let $\mathrm{F_j}$ and $\mathrm{g_j}$be the feature map from the j-th stage of double backbone network respectively. The TFPN composed of bottom-up pathway, top-down pathway with lateral connections, and convolutional layers. The bottom-up pathway composed of features from the dual backbone networks (i.e. GoogleNet and VGG-16). The features are provided as an input to the top-down pathway with lateral connections by upsampling and combining the higher and lower resolution feature maps. The mathematical steps involved in the top-down pathway are provided as follows,

$$\mathrm{Q_5 = 2DConv(F_5, filter = 512, K_S = 2 \times 2, stride = 2, padd = same)} \quad (15)$$

$$\mathrm{Q_4 = ADD\big(UPS(P_5, size = 3, IP = Near), 2DConv(F_4, filter = 512, K_S = 2 \times 2, stride = 2, padd = same)\big)} \quad (16)$$

$$\mathrm{Q_3 = ADD\big(UPS(P_4, size = 3, IP = Near), 2DConv(F_3, filter = 512, K_S = 2 \times 2, stride = 2, padd = same)\big)} \quad (17)$$

$$\mathrm{Q_2 = ADD\big(UPS(P_3, size = 3, IP = Near), 2DConv(F_2, filter = 512, K_S = 2 \times 2, stride = 2, padd = same)\big)} \quad (18)$$

$$\mathrm{Q_1 = ADD\big(UPS(P_2, size = 3, IP = Near), 2DConv(F_1, filter = 512, K_S = 2 \times 2, stride = 2, padd = same)\big)} \quad (19)$$

Finally, the side features are obtained by processing the generated pyramidal feature maps $(\mathrm{Q_1, Q_2, Q_3, Q_4, Q_5})$ with the additional convolutional layers. The corresponding mathematical steps are provided as below,

$$\mathrm{Q_1' = 2DConv(Q_1, filter = 256, K_S = 4 \times 4, strid = 2, padd = same)} \quad (20)$$

$$\mathrm{Q_2' = 2DConv(Q_2, filter = 256, K_S = 4 \times 4, strid = 2, padd = same)} \quad (21)$$

$$\mathrm{Q_3' = 2DConv(Q_3, filter = 256, K_S = 4 \times 4, strid = 2, padd = same)} \quad (22)$$

$$\mathrm{Q_4' = 2DConv(Q_4, filter = 256, K_S = 4 \times 4, strid = 2, padd = same)} \quad (23)$$

$$\mathrm{Q_5' = 2DConv(Q_5, filter = 256, K_S = 4 \times 4, strid = 2, padd = same)} \quad (24)$$

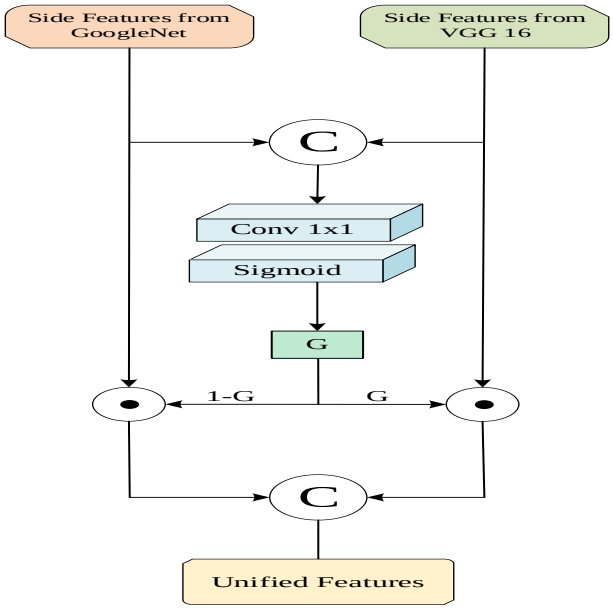


Figure 2. Illustration of GFM

From the above equations, K_S denotes the kernel size, strid denotes the stride value, padd denotes padding, and 2DConv denotes the convolution operation. Furthermore, ADD , UPS and IP denotes the adding, upsampling, and interpolation operation respectively. Therefore, the attained side feature maps from the GoogleNet are obtained as $\{Q_1', Q_2', Q_3', Q_4', Q_5'\}$. In similar manner the side features attained for the VGG-16 is obtained as $\{B_1', B_2', B_3', B_4', B_5'\}$. Note that, the computations for the attaining VGG-16 side features are similar as GoogleNet side features computations.

The side features $\{Q_1', Q_2', Q_3', Q_4', Q_5'\}$ and $\{B_1', B_2', B_3', B_4', B_5'\}$ from both the TFPN are amalgamated by the Gated Fusion Module (GFM). The GFM combines them by considering the similar spatial size and channels respectively. Figure 2 illustrates the GFM. Within the GFM, every side feature maps are amalgamated to form unified feature map as $U_{fea} = [Q_1' \oplus B_1';\ Q_2' \oplus B_2';\ Q_3' \oplus B_3';\ Q_4' \oplus B_4';\ Q_5' \oplus B_5']$ with 2×2 convolution is applied for building cross modal relationship among the side feature inputs. To be more clear, sigmoid layer is adopted for getting weight matrix Z for Q_j' and B_j' respectively by performing element wise multiplication. Finally, the weighted feature is amalgamated to form final unified feature map U_{fea}. Finally, the U_{fea} is processed through three convolutional layers with reduced kernel size as $\text{Conv } 3 \times 3, \text{Conv } 2 \times 2, \text{and Conv } 1 \times 1$ to precise multi organ and tumor output.

4. Numerical results and discussion

A. Dataset and evaluation measures

A development of the FLARE 2021-2022 struggle and the FLARE 2023 database disputes that focused to advance the improvement of charity models in the study of abdominal diseases. The segmentation goals consist of 13 organs such as spleen, liver, pancreas, right and left kidneys, stomach, gallbladder, esophagus, right and left adrenal glands, IVC and duodenum with a selection of a abdominal lesion which consist of various types of abdominal cancer and consisting those which involve the liver, pancreas, gastric, kidneys and additional organs. There are 4000 CT image of the abdomen in the training set, 2200 images have biased labels and 1800 have none. There are 100 and 400 CT scans in the testing and confirmation sets. Unlabeled images were not utilized in this research. Only 2200 scans with biased labels have been implemented because of the scarcity of computer possessions and 1800 unlabeled images have not been used. Table 1 suggest the organ and tumor explanation frequency information for the 2200 patients. Fivefold cross validation has been taken with every fold's internal validation dataset including 440 instances and the training dataset with 1760 cases.

Table 1: Tumor and Organ Frequency Representation

Target frequency	Liver 59.7	Pancreas 59.7	RK 59.2	Spleen 59.5	RAG 11.2	Aorta 11.2	IVC 11.2
Target frequency	Gallbladder 10.1	Stomach 11.2	Duodenum 11.2	LK 59.1	Tumor 67.0	Esophagus 11.2	LAG 11.1

The evaluation criteria contain two efficiency measurements including running time and area under the GPU memory time curve to two accuracy measures such as Dice similarity coefficient or DSC and Normalized surface DICE or NSD. All of these metrics insert up to the ranking computation. Moreover, the GPR memory consumption and running time area are carried out into the account surrounded by tolerances of 4 GB and 15 seconds.

B. Implementation details

The system operates Ubuntu 23.4 with an Intel (R) Core ™ i9-10900Xcpu noticed at 3.70GHz. It has a large number of memories with 4x32GB of RAM which operates at 2933MT/s. It employs CUDA version 12.0 for accelerated computing activities and an NVIDIA Geforce RTXTM 3090 GPU with 24GB of RAM for graphic processing. Python 3.13 is the main programming language utilized and the system constructs utilize of the PyTorch architecture more especially, Torch 2.01 for DL purposes.

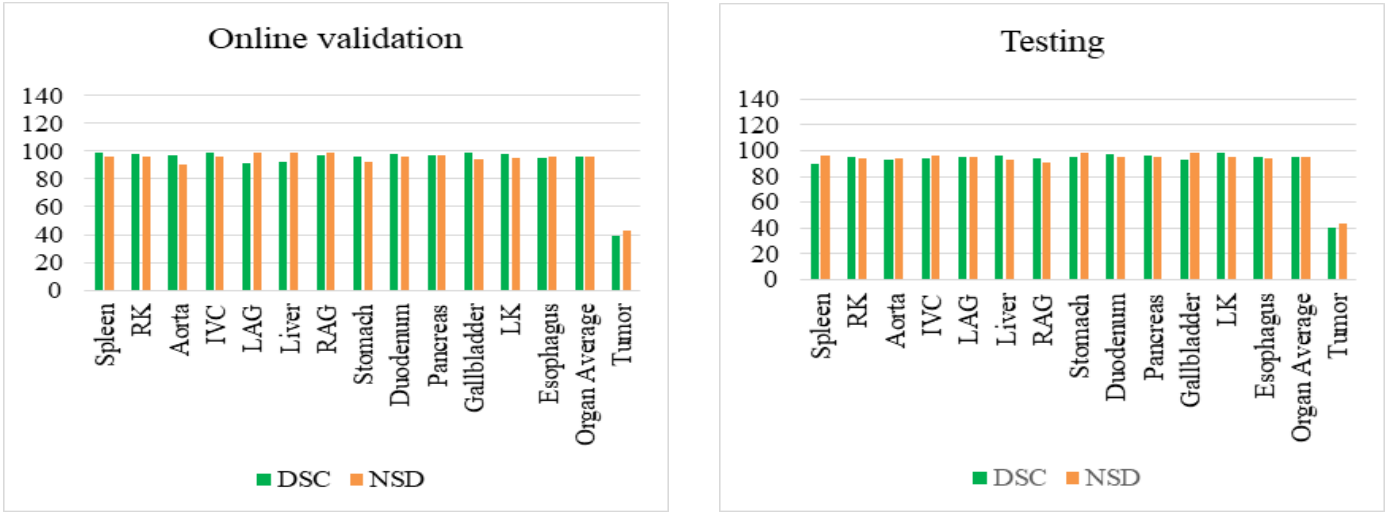
Instructional procedure we selected all of the samples within every epoch at random and set the batch size to 2 throughout the training phase. We also implemented the random patch cropping with patch sizes of (96,128,160) for each sample. We employ AdamW as the optimizer that has a weight decay of 1e-5 and a learning rate of 1e-3. The AdamW default approach is utilized for the learning rate updating. More information is shown in Table 2.

Table 2: Representation of Implementation Details

Network initialization	**"he" normal initialization**
Patch size	96x128x160
Batch size	3
Optimizer	AdamW with weight decay (*u=1e-5*)
Total epochs	120
Training time	12 hours each fold
Initial Learning rate (lr)	0.002
Lr decay schedule	Halved by 200 epochs
Loss function	Adaptive loss
No of flops	837.7117 G
No of model parameters	30.8M
CO2eq	3.92918Kg

C. Results and validation

The best fold was chosen based on the results of public validation as indicated in table 3 and compared with other existing works such as AdaptNet [21], DoDNet [27], CRCN [31], Tgnet [35] and MS-KD [36]. Out of the 100 validation sets, 50open cases are implemented to create the online validation result. Table 4 encloses a list of the validation results and Testing. Also figure 3 shows the graphical illustration of organs online validation and testing set.



(a) (b)

Figure 3. (a) Representation of Online Validation Set, (b) Representation of Testing Set

Table 3:10 Fold Comparison among Proposed Vs Existing Works

Target	AdaptNet		DoDNet		CRCN		Tgnet		MS-KD		Proposed	
	Organ	Tumor	Organ	Tumor	Organ	Tumor	Organ	Tumor	Organ	Tumor	Organ	Tumor
fold0	54.64	60.36	60.36	56.39	64.96	70.36	69.38	75.36	72.64	84.25	92.32	90.64
fold1	60.67	64.98	62.36	50.68	63.95	72.65	65.96	78.69	85.65	80.36	95.64	94.63
fold2	72.98	62.35	61.39	64.36	69.78	65.36	64.53	77.35	89.68	83.97	94.36	97.64
fold3	64.35	69.78	64.36	60.36	67.86	60.39	72.65	78.15	84.68	80.36	95.32	98.65
fold4	62.38	60.35	67.98	63.25	60.36	72.36	74.65	74.68	86.38	82.65	97.36	92.36
fold5	67.98	72.65	60.36	67.98	61.35	75.65	70.65	69.89	78.98	86.65	98.55	94.67
fold6	64.36	70.39	67.00	60.36	62.68	72.69	78.68	78.69	87.64	82.36	98.67	97.88
fold7	72.98	67.36	72.36	72.35	70.65	64.39	75.64	75.36	77.02	89.68	96.35	99.65
fold8	70.56	70.69	74.65	70.65	72.65	69.85	70.36	78.65	78.69	83.69	98.78	99.08
fold9	72.68	72.36	73.69	65.95	70.25	75.69	71.36	73.68	87.65	87.65	97.86	97.36
mean	66.36	67.13	66.45	63.23	66.45	69.94	71.39	76.05	82.90	84.16	96.52	96.26

Table 4: Quantitative Comparison on Online and Testing Set

Target	Online validation		Testing	
	DSC (%)	NSD (%)	DSC (%)	NSD (%)
Spleen	98.65	95.67	89.65	96.65
RK	97.69	96.38	95.65	93.87
Aorta	96.65	90.36	92.67	93.97
IVC	98.75	95.65	93.84	95.74
LAG	91.65	98.67	95.67	95.28
Liver	92.56	98.60	96.34	93.45
RAG	96.85	99.37	93.87	91.36
Stomach	95.60	92.36	95.67	98.65
Duodenum	97.67	95.64	97.36	94.98
Pancreas	96.92	97.36	96.67	95.68
Gallbladder	98.63	94.36	93.36	98.67
LK	97.69	95.37	98.67	95.69
Esophagus	94.65	96.37	95.69	93.69
Organ Average	96.47	96.15	95.10	95.25
Tumor	38.67	42.65	40.68	42.98

D. Quantitative Results on Validation Set

The measurable experimented were accomplished over the extraction research as indicated in Table 5. The presented approach performed better while compared to other existing works for tumor segmentation and exploring the improvements in DSC and NSD scores. The segmentation results of the existing works for organ segmentation are performance poor while compared to our approach. The aorta, IVG, gallbladder, LAG, stomach, esophagus and duodenum are between the organs with low frequency which is unable to segment by some of the existing works. Table show the performance of the existing works which achieved small while comparing to our suggested approach. The capacity of the approach is improved in malignancies and frequently arising organs like spleen, liver etc. Moreover, the suggested approach enhanced the segmentation of the object with high frequency while compared to the existing works which utilized the low frequency for the segmentation results. Additionally, our suggested model has achieved better performance with high frequency in segmentation results while compared to the existing works. Figure 4 illustrates accurate segmentation comparison among proposed and existing works.

Table 5: Quantitative Results Comparison among Proposed Vs Existing Works

Target	AdaptNet		DoDNet		CRCN		Tgnet		MS-KD		Proposed	
	DSC (%)	NSD (%)	DSC (%)	NSD (%)	DSC (%)	NSD (%)	DSC (%)	NSD (%)	DSC (%)	NSD (%)	DSC (%)	NSD (%)
Spleen	54.64	60.36	64.36	60.36	64.36	60.36	69.38	64.36	85.65	80.36	98.55	94.67
RK	60.67	64.98	67.98	63.25	67.98	63.25	65.96	67.98	89.68	83.97	98.67	97.88
Aorta	72.98	62.35	60.36	67.98	60.36	67.98	64.53	60.36	84.68	80.36	96.35	99.65
IVC	64.35	69.78	67.00	60.36	67.00	60.36	72.65	67.00	86.38	82.65	98.78	99.08

LAG	62.38	60.35	72.36	72.35	72.36	72.35	74.65	72.36	78.98	86.65	97.67	95.64
Liver	67.98	72.65	74.65	70.65	69.89	70.65	69.89	74.65	87.64	82.36	96.92	97.36
RAG	70.36	69.38	72.36	78.68	78.69	78.68	78.69	64.98	77.02	89.68	98.63	94.36
Stomach	72.65	65.96	75.65	75.64	75.36	75.64	75.36	62.35	64.35	69.78	97.67	95.64
Duodenum	65.36	64.53	64.35	70.36	78.65	70.36	78.65	69.78	62.38	60.35	96.67	95.68
Pancreas	60.39	72.65	62.38	71.36	73.68	71.36	73.68	60.35	64.36	60.36	98.55	94.67
Gallbladder	64.36	60.36	70.36	71.39	76.05	71.39	76.05	69.38	67.98	63.25	98.67	97.88
LK	67.98	63.25	72.65	65.96	67.98	63.25	72.65	65.96	60.36	67.98	96.35	99.65
Esophagus	60.36	67.98	65.36	64.53	60.36	67.98	65.36	64.53	85.65	80.36	98.78	99.08
Organ Average	66.05	65.85	68.46	69.65	71.20	69.73	72.12	67.45	76.55	76.25	97.86	97.01
Tumor	36.50	29.65	35.84	30.65	40.62	32.75	38.64	32.71	35.15	33.62	42.61	39.23

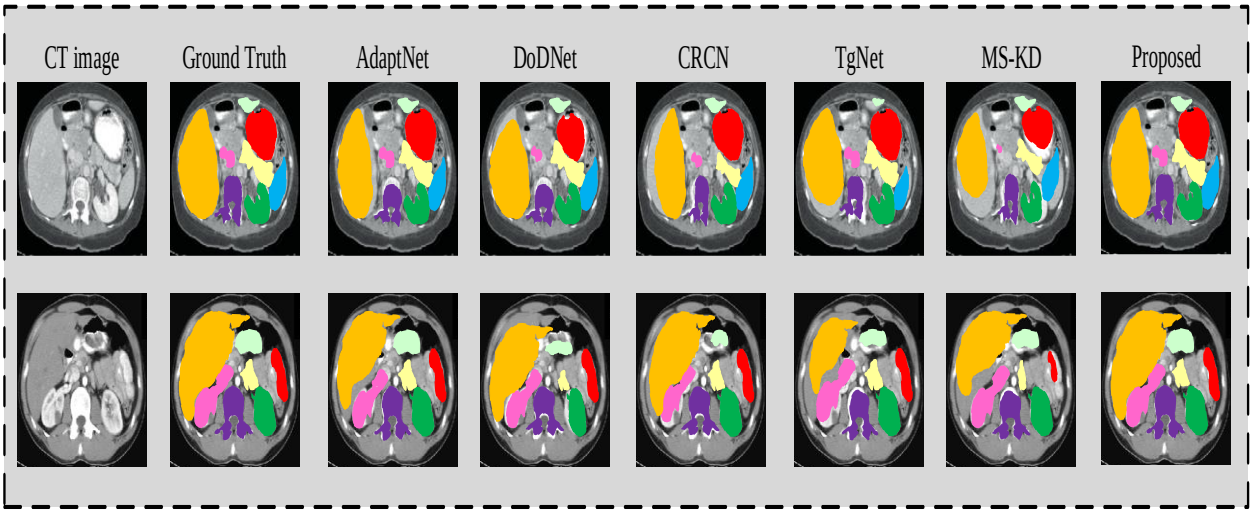


Figure 4. Segmentation Performance Comparison among Proposed Vs Existing Works

E. Segmentation Case Studies

We suggested two successful segmentation cases and two unsuccessful segmentation cases in this part.

(a) Good segmentation cases

Outstanding illustration of segmentation is in the IVC, aorta, stomach, duodenum and RAG will not segmented by implementing AdaptNet approach. A number of LK is improperly classifying as the spleen by the AdaptNet approach. The duodenum will only segment moderately by utilizing the suggested approach. Our technique performs better in the tumor than the existing works. The tumors in the RK, stomach, aorta, LK and IVC were not segmented using the existing works. The small part of LK is misclassified as pancreas while the suggested approach can also segment the tumor in LK. Our approach displays improved tumor segmentation performance and a better capacity to emphasize tumor segmentation compared to other approaches. The part of LK which is misclassified as part of the tumor and lesion in LK are under segmentation by the existing approaches.

(b) Bad segmentation cases

The IVC and aorta are difficult to other existing approaches to segment. Moreover, none of three methods can classify the esophagus. It can be extract that all of the ways are confusing due to the esophagus location. AdaptNet does not segment the duodenum, IVC, gallbladder or aorta as indicated in fig. All other three methods improperly detect the LK as the tumor. And also, the duodenum and pancreas have ambiguous boundaries in the predictive segmentation.

(c) Segmentation efficiency results on validation set

We have curved our approach encapsulating Docker container for the official challenge. We tested it on 20 illustrations and the average implementation time was 40.673 seconds, the average maximum GPU memory procedure was 4499.8 MB and the average area under the CPU cure was 124628 seconds. These are the effectiveness measures. Table 6 displays the eight situations which are well-organized. Figure 5shows the confusion matrix on segmentation efficacy.

Table 6: Computational Performance on Various Cases

Case ID	Image Size	Running Time (s)	Max GPU (MB)	Total GPU (MB)
0051	(512, 512, 101)	44.82	4950	154155
0019	(512, 512, 216)	42.33	4390	122557
0048	(512, 512, 500)	60.0	4750	186000
0029	(512, 512, 555)	75.40	5101	241407
0017	(512, 512, 156)	47.09	4528	171903
0001	(512, 512, 56)	34.40	4099	76885
0063	(512, 512, 449)	54.15	4774	164349
0099	(512, 512, 335)	52.90	4776	166711

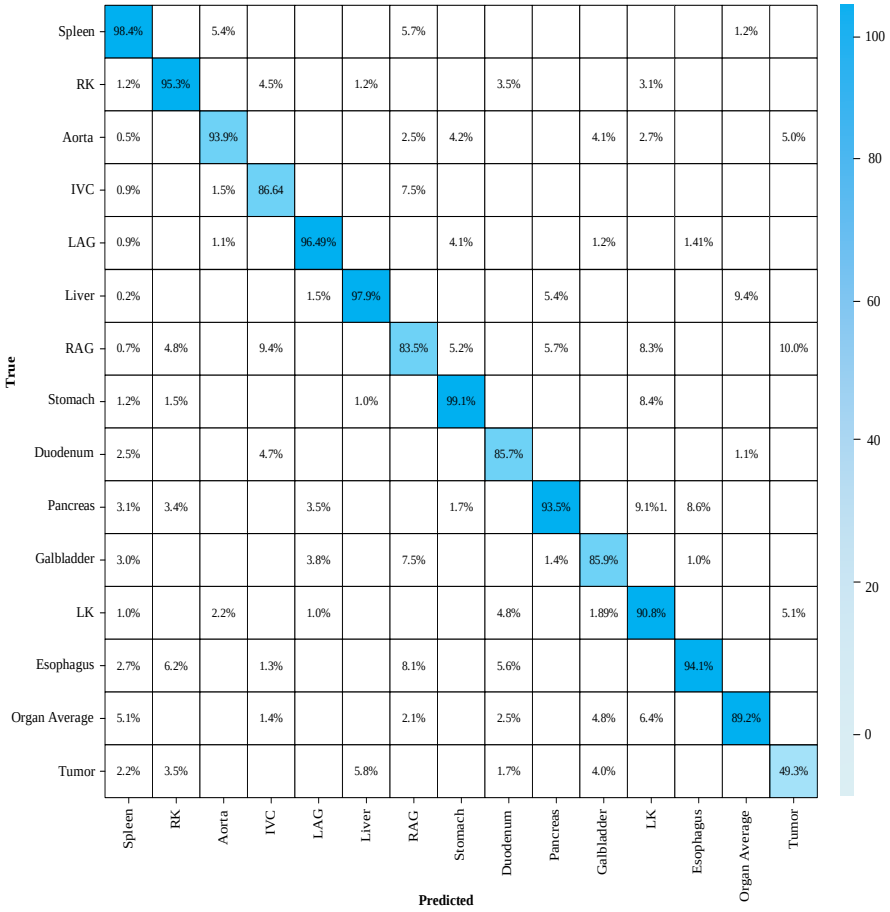


Figure 5. Confusion Matrix Representation on Various Organs and Tumor Segmentation

5. Conclusion

Higher computational complexity, and lack of annotation labels are the major issues employed in multi organ and tumor segmentation respectively. To resolve these issues, we propose a novel segmentation model named Intelligent Segmentor which segments both the multi organ and tumor from the CT images. Initially, the acquired CT images from the FLARE 23 datasets are pre-processed in terms of geometric standardization, noise removal, and intensity normalization respectively. The pre-processed image is then considered as labelled image which are provided for semi supervised training. Whereas the unlabeled image is provided for dual view training results in better pseudo label generation. Furthermore, the pseudo labels also fed to train the model which enhance the model performance. The test images are then provided as an input to the designed model which composed of two parallel encoders (i.e. GoogleNet and VGG-16) for contextual and spatial feature extraction. From the extracted features, side features and dimensionality reduction are achieved by Tweaked Feature Pyramidal Network (TFPN). The side features from the both the inputs are fused using Gated Fusion Module (GFM) to form unified feature map. The unified feature map are fed as input to the three consecutive convolution layers for robust multi organ and tumor detection.

References

[1] Wu, J., Li, G., Lu, H., & Kamiya, T. (2021). A supervoxel classification based method for multi-organ segmentation from abdominal ct images. Journal of Image and Graphics, 9(1), 9-14.

[2] Chen, S., Zhong, X., Hu, S., Dorn, S., Kachelrieß, M., Lell, M., & Maier, A. (2022, July). Automatic multi-organ segmentation in dual energy CT using 3D fully convolutional network. In Medical Imaging with Deep Learning.

[3] Lei, Y., Fu, Y., Wang, T., Qiu, R. L., Curran, W. J., Liu, T., & Yang, X. (2021). Deep learning architecture design for multi-organ segmentation. In Auto-Segmentation for Radiation Oncology (pp. 81-112). CRC Press.

[4] A. A. Khan, R. K. Mahendran, K. Perumal and M. Faheem, "Dual-3DM3AD: Mixed Transformer Based Semantic Segmentation and Triplet Pre-Processing for Early Multi-Class Alzheimer's Diagnosis," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 32, pp. 696-707, 2024, doi: 10.1109/TNSRE.2024.3357723.

[5] Jain, R., Sutradhar, A., Dash, A. K., & Das, S. (2021, December). Automatic Multi-organ Segmentation on Abdominal CT scans using Deep U-Net Model. In 2021 19th OITS International Conference on Information Technology (OCIT) (pp. 48-53). IEEE.

[6] Sherubha, "Graph Based Event Measurement for Analyzing Distributed Anomalies in Sensor Networks", Sådhanå(Springer), 45:212, https://doi.org/10.1007/s12046-020-01451-w

[7] Piyush K. Pareek, Pixel Level Image Fusion in Moving objection Detection and Tracking with Machine Learning ",Fusion: Practice and Applications, Volume 2 , Issue 1 , PP: 42-60, 2020

[8] Shivam Grover, Kshitij Sidana, Vanita Jain, "Egocentric Performance Capture: A Review", Fusion: Practice and Applications, Volume 2, Issue 2 , PP: 64-73, 2020.

[9] Abdel Nasser H. Zaied, Mahmoud Ismail and Salwa El- Sayed, A Survey on Meta-heuristic Algorithms for Global Optimization Problems, Journal of Intelligent Systems and Internet of Things,Volume 1 , Issue 1 , PP: 48-60, 2020

[10] Mahmoud H.Alnamoly, Ahmed M. Alzohairy, Ibrahim M. El-Henawy, "A survey on gel images analysis software tools, Journal of Intelligent Systems and Internet of Things,Volume 1 , Issue 1 , PP: 40-47, 2021.

[11] Lem, H., & Zhang, L. (2023, October). Mask R-CNN Transfer Learning Variants for Multi-Organ Medical Image Segmentation. In 2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (pp. 1209-1216). IEEE.

[12] Chen, D., Bai, Y., Shen, W., Li, Q., Yu, L., & Wang, Y. (2023). Magicnet: Semi-supervised multi-organ segmentation via magic-cube partition and recovery. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 23869-23878).

[13] Jia, D. (2022). Semi-supervised multi-organ segmentation with cross supervision using siamese network. In MICCAI Challenge on Fast and Low-Resource Semi-supervised Abdominal Organ Segmentation (pp. 293-306). Cham: Springer Nature Switzerland.

[14] Lei, Y., Wang, T., Tian, S., Fu, Y., Patel, P., Jani, A. B., ... & Yang, X. (2021). Male pelvic CT multi-organ segmentation using synthetic MRI-aided dual pyramid networks. Physics in Medicine & Biology, 66(8), 085007.

[15] Wang, T., Lei, Y., Roper, J., Ghavidel, B., Beitler, J. J., McDonald, M., ... & Yang, X. (2021). Head and neck multi-organ segmentation on dual-energy CT using dual pyramid convolutional neural networks. Physics in Medicine & Biology, 66(11), 115008

[16] . Tang, H., Liu, X., Han, K., Xie, X., Chen, X., Qian, H., ... & Bai, N. (2021). Spatial context-aware self-attention model for multi-organ segmentation. In Proceedings of the IEEE/CVF winter conference on applications of computer vision (pp. 939-949).

[17] Shi, G., Xiao, L., Chen, Y., & Zhou, S. K. (2021). Marginal loss and exclusion loss for partially supervised multi-organ segmentation. Medical Image Analysis, 70, 101979.

[18] Li, S., Wang, H., Meng, Y., Zhang, C., & Song, Z. (2024). Multi-organ segmentation: a progressive exploration of learning paradigms under scarce annotation. Physics in Medicine & Biology, 69(11), 11TR01.

[19] Huang, Z., Jiang, Y., Zhang, R., Zhang, S., & Zhang, X. (2024). CAT: Coordinating Anatomical-Textual Prompts for Multi-Organ and Tumor Segmentation. arXiv preprint arXiv:2406.07085.

[20] Toosi, A., Chausse, G., Chen, C., Klyuzhin, I., Benard, F., & Rahmim, A. (2022). Multi-modal, multi-organ deep segmentation of salivary and lacrimal glands in PSMA PET/CT images.

[21] S. Hemamalini ,V. D. Ambeth Kumar ,R. Venkatesan,S. Malathi. (2023). Relevance Mapping based CNN model with OSR-FCA Technique for Multi-label DR Classification. Journal of Fusion: Practice and Applications, 11 (2), 90-110.

[22] C. S. Manigandaa,V. D. Ambeth Kumar,G. Ragunath,R. Venkatesan,N. Senthil Kumar. (2023). De-Noising and Segmentation of Medical Images using Neutrophilic Sets. Journal of Fusion: Practice and Applications, 11 (2), 111-123.

[23] Mao, L. (2023). Semi-Supervised Two-Stage Abdominal Organ and Tumor Segmentation Model with Pseudo-Labeling.

[24] Huang, Y., Zhu, J., Hassan, H., Su, L., & Li, J. (2024). Label-efficient Multi-organ Segmentation Method with Diffusion Model. arXiv preprint arXiv:2402.15216.

[25] AN, P., XU, Y., & WU, P. (2023). Attention mechanism-based deep supervision network for abdominal multi-organ segmentation.

[26] Sathya Preiya, V., and V. D. Ambeth Kumar. (2023). Deep Learning-Based Classification and Feature Extraction for Predicting Pathogenesis of Foot Ulcers in Patients with Diabetes. Diagnostics 13(12), 1983.

[27] Balakrishnan, Chitra, and V. D. Ambeth Kumar. (2023). IoT-Enabled Classification of Echocardiogram Images for Cardiovascular Disease Risk Prediction with Pre-Trained Recurrent Convolutional Neural Networks. Diagnostics 13(4), 775

[28] Hemamalini, Selvamani, and Visvam Devadoss Ambeth Kumar. (2022). Outlier Based Skimpy Regularization Fuzzy Clustering Algorithm for Diabetic Retinopathy Image Segmentation. Symmetry, 14(12), 2512.

[29] Xie, Y., Zhang, J., Xia, Y., & Shen, C. (2023). Learning from partially labeled data for multi-organ and tumor segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence.