Minimax Theory and its Applications
Volume 11 (2026), No. 3s, 798 to 813

CSDA-MOSNet: Circular Sparse Dual Attention Multi Sparse Transformer for Multi Organ Segmentation

P.Savitha[1], Dr. Laxmi Raja[2]

[1]Research scholar Department of Computer Science and Engineering Faculty of Engineering Karpagam Academy of Higher Education Coimbatore, Tamil Nadu, India, Email: saviinfo@gmail.com
[2]Assistant Professor Department of Computer Science and Engineering Faculty of Engineering Karpagam Academy of Higher Education Coimbatore, Tamil Nadu, India, Email: laxmi.raja@kahedu.edu.in

Abstract
Intelligent multi-organ segmentation is a decisive task in computer aided diagnosis for diagnosing disease in body regions. Numerous Deep Learning (DL) models have been widely utilized for multi-organ segmentation tasks whereas they hindered with locating the organs with varied sizes and sparse boundaries. To this end, we design a novel model named Circular Sparse Dual Attention Multi Sparse Transformer based Multi Organ Segmentation Network (CSDA-MOSNet). For testing and training, we collect data from the two benchmark multi organ datasets such as Beyond the Cranial Vault (BTCV), and Abdominal Multi-Organ Segmentation Challenge (AMOS). Different from the state of art works, the acquired images from the two datasets are simultaneously fed into our model from two ends individually. Initially, both of images are processed by the Circular Dilated Convolution Network (CDCN) for capturing the long-range dependencies and boundary details. Aftermath, the captured details are carried out by zig-zag connected Multi Sparse Transformer Network (MSTN) for extracting long range temporal details which results in global context feature map. At last, Dual Branch Spatial Attention Module (DBSAM) is utilized for processing the global context feature map to obtaining final segmented outcome. The proposed research is examined using two benchmark datasets in terms of Dice Similarity Score (DISC) with 98.96% and 99.70% respectively. The results shows that the proposed CSDA-MOSNet accomplishes improved performance over existing works.

Keywords: Multi Organ Segmentation, Deep Learning (DL), Magnetic Resonance Imaging (MRI), Computed Tomography (CT), Circular Mixing, and Transformer.

Introduction

Accurate medical picture segmentation is a key component of automated medical image analysis and forms the basis for many important applications, including prognosis, diagnosis, treatment planning [1], image reconstruction, and therapy response evaluation. Many large datasets have been collected and made accessible expressly for the purpose of segmenting various organ types and tumour structures in order to assist these applications. Every dataset is customised to fulfil a specific role, such as segmenting a single organ or a portion of body tumours [2]. Because of this specialisation, creating and implementing algorithms for every use case would require a large number of models, maybe up to hundreds of models for different kinds of cancer, wounds, and other illnesses. This difficulty is further exacerbated by the time-consuming and expensive procedure of manually annotating data in order to generate extensive multi-organ datasets [3] on a large scale. Manual organ segmentation in clinical settings is typically carried out by radiation oncologists or radiologists, a process that is time-consuming and prone to errors. Annotators must meticulously delineate and verify slices one by one, often spending several hours per case. The rapid delineation [4] of multiple organs presents a daunting challenge, especially for junior oncologists, due to variations in imaging protocols, image quality, and anatomical structures. Multi-organ segmentation can be performed on various types of medical images acquired from different imaging modalities [5], including:

- Computed Tomography (CT),
- Magnetic Resonance Imaging (MRI),
- Ultrasound (US),
- Positron Emission Tomography (PET),
- Single-Photon Emission Computed Tomography (SPECT),
- X-ray Imaging.

There are several advantages to combining the MRI and CT modalities for multi-organ segmentation in medical image processing. MRI and CT provide complimentary information about tissues and organs: Whereas CT excels in high-resolution anatomical features [6]-[8] and bone visualisation, MRI is best at soft tissue contrast and functional imaging. The organs are better understood by the segmentation process by combining data from both modalities. Furthermore, combining MRI and CT data might lessen the impact of artefacts, improving the stability and dependability of segmentation algorithms, as various artefacts may be present in the images depending on the patient or the acquisition technique [9][10].

Over the decadeDue to deep convolutional neural networks' (DCNN) effectiveness in a variety of visual tasks, many DCNN-based techniques for abdominal organ segmentation have been put forth. Medical picture segmentation, particularly multi-organ segmentation, has been transformed by Deep Convolutional Neural Network (DCNN)-based techniques [11], but they also have significant drawbacks.A significant quantity of annotated data is needed for DCNN training, particularly when segmenting many organs, each of which may have unique properties and variances. Getting hold of these massively annotated datasets can be difficult and time-consuming. DCNNs are prone to overfitting, particularly when working with unbalanced classes or little training data[12]. Overfitting can lead to poor generalization and reduced performance on unseen data. To overcome the shortcomings of DCNN-based techniques, we created a unique multi-organ segmentation [13]-[15] model known as the Circular Sparse Dual Attention Multi-Organ Segmentation Network (CSDA-MOSNet). We used two benchmark datasets, the Abdominal Multi-Organ Segmentation Challenge (AMOS) and Beyond the Cranial Vault (BTCV), to train and assess our model. We divided the CT and MRI images in these datasets into patches, which we then fed into our Circular Dilated Convolution Network (CDCN) in two directions. Because of its capacity to record boundary information [16][17] and long-range interdependence across imaging modalities, the CDCN was selected.We next fed the interdependency-capturing outputs from the CDCN into our Multi-Sparse Transformer Network (MSTNet). The MSTNet uses zigzag-arranged transformers to extract temporal correlations across vast distances from the input pictures. After that, we passed these temporal relationships [18] into our Dual Branch Spatial Attention Module (DBSAM), which collected and rebuilt edge characteristics and spatial dependencies. Due to the DBSAM's ability to efficiently handle the spatial dependencies and edge information [19][20] required for precise organ segmentation in medical pictures, the method produced a final segmented output with increased segmentation accuracy. The major contribution of this work are listed below,

To the best of our knowledge, this the first work to utilize Circular Dilated Convolution Network (CDCN) for multi organ segmentation. The adoption of circular dilated convolution captures the long-range dependencies and boundary information from both the imaging modalities. Whereas the existing works utilized conventional dilated convolution fails to captures the long-range dependencies.

By adopting and connecting the Multi Sparse Transformer Network (MSTN) in a zig-zag fashion the temporal dependencies are easily captured and analysed to provided global context output.

The remaining of the work is organized as follows; section II explains the literature survey with existing research gaps. Section III emphases the proposed work with suitable theoretical, and mathematical equations. Section IV explains the experimental and implementation results, section V summarizes the proposed work with discussion. Section VII concludes the proposed work with future direction.

Literature Survey

Li B et.al [21] introduced a novel method for segmenting medical images using a dual encoder network called DECTNet. In particular, the DECTNet consists of four parts: a deep supervision module, a feature fusion decoder, a transformer-based encoder, and a convolution-based encoder. Additionally, they have eliminated local and global information by using a convolution encoder and a Swin Transformer encoder. In the end, they combined and chose the representation that came from the convolution and Swin Transformer encoders using a feature fusion decoder. Ye Q et.al [22] introduced an extensive approach, termed MIMO (Multi-Indicator and Multi-Organ), for evaluating medical image segmentation models. MIMO enables models to create distinct thresholds independently, which are then integrated with multi-indicator assessment and confidence estimation to assess and quantify each organ. This methodology provides in-depth insights into the segmentation of individual organs within each sample, assisting developers in analyzing and enhancing model performance. Furthermore, MIMO facilitates the generation of succinct usability and comprehensiveness ratings for various models. Models achieving higher scores are identified as excellent, streamlining their clinical evaluation process. Rickmann M et.al [23] introduced AbdomenNet, a deep neural network for precise abdominal organ segmentation in epidemiologic imaging studies. AbdomenNet combines convolutional and recurrent layers to capture spatial and contextual information efficiently. Trained on a large annotated dataset of abdominal CT

scans, AbdomenNet achieves state-of-the-art performance in segmenting organs like the liver, pancreas, kidneys, and spleen. Its efficacy highlights its potential for enhancing epidemiologic imaging studies.Chen D et.al [24] developed a partition-and-recovery N3 cube-based data augmentation method for both labelled and unlabeled photos. Our method improves the learning ability for tiny organs (within-branch) and promotes unlabeled pictures to understand organ semantics within their relative placements as compared to labelled images (cross-branch). We propose to refine pseudo labels by combining learnt representations from tiny cubes to add local features, hence improving within-branch performance. Our approach, which we call MagicNet, handles the CT volume as if it were a magic cube, lining up with the N3 cubes' partition-and-recovery mechanism, which mimics the dynamics of a magic cube game.Kanhere A et.al [25] suggested SegViz, a federated learning system that may be used to train segmentation models on partially annotated, distributed non-i.i.d datasets. The effectiveness of SegViz was contrasted with building separate models for each dataset and centrally combining datasets to train a single model. With the exception of the spleen segmentation, SegViz's aggregation technique of FedBN allowed for exceptional performance on the external BTCV set, with considerably higher dice scores ($p < 0.05$) for the liver, pancreas, and kidneys as compared to baseline models.Ji Y et.al [26] Presented AMOS, an extensive clinical dataset for the segmentation of abdominal organs. AMOS is made up of 500 CT and 100 MRI scans that were obtained from various suppliers, centres, phases, and illnesses. Every scan has voxel-level annotations for fifteen different abdominal organs, providing hard examples and acting as a testing ground for reliable segmentation algorithms on a variety of targets and conditions.Hu S et.al [27] provided a boundary-aware network (BA-Net) for multi-organ segmentation on CT and MRI scans. A common encoder, border, and segmentation decoder are features of this framework. Both decoders implement the multi-scale deep supervision technique. The boundary decoder produces boundary probability maps at various scales, acting as attention mechanisms to improve the segmentation feature maps. This method lessens the difficulties brought on by different organ sizes.Huang Z et.al [28] presented A-Eval, a benchmark for cross-dataset evaluation of abdominal multi-organ segmentation. It makes use of five datasets total—training data from FLARE22, AMOS, WORD, and Total Segmentator, together with validation data from BTCV. Additionally, evaluate the generalizability of the model using different data situations and investigate how the size of the model affects cross-dataset performance.Ciausu C et.al [29] developed a modality-agnostic domain randomization strategy to improve a U-Net segmentation network for abdominal organ segmentation by creating synthetic pictures from CT label maps during training. Comparable outcomes to fully-supervised segmentation models trained on MR data are obtained using this approach.Salimi Y et.al [30] introduced a deep learning approach for automatically segmenting multiple organs from whole-body CT images. The method utilizes deep neural networks to accurately segment organs like the liver, pancreas, kidneys, spleen, bladder, aorta, and gallbladder. By combining convolutional and attention mechanisms, the model captures spatial relationships and semantic features for precise segmentation. Extensive experiments show the approach's robustness across imaging modalities and anatomical variations, offering potential benefits for clinical diagnosis and epidemiologic research in medical imaging.Pan S et.al [31] suggested purpose of segmenting various organs in CT images, our unique deep learning model combines a U-shaped architecture with a Multi-Layer Perceptron Mixer (MLP-Mixer) and a Convolutional Neural Network (CNN). Our model replaces conventional convolutional blocks with an innovative MLP-Convolutional block, which is inspired by the V-net structure. Two datasets, one from an institution with 60 patient cases and the other from a public source (BCTV) with 30 patient cases, were used to assess our methodology. This analysis shows that our suggested network is capable of correctly segmenting many organs in CT images.Lian S et.al [32] presented the Partial- and Mutual-Prior Incorporated Framework (PRIMP), which aims to create a solid multi-organ segmentation model by utilizing knowledge from datasets with a single organ.They suggest using multi-organ statistics to incorporate the mutual-prior of organs and single-organ statistics to learn the partial-prior of each organ in order to operationalize these priors. This methodology guarantees that our model encompasses the distinct characteristics of organs as well as their interactions within the anatomical setting, resulting in a multi-organ segmentation framework that is more reliable and accurate.Jeon S et.al [33] Using a specialised 3D nnU-Net-based methodology, a completely automated multi-organ segmentation system for abdominal CT has been created. This method works well with both non-contrast and post-contrast photos. Portal venous phase (PVP) pictures were used to train the system, and they were combined with virtual non-contrast images from dual-energy CT (DECT) that were spatiotemporally matched. Next, a single-energy (SE) CT dataset including PVP and true non-contrast (TNC) images was used to assess the algorithm's performance.The findings demonstrated that, even in the absence of contrast enhancement, the liver, spleen, right kidney (RK), and left kidney (LK) could all be accurately segmented with mean Dice similarity coefficients (DSCs) more than 0.94 for each organ.Shen N et.al[34] innovated a model includes deformable receptive fields and makes use of organ features in terms of their positions and sizes.In order to do this, a spatial attention block is added, which learns spatial attention maps through explicit external supervision and emphasizes organ areas of interest during feature extraction. Furthermore, to ensure adequate segmentation across various organ types, a deformable convolution block is built to change receptive fields with trainable offsets to manage differences in organ forms and

sizes.Additionally, multi-scale attention maps and high-level semantic information are added to the U-Net's skip-connection structure, strengthening its segmentation performance and resilience in difficult situations.Li X et.al [35] suggested SUnet, a unique and effective medical picture segmentation method created especially for multi-organ segmentation in the belly and thorax, is presented in this work. SUnet is a neural network that is entirely focused on attention and consists of two essential parts: The Efficient Spatial Reduction Attention (ESRA) module was developed to improve the extraction of features from images while lowering the total number of model parameters and preventing overfitting. The Multiple Attention-Based Feature Fusion Module is a component of SUnet that enhances the network's capacity to efficiently fuse data from various sizes by facilitating cross-scale feature integration.

Through the integration of these components, SUnet is able to retain computational efficiency and lower the danger of overfitting, all while achieving robust and accurate multi-organ segmentation. Liu H et.al [36] provided The partial-label segmentation problem and systematically investigated by the authors in this study through the use of both theoretical analysis and actual assessments of current methods. The study's authors examined the problem of partial label supervision signals, in particular, how to distinguish signals based on ground truth data from those obtained from pseudo labels. Their suggested approach presents a brand-new COSST (Comprehensive Supervision Signals with Self-Training) two-stage architecture. To train a single model, they first use the two signals based on ground truth data. They then use several supervision signals to efficiently and successfully integrate the pseudo label signal into the original model through repeated self-training. Xu X et.al [37]Federated Multi-Encoding U-Net (Fed-MENU) is a revolutionary strategy that is introduced in the suggested multi-organ segmentation method. To extract organ-specific characteristics across many encodings sub-networks, this approach uses a Multi-Encoding U-Net. Every sub-network concentrates on a particular organ and receives customised training for every client application. An Auxiliary Generic Decoder (AGD) directs the training of the MENU-net in order to guarantee that the features retrieved from various sub-networks are both identifiable and informative.Zhang Y et.al [38The standard output layer in the suggested design is swapped out for a series of lightweight, class-specific heads. This layout improves flexibility, especially when managing recently created classes. These heads allow for independent predictions for both recently introduced and previously learned classes, which reduces the impact of new classes on old ones during continuous learning. We also propose to add Contrastive Language-Image Pre-training (CLIP) embeddings to the organ-specific heads in order to improve performance. Salimi Y et.al [39] The liver (157 instances), pancreas (390 cases), kidneys (310 cases), spleen (70 cases), bladder (112 cases), aorta (30 cases), and gallbladder (295 cases) were the organs for which they automated organ segmentation using CT scans. The data was divided into two groups, 20% for testing and 80% for training and validation, after preprocessing operations including scaling and resizing. Creating a residual deep neural network for picture segmentation was their goal. Using measures such as the Dice and Jaccard Coefficients, performance evaluation involved comparing the model's output with ground truth segmentations for each organ. Dong J et.al [40] By combining the best single-organ segmentation models currently available into a single multi-organ segmentation model for the given dataset, the method seeks to address the multi-organ segmentation problem. This approach lessens the need for labelled data, particularly in jobs involving the segmentation of many organs. To this end, the authors provide a brand-new two-stage technique that consists of a Model Ensemble step and a Model Adaptation stage. First, they improve each single-organ segmentation model's adaption and generalisation to the target domain. In the Model Ensemble step, they then combine and extract knowledge from several improved single-organ segmentation models, producing a strong multi-organ segmentation.

Proposed Model

In this section, we present the proposed model for multi organ segmentation with novel processes involved. Fig 1 illustrates the overall architecture of the proposed model. As the proposed research adopts both the CT and MRI images from the respective datasets, we split those images into patches and provide them to the Circular Dilated Convolution Network (CDCN) in two directions. The adoption of CDCN firmly captures the long-range dependencies and boundary information for both the imaging modalities. The captured dependencies are then fed as an input to the Multi Sparse Transformer Network (MSTNet) where the transformers are connected in zig-zag fashion to extract the long-range temporal dependencies for the input image. The temporal long-range dependencies are then provided to the Dual Branch Spatial Attention Module (DBSAM) where the spatial dependencies and edge features are extracted and reconstructed to form the final segmented output with better segmentation accuracy. The detailed processes are provided below.

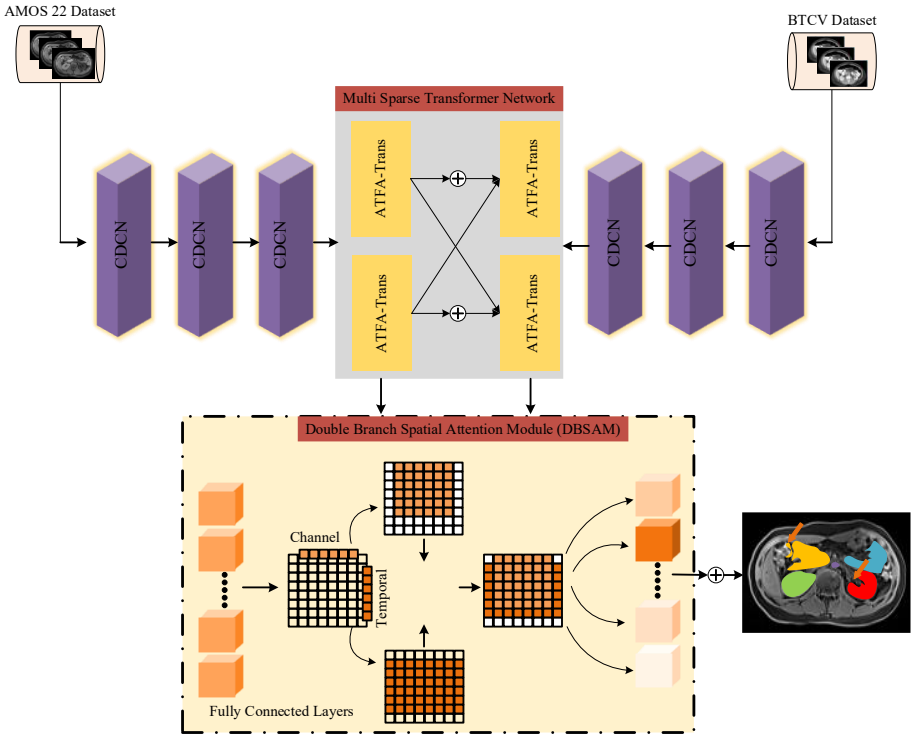


Fig.1. Overall Architecture of the Proposed CSDA-MOSNet

Pre-Processing

The acquired multi organ images are initially subjected to pre-processing in terms of noise reduction, artifact removal, and intensity normalization respectively. The detailed mathematical equation is provided as below,

Noise Reduction

For noise reduction in the input image (Im), we adopt advanced noise reduction algorithm named Non-Local Means (NLM) denoising algorithm which is entrenched on noise similarity technique. To be clearer, the redundancy noisy patches are cleansed by averaging the patches. The mathematical equations for the proposed denoised algorithm is provided below,

For an image (Im) with pixel location (loc) patch similarity is computed. The similarity among the patches p_1 and p_2 is computed based on Euclidean distance among the intensity value of the patches as,

$$\mathrm{dis}(p_1, p_2) = \sum_j (\mathrm{Im}(p_1 + j) - \mathrm{Im}(p_2 + j))^2 \qquad (1)$$

After that, we compute the weighted average of the every pixel in the image. The pixel weights are calculated by computing the similarity among the centered and other pixel patches and can be formulated as,

$$\widehat{\mathrm{Im}}(\mathrm{CP}) = \frac{1}{\mathrm{Nor}(\mathrm{CP})} \sum_{p_2} \mathrm{we}(p_1, p_2).\, \mathrm{Im}(p_2) \qquad (2)$$

From the above equation, $\mathrm{Nor}(\mathrm{CP})$ is the normalization factor, $\mathrm{we}(p_1, p_2)$ denotes the pixel p_2 weight based on the similarity among the CP and p_2.

(ii) Artifact Removal

The noise removed image patches (Im_{pt}) are then fed for artifact removal where we adopt DL model named patch entrenched Convolutional Neural Network (CNN). Initially, we extract patches from the Im_{pt} where the extracted patches are fixed in size. The extracted patches are provided to the CNN model where there are multiple layers involved such as pooling and activation function respectively.

For the denoised image Im_{pt} which also compose of artifacts, the proposed research extract patches Pc_j with size $N \times N$ in which the set of patches are denoted as $\{Pc_1, Pc_2, \dots, Pc_M\}$. The extracted patches are provided to the CNN where it performs mapping with function $f: Pc_j \rightarrow \widehat{Pc_j}$ in which $\widehat{Pc_j}$ denotes the artifact free patch. To optimize the artifact removal process, we reduce the reconstruction error among the $\widehat{Pc_j}$ and ground truth patches $\widetilde{Pc_j}$ from the training dataset. At last, the artifact removed image $\widehat{Im}_{pt}$ gets reconstructed.

(iii) Intensity Normalization

The $\widehat{Im}_{pt}$ are then provided for intensity normalization which is adopted for standardize the intensity values by confirming their value range distribution. For intensity normalization, we utilize Z-score normalization method which performs transformation of image intensity with a standard deviation to one and mean to zero. The actual intensity of the artifact removed image is denoted as $\widehat{Im}_{pt}^{(Ar)}$ whereas the normalized intensity is denoted as $\widehat{Im}_{pt_{Norm}}^{(Ar)}$. The formulation of z-score normalization is initialized by computing standard deviation ($\propto$) and mean (τ) as,

$$\propto = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(\widehat{Im}_{pt}{}^{(Ar)}_{j} - \tau)^2} \qquad (3)$$

$$\tau = \frac{1}{N}\sum_{j=1}^{N} \widehat{Im}_{pt}{}^{(Ar)}_{j} \qquad (4)$$

From the above equations () & (), the voxels in the image are represented as N. After that, the normalized intensity values are determined by the subtracting the τ and divide it by the $\propto$ in the every image voxel as,

$$\widehat{Im}_{pt_{Norm}}^{(Ar)} = \frac{\widehat{Im}_{pt}{}^{(Ar)} - \tau}{\propto} \qquad (5)$$

From the aforementioned steps, we have pre-processed the acquired medical images. After that we have perform several novel operations to effectively segment the multi organ medical images.

Circular Dilated Convolution Network

The pre-processed medical images from the two datasets are represented as $Im_{D_1}^{pr}, Im_{D_2}^{pr}$. Both the images are fed into two directions to the CDCN where the images are provided with equal and fair chance in every possible position to collect the information from the succeeded layers.

In order to capture the interactions, we adopt symmetric dilated convolution layers. The size of dilation is increase in number based on the network depth. The reason for adopting dilated convolution is that, the dense image regions are predicted easier with its multi scale information. The image receptive field can be improved by adopting dilated convolution. The size of dilation can be increased exponentially $dia_l = 2^{l-1}$ in which dia_l denotes the size of dilation at the l-th convolution layer. The scalability of model can be expanded firmly by combining the dilated convolutions with deep learning model. To resolve the issue of notion clutter, the 1D input sequence of the l-th convolution layer is denoted as $[u_1, u_2, \dots, u_M]$. The output of the convolution is denoted as v_p at the p-th position ($1 < p < M$) that is calculated by $v_p = \sum_{h=0}^{h-1} q_h^{(l)} \cdot u_{p+\left(h-\frac{h-1}{2}\right).dia_l}$ in which the size of kernel is denoted as p with odd number, the co-efficient of the l-th convolution layer is denoted as $q^{(l)}$.

To further enhance the position information extraction, we utilize circular mixing or padding strategy where the blinds spots in the traditional zero-padding can be resolved. To be clearer, the position signals in the boundaries are mixed up to form the cluttered blind spots. The issues can be resolved in the circular padding by declined the boundary signals to convolved with zero whereas the opposite end boundary position are convolved with each other. On the whole, the circular padding strategy enhances the boundary position information by reducing the sensitivity to data shift issues. The formulation of circular dilated convolution is provided as below,

$$v_p = \sum_{h=0}^{h-1} q_h^{(l)} \cdot u_{\left[p+\left(h-\frac{h-1}{2}\right).dia_l\right] \bmod M} \qquad (6)$$

From the above equation, the boundary position can be easily connected with long tenure dependencies. For the provided input sequence, every boundary position can receive an equal chance resulting in accurate feature extraction.

Multi Sparse Transformer Network (MSTN)

The capture long tenure dependencies and boundary features $fea^{(bo,lo)}$ from the CDCN is fed to the MSTN which composed of four transformer modules that are connected in zig-zag fashion. In addition to that, the Hierarchical Attention Module (HAM) also presented in the MSTN. Fig 2 represents the MSTN illustration.

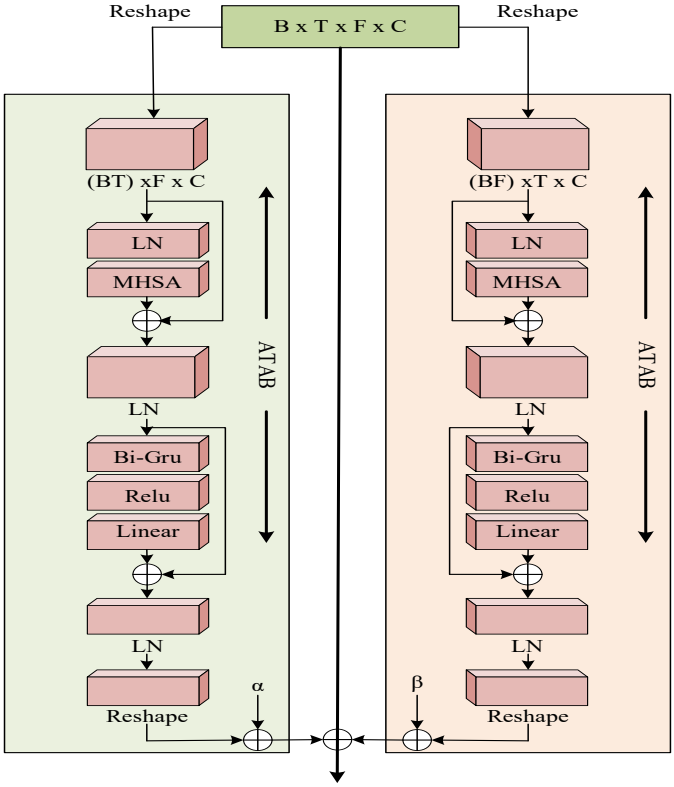


Fig.2. Illustration of MSTN

The MSTN firmly extract the temporal dependencies features, intermediate multi scale information features with lesser computational cost. The outputs from the CDCN on the both sides are provided to the MSTN which consist of two branches named Adaptive Regularity Consideration Branch (ARCB) and Adaptive Temporal Consideration Branch (ATCB). Both of the module's apprehensions the frequency and temporal features using dual adaptive weights ∃ & ∄ respectively. Furthermore, the transformer module composed of Normalization Layer (NL), residual connection, feed-forward network, and Multi Head Self Attention (MHSA). Different from the conventional transformer architecture, the feed forward network is enhanced by Bi-directional Gated Recurrent Network (Bi-GRU). For retaining the channel number of kernel size (2,2) 2D convolution and PReLU is employed.$R^{B\times1\times1\times1\times1}$ The Adoption of HAM tends to amalgamates the extracted feature maps from the MSTN that outputs the $Fe = \{Fe_n\}_{n=1}^{N}, Fe_n \in R^{Ba\times Te\times Fre'\times C}$ in which dimension of halved frequency is denoted by Fre', batch size is denoted by Ba, and transformer numbers is denoted by N. The feature output of the every MSTN is then employed to the average pooling layer (avg_{pool}) with 1×1 convolution layer. The mentioned layers are squeezed the features to form the global feature representation that can be denoted as $Po_n^b = avg_{pool}(Fe_n) * Co_n \in R^{B\times1\times1\times1\times1}$..After that all the pooled outputs are amalgamated that can be denoted as $Po^b \in R^{B\times1\times1\times1\times N\times1}$ and are provided to the softmax layer for extracting the HAM $Co_n \in R^{B\times1\times1\times N\times1}$. The correlated outputs from the MSTN are amalgamated to acquire a global feature $Fe^b \in R^{B\times Te\times Fre'\times C\times N}$. Posteriorly, the multi scale contextual features are obtained by computing matrix multiplication among Fe^b and Co^b that can be computed as,

$$W_N = Fe^b Co^b = \sum_{n=1}^{N} \frac{\exp(avg_{pool}(Fe_n) * Co_n)}{\sum_{n=1}^{N} \exp(avg_{pool}(Fe_n) * Co_n)} Fe_n \quad (7)$$

From the above equation, transitional output of the MTSN can be denoted as Fe_n, the contextual global feature map can be denoted as $W_N \in R^{B\times Te\times Fre'\times C}$. At last, the output can be obtained by performing element wise sum

operation among MSTN feature map output $\mathrm{fea_N}$ and contextual global feature map $\mathrm{Glo_N}$. The formulation is provided as below,

$$\mathrm{OP^{MSTN}} = \mathrm{Glo_N} + \Phi \mathrm{fea_N} \qquad (8)$$

From the above equation, scalar learnable co-efficient is denoted as Φ and are set to 0. The adaptive merging technique intelligent allocates an appropriate weight value for merging both the extracted feature information.

Dual Branch Spatial Attention Module (DBSAM)

The output from the MSTN holds the global and contextual information rich feature $\mathrm{OP^{MSTN}}$ which is then fed to the DBSAM for capturing the channel and temporal correlations. For that, we took fully connected layer for enabling the relationship among the channel and temporal information. More clearly, when the number channel increases the parameters utilization rate also increases which is shown in fig 3. In order to reduce the high utilization of parameters and amplifies the amenable field of important details, we employ DBSAM. The phases with mathematical validation involved in the DBSAM are explained as below,

Squeezing Phase: The correlations among the channel and temporal information in the $\mathrm{OP^{MSTN}}$ can be obtained by performing squeezing operation on input feature as $\mathrm{OP^{MSTN}} \in \mathrm{R^{Tem \times He \times We \times Ch}}$ in which Tem denotes the temporal time step, Ch denotes the channel size. The squeeze phase computes the middling matrix $\mathfrak{V} \in \mathrm{R^{Tem \times Ch}}$ based on every element in the $\mathfrak{V}_{(\mathrm{ch,tem})}$ as,

$$\mathfrak{V}_{(\mathrm{ch,tem})} = \frac{1}{\mathrm{He \times We}} \sum_{\mathrm{j=1}}^{\mathrm{He}} \sum_{\mathrm{i=1}}^{\mathrm{We}} \mathrm{OP^{MSTN}}_{\mathrm{j,i}}^{(\mathrm{ch,tem})} \qquad (9)$$

From the above equation, $\mathrm{OP^{MSTN}}^{(\mathrm{ch,tem})}$ denotes the ch-th channel when the t-th time step.

(ii) Temporal Entrenched Local Attention (TELA): For founding temporal associations, the TELA is adopted succeeded by squeezing phase. We employ one dimensionalconvolution operation to extracts the local association among the temporal data. To be clearer, for every squeezing frame we intend to capture the temporal associations on every row of middling matrix $\mathfrak{V}$. After that, convolving operations is done for the gathering feature maps of the $\mathfrak{V}$. The formulation of TELA is provided as below,

$$\mathrm{Tem_{j,i}} = \sum_{\mathrm{n=1}}^{\mathrm{Ch}} \sum_{\mathrm{m=0}}^{\mathrm{ker_{Tem-1}}} \omega_{(\mathrm{n,i})}^{\mathrm{m}} \mathfrak{V}_{(\mathrm{n,i+m})} \qquad (10)$$

From the above equation, convolution kernel size is denoted as $\mathrm{ker_{Tem}}$ ($\mathrm{ker_{Tem}} < \mathrm{Tem}$) that designates the convolution operation time step. The learnable parameter is denoted as $\omega_{(\mathrm{n,i})}^{\mathrm{m}}$ that denotes the i-th convolution operation of the m-th parameter. Finally, the matrix of attention score based on the $\mathfrak{V}$ can be denoted as $\mathrm{Tem} \in \mathrm{R^{Ch \times Tem}}$.

(iii) Channel Entrenched Local Attention (CELA): For temporal correlation, the saliency score did not adopt thereby channel dimension estimation did not extracted. This research adopts temporal channel one dimensional convolution operations with $\mathfrak{V}$ matrix, and the row wise convolution operation can be formulated as,

$$\mathrm{Ch_{j,i}} = \sum_{\mathrm{n=1}}^{\mathrm{Tem}} \sum_{\mathrm{m=0}}^{\mathrm{ker_{Tem-1}}} \mathcal{E}_{(\mathrm{n,j})}^{\mathrm{m}} \mathfrak{V}_{(\mathrm{j+n,i})} \qquad (11)$$

From the above equation, $\mathcal{E}_{(\mathrm{n,j})}^{\mathrm{m}}$ is the learnable parameter that denotes the j-th channel of the one-dimensional convolution in the n-th column of $\mathfrak{V}$. The attention score matrix is denoted for the CLEM is denotes as $\mathrm{Ch} \in \mathrm{R^{Ch \times Tem}}$.

(iv) Convolution Fusion Phase (CFP): The channel and temporal attention scores for the feature inputs are obtained as Ch & Tem respectively. The major reason for adopting CFP is to compute the CFP matrix $\mathfrak{A}$ which is utilized for computing the relationship among the temporal and channel correlations. Precisely, the relationship among the channel and temporal relation as below,

$$\mathfrak{A}_{\mathrm{j,i}} = \eth\left(\mathrm{Tem_{j,i}}, \mathrm{Ch_{j,i}}\right) = \eth\left(\sum_{\mathrm{n=1}}^{\mathrm{Ch}} \sum_{\mathrm{m=0}}^{\mathrm{Ker_{Tem-1}}} \omega_{(\mathrm{n,i})}^{\mathrm{m}} \mathfrak{V}_{(\mathrm{n,i+m})} \cdot \sum_{\mathrm{n=1}}^{\mathrm{Tem}} \sum_{\mathrm{m=0}}^{\mathrm{Ker_{Ch-1}}} \mathcal{E}_{(\mathrm{n,j})}^{\mathrm{m}} \mathfrak{V}_{(\mathrm{n,i+m})}\right) \qquad (12)$$

From the above equation, the sigmoid function is denoted as $\eth$. The output of the DBSAM provides the reconstructed multi organ segmented output

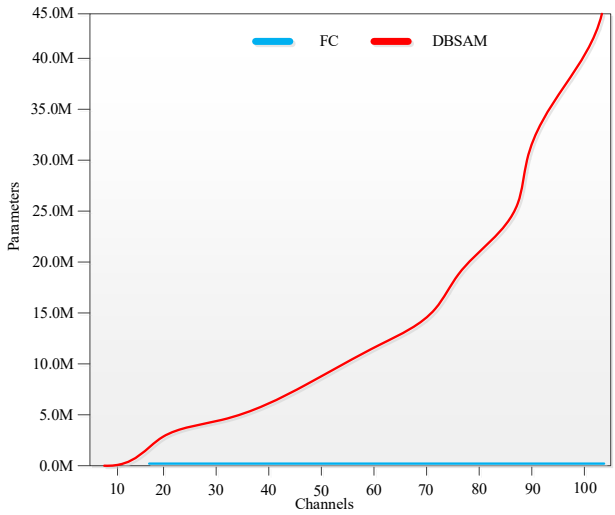


Fig.3. Parameter Utilization Rate

Experimental Results

Dataset Details

AMOS 2022 Dataset: Although autonomous abdominal multi-organ segmentation from CT/MRI images has made substantial progress in recent years, the lack of a large-scale benchmark covering a variety of clinical settings hinders a thorough assessment of model capabilities. The majority of deep learning models are trained on datasets with a few numbers of samples or organs since it is expensive to gather and annotate 3D medical data. This constraint limits the capabilities of current deep models and makes it more difficult to carry out an impartial and thorough analysis of various approaches.We seek to widen the area of assessment by introducing the Abdominal Multi-Organ Segmentation (AMOS) dataset in order to solve this problem. AMOS is made up of five hundred CT and MRI scans that were acquired from various vendors, centres, phases, and disordersEvery scan has voxel-level annotations for 15 abdominal organs, which provide difficult examples and act as a testing ground for dependable segmentation algorithms under many circumstances and goals.The AMOS dataset provides a thorough and varied set of clinical data for abdominal organ segmentation, therefore mitigating the shortcomings of previous datasets. AMOS is an invaluable resource for researching effective segmentation algorithms, containing 550 CT and MRI images from patients with a wide range of disorders, centres, vendors, modalities, phases, and levels of difficulty. Furthermore, we use this difficult and distinct dataset to perform benchmark evaluations of other cutting-edge medical segmentation algorithms to determine their effectiveness.

BTCV Dataset:A well-known set of medical pictures used for brain tumour segmentation is called the BTCV (Brain Tumour Core Value) dataset. It serves as a reference dataset for testing and developing medical image analysis algorithms, with an emphasis on brain tumour identification and classification. The BTCV dataset includes MRI scans with a variety of modalities, including FLAIR, T1-weighted, and T2-weighted scans, which provide important information on brain anatomy and pathology that is necessary for precise tumour segmentation. Every scan is meticulously labelled with brain tumour information, acting as a standard for segmentation algorithms. Robust algorithm assessment is ensured by the dataset's diversity in patient demographics, tumour kinds, and clinical circumstances. Scans are subjected to standard pre-processing for uniformity and quality before inclusion. Algorithm Development: It offers a framework for creating and refining algorithms for automatically identifying and classifying brain tumours. Performance Evaluation: To measure accuracy and dependability, algorithms are evaluated against ground truth annotations using metrics such as Hausdorff distance, sensitivity, specificity, and Dice coefficient. Comparative Studies: By comparing several algorithms, the dataset makes it possible to identify state-of-the-art methods and developments in brain tumour segmentation.

Implementation Details

By utilizing Python 3.10.0 and PyTorch 2.1 we conducted an experiment through a CentOS 7. Furthermore, SGD optimizer was utilized with a Nesterov momentum by 0.99 and 1e-3 weight decay. This approach set aside size of batch constant through 4 and premeditated individually epoch towards include 370 reprises with an early

learning rate of 0.01 to train the data from the scratch and were adjusted according to a poly learning rate policy over time. Training duration were accepted by different approaches adapted to each dataset by recognizing essential training time to congregate across various datasets. AMOS and BTCV training period were extended to 4000 epochs on the large and complex dataset to fully to leverage its potential. For the other datasets BTCV and AMOS, we maintained standard configuration of 1000 epochs as our assessments specified that spreading training beyond this period did not enhance performance.

Cross-Dataset Evaluation for Models Trained on Individual Datasets

The generalizability of models trained independently on the BTCV and AMOS 2022 datasets is evaluated in this section. Table 1 demonstrates how these datasets' model performances varied significantly from one another. It is clear that models that have been trained on a particular dataset typically perform better when tested on that dataset, but they frequently perform worse when tested on other datasets. These patterns are graphically shown in Figure 4, which shows which models have the lowest standard deviations (89.42% for NSTD and 87.37% for DICE). This suggests that the model is more stable and broadly applicable to a variety of test datasets. When it comes to average performance and stability, the AMOS model falls short, receiving highest scores (DISC of 90.32%, NSD of 91.34%). It is noteworthy that a clear relationship appears to exist between the training dataset size and the generalizability of the model, with the order of AMOS 2022 < BTCV. The BTCV dataset has low variability (STD for DISC of 86.37 % and for NSTD of 87.38 %) when seen through the lens of testing, suggesting consistent but less distinguished assessments. demonstrating a consistent but less critical assessment. It manifests as increased performance variability in AMOS 2022. This shows that the validation samples in these datasets are distinct, difficult, and resistant to simple generalisation

Table 1. Cross -Dataset Validation

Train/Test	AMOS-2022		BTCV		Mean ± STD	
	DISC	NSTD	DISC	NSTD	DISC	NSTD
AMOS-2022	90.32	91.34	87.78	88.45	**90.37 ±3.94**	**91.37 ± 1.94**
BTCV	89.56	88.43	85.67	84.76	88.37 ± 4.56	87.18 ±4.04
Mean± STD	86.37 ± 3.14	90.37 ± 2.04	86.37 ± 2.34	87.38 ± 2.94	87.37 ± 1.94	89.42 ± 3.23

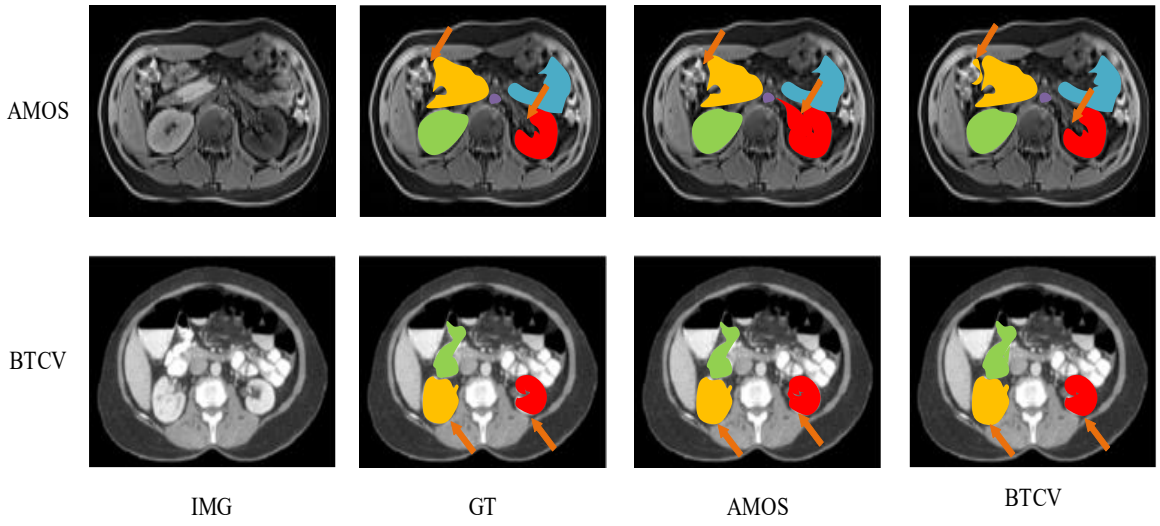


Fig.4. Performance Comparison on Samples from Two Different Datasets (AMOS and BTCV)

Quantitative Analysis

Precision, we compared four alternative models in the BTCV dataset: MIMO [22], Magic-Net [24], SegViz[25], and MLP-Mixer [31]. The performance of the models was assessed using the average dice coefficient in each dataset. Table 2 and fig 5 illustrates compared to previous studies, our model performs exceptionally well in average dice coefficient 0.843, and usability comprehensiveness 0.685. Additionally, MIMO Outperformed the other three model by a slight margin in the BTCV dataset. In AMOS 2022, We are evaluating the performance

of four different models: MIMO [22], BA-Net [27], A-Eval [28], and A U-Net [29]. The high dice coefficient of 0.893 and high usability comprehensiveness 0.685 that our model attained. In both datasets, our model outperformed other models by a significant margin. Fig 6 (a) & (b) shows the DSC comparison on BTCV and AMOS dataset respectively.

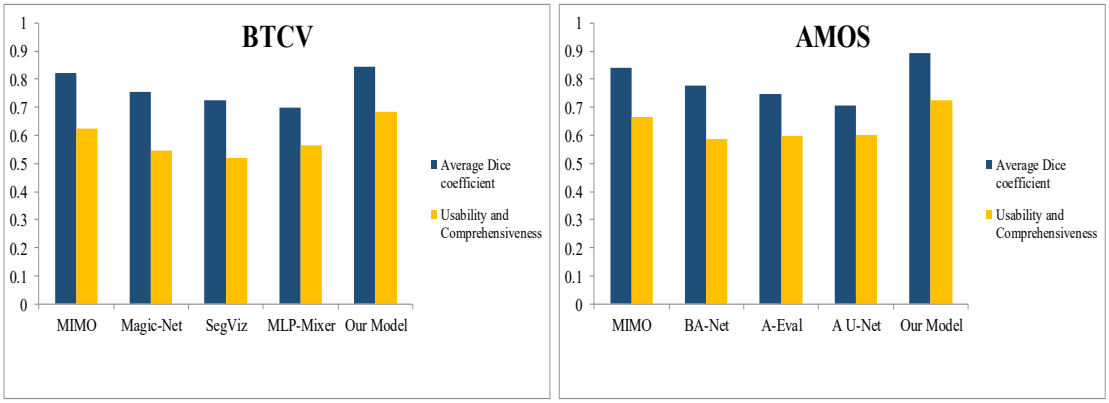


Fig.5. Quantitative Comparison on Proposed Vs Existing Works

Table 2. Validation on BTCV Dataset

Models	**BTCV Dataset**	
	Average Dice coefficient	**Usability and Comprehensiveness**
MIMO	0.821	0.625
Magic-Net	0.756	0.546
SegViz	0.725	0.521
MLP-Mixer	0.698	0.565
Our Model	**0.843**	**0.685**
Models	**AMOS 2022 Dataset**	
	Average Dice coefficient	**Usability and Comprehensiveness**
MIMO	0.839	0.664
BA-Net	0.778	0.587
A-Eval	0.749	0.598
A U-Net	0.707	0.601
Our Model	**0.893**	**0.725**

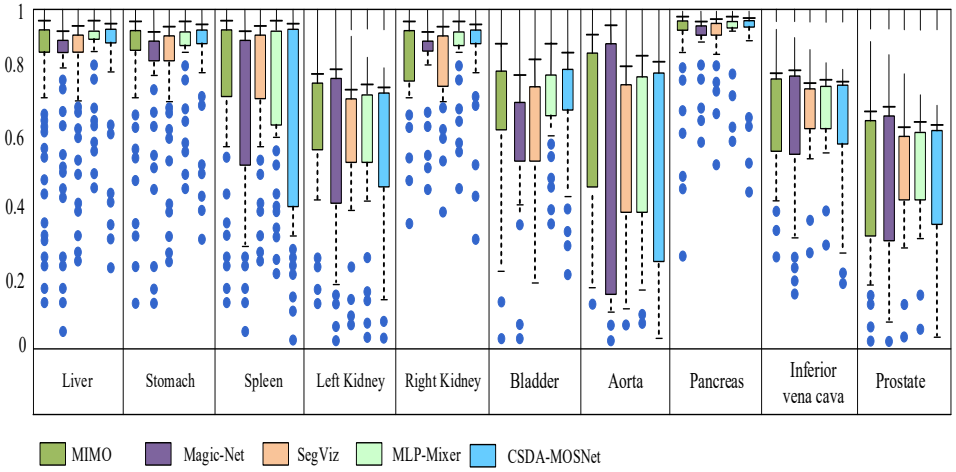


Fig. 6(a). DSC Comparison on BTCV Dataset

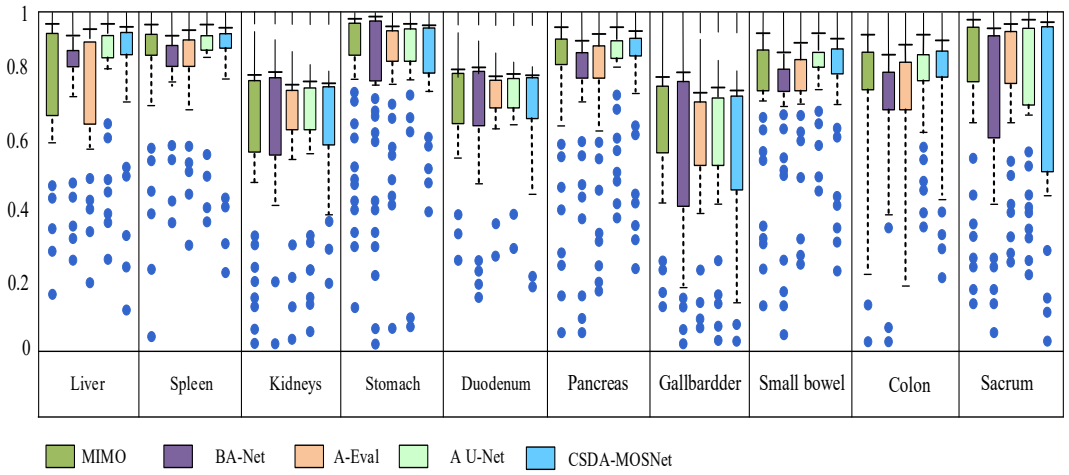


Fig. 6(b). DSC Comparison on AMOS Dataset

Confidence Estimation

Our model demonstrates superior calibration levels with lower Expected Calibration Error (ECE) and Misclassification Error (MCE) scores across both datasets. Conversely, Multi Sparse Transformer Network shows the poorest calibration with the highest scores in these metrics across both indicators. However, it's important to note that the relative ranking of other models may vary depending on whether ECE or MCE is considered, as well as the specific datasets used. Furthermore, there's an interesting observation where Dual Branch Spatial Attention Module outperforms e Circular Dilated Convolution Network in correctness scores but lags in reliability. This poses a dilemma in clinical model selection, where the choice between prioritizing correctness or reliability becomes crucial. Consequently, directly comparing models for clinical deployment presents challenges due to these conflicting performance aspects.

$$ECE = \sum_{i=1}^{N} \frac{|B|}{N} |acc(B_i) - conf\,(B_i)| \qquad (13)$$

$$MCE = \frac{\text{Total Misclassified Pixels /Voxels}}{\text{Total Pixels/Voxels}} \qquad (14)$$

Where, N is the number of bins, B_i is the i-th bin, $|B_i|$is the number of samples in the i-th bin, acc (B_i) is the accuracy of the i-th bin, and conf (B_i)is the average confidence of the i-th bin. Whereas "Total Pixels/Voxels" shows the overall number of pixels or voxels in all segmented pictures, "Total Misclassified pixels/Voxels" indicates the total amount of misclassified pixels or voxels across all segmented images. Table 3 & 4 shows the confidence validation on BTCV and AMOS dataset respectively. Fig 7 shows the confidence estimation comparison on BTCV and AMOS dataset.

Table 3. Confidence Validation on BTCV Dataset

Models	**BTCV Dataset**	
	ECE(%)	**MCE(%)**
MIMO	61%	54%
Magic-Net	67%	58%
SegViz	53%	42%
MLP-Mixer	46%	32%
Our Model	**39%**	**28%**

Table 4. Validation on AMOS Dataset

Models	**AMOS 2022 Dataset**	
	ECE(%)	**MCE(%)**
MIMO	62%	57%
BA-Net	58%	50%
A-Eval	48%	45%
AU-Net	50%	40%
Our Model	**33%**	**30%**

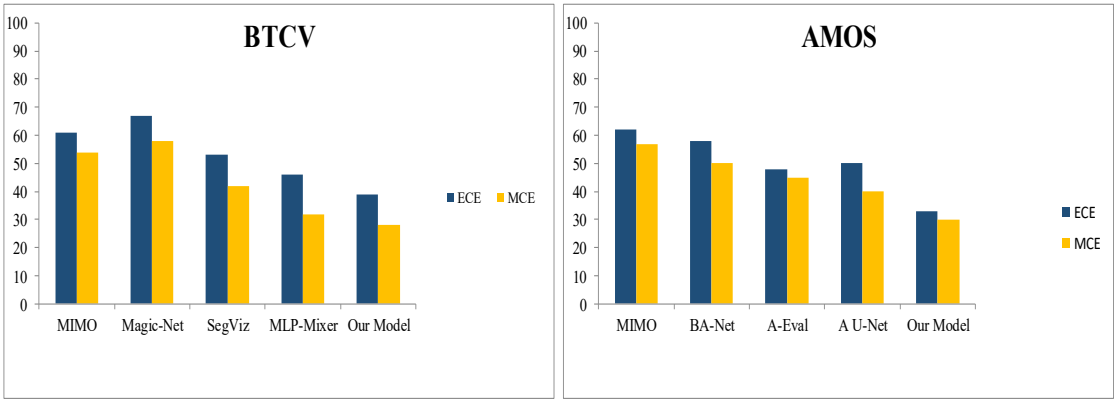


Fig.7. Confidence Estimation Comparison on Proposed Vs Existing Works on BTCV and AMOS Dataset

Discussion

The BTCV and AMOS 2022 datasets showed outstanding performance from the CSDA-MOSNet suggested technique. With little to no decrease in the global meta-model's average dice score performance, our method effectively combined knowledge from all image nodes. There are several potential advantages to image segmentation from diverse datasets with imperfect annotations. For instance, by enabling knowledge sharing across models, CSDA-MOSNet has the potential to cut labelling time by 1/η, where η is the number of different labels in the distributed data sets.We think that a number of intrinsic benefits in CSDA-MOSNET implementations like the use of random affine transformations during training and a learning rate decay are responsible for the system's effectiveness in handling non-i.i.d data.

CDCN:A Circular Dilated Convolutional Neural Network (CDCNN) is a type of convolutional neural network (CNN) architecture that incorporates circular dilated convolutions.

Dilated Convolution: In a standard convolution operation, a filter/kernel moves over an input image in a sequential manner, typically with a stride of 1 pixel at a time. Dilated convolution introduces gaps (dilation rate > 1) between the elements of the filter, effectively increasing the receptive field without increasing the number of parameters. This helps capture larger contextual information.

Circular Padding: Circular padding, also known as periodic padding or circular convolution, treats the input data as if it were periodic. Instead of padding with zeros or a constant value, circular padding wraps the input data around, allowing features at the edges to interact as if they were adjacent.

Advantages of Circular Dilated Convolutions:

They efficiently record long-range relationships, which is essential for activities involving the comprehension of global context.

By minimizing boundary effects, circular padding enhances feature extraction in the vicinity of the input data's boundaries.

Dilated convolutions are effective for collecting both local and global characteristics because they introduce a broader receptive field without raising computing expense

(ii) Model Complexity: Our proposed CSDA-MOSNet architecture for multi-organ segmentation combines advanced components for efficient feature extraction, context integration, and spatial attention. Each module contributes to the overall complexity and performance of the segmentation model.

Circular Dilated Convolution Network

Circular dilated convolutions are integrated by the CDCNN to improve feature extraction and capture global context.The computational complexity of circular dilated convolutions in a CDCNN is typically represented as O(M×K^2×D), where: M is the number of input feature maps. K^2is the kernel size. D is the dilation rate.CDCNN optimizes its parameters to effectively extract both local information and long-range interdependence.

Multi Sparse Transformer Network

Transformer-based methods for feature refining and context aggregation are introduced by MSTN.The computational complexity of a transformer layer in MSTN can be expressed O(M^2×AL) where: M^2 is the sequence length (number of features).AL is the number of attention layers.To improve segmentation accuracy and capture intricate spatial connections, the transformer encoder-decoder design is used.

Dual Branch Spatial Attention Module

DBSAM uses dual-branch attention techniques to improve feature selection and spatial awareness.The computational complexity of a dual-branch spatial attention module depends on its design but is typically represented as O(M^2×AH), where: M^2 is the number of input feature maps, AH is the number of attention heads.The model incorporates spatial attention methods to reduce noise and highlight pertinent locations.

Conclusion

Poor detection accuracy, higher complexity, and varied sizes are the major issues in the multi organ segmentation task. So that, we design a novel model namedCSDA-MOSNet. Initially, we perform image pre-treatment for the acquired multi-organ images. The pre-processed images is fed in dual direction simultaneously to the CDCN where the boundary information, and long range dependencies are captured. The outputs are then fed to the zig-zag connected MSTN module where the long range temporal dependencies are captured to form global context feature map. The feature map is then processed by DBSAM where the channel and temporal correlations are extracted to provide final segmented output. We utilize two benchmark dataset such as BTCV and AMOS 22 respectively for examining the performance of the suggested project using a range of performance indicators. The findings demonstrate that the suggested work outperforms the current works in terms of performance.

In future we have planned to integrate reinforced models to the existing DL and transformer models. In addition to that, the security practices in the multi organ segmentation also have worth future impact.

Reference

1. Khanagar, S. B., Al-Ehaideb, A., Vishwanathaiah, S., Maganur, P. C., Patil, S., Naik, S., ... & Sarode, S. S. (2021). Scope and performance of artificial intelligence technology in orthodontic diagnosis, treatment planning, and clinical decision-making-a systematic review. Journal of dental sciences, 16(1), 482-492.

2. Peña-Romero, A. C., & Orenes-Piñero, E. (2022). Dual effect of immune cells within tumour microenvironment: pro-and anti-tumour effects and their triggers. Cancers, 14(7), 1681.

3. Zhang, J., Xie, Y., Xia, Y., & Shen, C. (2021). Dodnet: Learning to segment multi-organ and tumors from multiple partially labeled datasets. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 1195-1204).

4. Dai, X., Lei, Y., Wang, T., Dhabaan, A. H., McDonald, M., Beitler, J. J., ... & Yang, X. (2021). Head-and-neck organs-at-risk auto-delineation using dual pyramid networks for CBCT-guided adaptive radiotherapy. Physics in Medicine & Biology, 66(4), 045021.

5. Taleb, A., Lippert, C., Klein, T., & Nabi, M. (2021, June). Multimodal self-supervised learning for medical image analysis. In International conference on information processing in medical imaging (pp. 661-673). Cham: Springer International Publishing.

6. McHanwell, S., Davies, D. C., Morris, J., Parkin, I., Whiten, S., Atkinson, M., ... & Wilton, J. (2021). A core syllabus in anatomy for medical students-Adding common sense to need to know. European Journal of Anatomy, 11(S1), 3-18.

7. Yan, K., Cai, J., Jin, D., Miao, S., Guo, D., Harrison, A. P., ... & Lu, L. (2022). SAM: Self-supervised learning of pixel-wise anatomical embeddings in radiological images. IEEE Transactions on Medical Imaging, 41(10), 2658-2669.

8. Weiskopf, N., Edwards, L. J., Helms, G., Mohammadi, S., & Kirilina, E. (2021). Quantitative magnetic resonance imaging of brain anatomy and in vivo histology. Nature Reviews Physics, 3(8), 570-588.

9. Alhassan, E. A., Asaleye, J. A., Biniyat, J. K., Alhassan, T. R., & Olaoye, J. O. (2024). Tractor acquisition and agricultural performance among Nigerian farmers: Evidence from Co-integration Modeling technique. Heliyon, 10(2).

10. Mendoza Ladd, A., Casner, N., Cherukuri, S. V., Garcia, C., Padilla, O., Dwivedi, A., & Hakim, N. (2022). Fine needle biopsies of solid pancreatic lesions: tissue acquisition technique and needle design do not impact specimen adequacy. Digestive Diseases and Sciences, 1-8.

11. Malhotra, P., Gupta, S., Koundal, D., Zaguia, A., &Enbeyle, W. (2022). Deep neural networks for medical image segmentation. Journal of Healthcare Engineering, 2022.

12. Chen, Z., Duan, J., Kang, L., & Qiu, G. (2021). Class-imbalanced deep learning via a class-balanced ensemble. IEEE transactions on neural networks and learning systems, 33(10), 5626-5640.

13. Conze, P. H., Kavur, A. E., Cornec-Le Gall, E., Gezer, N. S., Le Meur, Y., Selver, M. A., & Rousseau, F. (2021). Abdominal multi-organ segmentation with cascaded convolutional and adversarial deep networks. Artificial Intelligence in Medicine, 117, 102109.

14. 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).

15. Feng, S., Zhou, Y., Zhang, X., Zhang, Y., & Wang, Y. (2021). Ms-kd: Multi-organ segmentation with multiple binary-labeled datasets. arXiv preprint arXiv:2108.02559.

16. Zhu, Y., Liang, Z., Yan, J., Chen, G., & Wang, X. (2021). ED-Net: Automatic building extraction from high-resolution aerial images with boundary information. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 4595-4606.

17. Chen, T., Gil-Garcia, J. R., & Gasco-Hernandez, M. (2022, June). Cross-Boundary Information Sharing Flows in Emergency Management: Proposing a Conceptual Framework. In DG. O 2022: The 23rd Annual International Conference on Digital Government Research (pp. 410-415).

18. Ulfers, A., Zeeden, C., Voigt, S., Abadi, M. S., &Wonik, T. (2022). Half-precession signals in Lake Ohrid (Balkan) and their spatio-temporal relations to climate records from the European realm. Quaternary Science Reviews, 280, 107413.

19. Zhu, Z., He, X., Qi, G., Li, Y., Cong, B., & Liu, Y. (2023). Brain tumor segmentation based on the fusion of deep semantics and edge information in multimodal MRI. Information Fusion, 91, 376-387.

20. Chen, H., Wu, C., Xu, Y., & Du, B. (2021). Unsupervised domain adaptation for semantic segmentation via low-level edge information transfer. arXiv preprint arXiv:2109.08912002E

21. Li, B., Xu, Y., Wang, Y., & Zhang, B. (2024). DECTNet: Dual Encoder Network combined convolution and Transformer architecture for medical image segmentation. Plos one, 19(4), e0301019.

22. Ye, Q., & Guo, L. (2023). Evaluation of Multi-indicator And Multi-organ Medical Image Segmentation Models. arXiv preprint arXiv:2306.00446.

23. Rickmann, A. M., Senapati, J., Kovalenko, O., Peters, A., Bamberg, F., &Wachinger, C. (2022). AbdomenNet: deep neural network for abdominal organ segmentation in epidemiologic imaging studies. BMC medical imaging, 22(1), 168.

24. 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).

25. Kanhere, A. U., Kulkarni, P., Yi, P. H., & Parekh, V. S. (2023). SegViz: A federated-learning based framework for multi-organ segmentation on heterogeneous data sets with partial annotations. arXiv preprint arXiv:2301.07074.

26. Ji, Y., Bai, H., Ge, C., Yang, J., Zhu, Y., Zhang, R., ... & Luo, P. (2022). Amos: A large-scale abdominal multi-organ benchmark for versatile medical image segmentation. Advances in Neural Information Processing Systems, 35, 36722-36732.

27. Hu, S., Liao, Z., & Xia, Y. (2022). Boundary-Aware Network for Abdominal Multi-Organ Segmentation. arXiv preprint arXiv:2208.13774.

28. Huang, Z., Deng, Z., Ye, J., Wang, H., Su, Y., Li, T., ... & Qiao, Y. (2023). A-Eval: A Benchmark for Cross-Dataset Evaluation of Abdominal Multi-Organ Segmentation. arXiv preprint arXiv:2309.03906.

29. Ciausu, C., Krishnaswamy, D., Billot, B., Pieper, S., Kikinis, R., & Fedorov, A. (2024). Towards Automatic Abdominal MRI Organ Segmentation: Leveraging Synthesized Data Generated From CT Labels. arXiv preprint arXiv:2403.15609.

30. Salimi, Y., Shiri, I., Mansouri, Z., & Zaidi, H. (2023). Deep learning-assisted multiple organ segmentation from whole-body CT images. medRxiv, 2023-10.

31. Pan, S., Chang, C. W., Wang, T., Wynne, J., Hu, M., Lei, Y., ... & Yang, X. (2023). Abdomen CT multi-organ segmentation using token-based MLP-Mixer. Medical Physics, 50(5), 3027-3038.

32. Lian, S., Li, L., Luo, Z., Zhong, Z., Wang, B., & Li, S. (2023). Learning multi-organ segmentation via partial-and mutual-prior from single-organ datasets. Biomedical Signal Processing and Control, 80, 104339.

33. Jeon, S. K., Joo, I., Park, J., Kim, J. M., Park, S. J., & Yoon, S. H. (2024). Fully-automated multi-organ segmentation tool applicable to both non-contrast and post-contrast abdominal CT: deep learning algorithm developed using dual-energy CT images. Scientific Reports, 14(1), 4378.

34. Shen, N., Wang, Z., Li, J., Gao, H., Lu, W., Hu, P., & Feng, L. (2023). Multi-organ segmentation network for abdominal CT images based on spatial attention and deformable convolution. Expert Systems with Applications, 211, 118625.

35. Li, X., Qin, X., Huang, C., Lu, Y., Cheng, J., Wang, L., ... & Yuan, C. A. (2023). SUnet: A multi-organ segmentation network based on multiple attention. Computers in Biology and Medicine, 167, 107596.

36. Liu, H., Xu, Z., Gao, R., Li, H., Wang, J., Chabin, G., ... & Grbic, S. (2024). Cosst: Multi-organ segmentation with partially labeled datasets using comprehensive supervisions and self-training. IEEE Transactions on Medical Imaging.

37. Xu, X., Deng, H. H., Gateno, J., & Yan, P. (2023). Federated multi-organ segmentation with inconsistent labels. IEEE transactions on medical imaging.

38. Zhang, Y., Li, X., Chen, H., Yuille, A. L., Liu, Y., & Zhou, Z. (2023, October). Continual learning for abdominal multi-organ and tumor segmentation. In International conference on medical image computing and computer-assisted intervention (pp. 35-45). Cham: Springer Nature Switzerland.

39. Salimi, Y., Shiri, I., Mansouri, Z., & Zaidi, H. (2022, November). Fully Automated Multi-Organ Segmentation in CT Images via Deep Neural Networks. In 2022 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC) (pp. 1-3). IEEE.

40. Dong, J., Cheng, G., Zhang, Y., Peng, C., Song, Y., Tong, R., ... & Chen, Y. W. (2023). Tailored multi-organ segmentation with model adaptation and ensemble. Computers in Biology and Medicine, 166, 107467