International Journal of Bioinformatics Research and Applications

ISSN online: 1744-5493 - ISSN print: 1744-5485
https://www.inderscience.com/ijbra

A modified UNet-based semantic segmentation architecture for pancreas tumour detection

Banavathu Sridevi, B. John Jaidhan

DOI: 10.1504/IJBRA.2024.10061615

Article History:
Received: 22 June 2023
Accepted: 19 July 2023
Published online: 14 March 2024

A modified UNet-based semantic segmentation architecture for pancreas tumour detection

Banavathu Sridevi* and B. John Jaidhan

Department of CSE,
GIT, GITAM (Deemed to be University),
Visakhaptnam, Andhrapradesh, India
Email: sbanavat@gitam.in
Email: jbeera@gitam.edu
*Corresponding author

Abstract: For computer aided diagnosis, computerised organ segmentation is a crucial but complicated task. The anatomy of the pancreas varies greatly and it is an abdominal organ. Especially when compared to other organs like the liver, heart, or kidneys, this prevents earlier segmentation approaches from obtaining high accuracy levels. To address this issue, we proposed a modification in UNet architecture called DAH-UNet that combines residual blocks, a rebuilt atrous spatial pyramid pooling (ASPP), and depth-wise convolutions. Also, a hybrid loss function which is explicitly aware of the boundaries is another thing we suggest. Experiments were conducted on two publicly available dataset and proved better in some metrics as compare to existing semantic segmentation models.

Keywords: pancreas tumour detection; UNet architecture; atrous spatial pyramid pooling; ASPP; depth-wise convolutions; semantic segmentation.

Reference to this paper should be made as follows: Sridevi, B. and Jaidhan, B.J. (2024) 'A modified UNet-based semantic segmentation architecture for pancreas tumour detection', *Int. J. Bioinformatics Research and Applications*, Vol. 20, No. 1, pp.1–20.

Biographical notes: Banavathu Sridevi is pursuing her PhD in GITAM (Deemed to be University), Visakhapatnam, AP, India; MTech and BTech from the AU College of Engineering and JNTU, Kakinada in 2016 and 2012 respectively. Her research area is image processing.

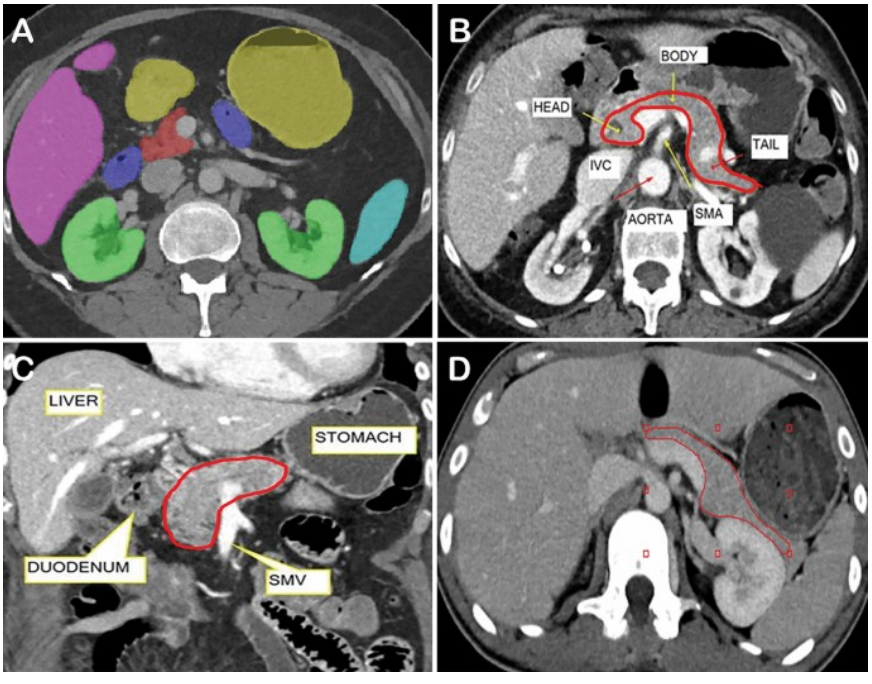
B. John Jaidhan received his PhD from the Andhra University College of Engineering, Visakhapatnam, AP, India. At present, he is working as an Associate Professor in GITAM (Deemed to be University), Visakhapatnam. He has teaching experience of 26 years. His interested areas are bio-informatics, embedded systems, IoT, cloud computing and image processing.

1 Introduction

For clinical applications, using computers to aid clinicians requires accurate pancreatic segmentation. Due of the incredibly small size, uneven shape, and border, it is not a simple process (Figure 1). Semantic segmentation is used to automatically segment the

pancreas in clinical abdominal computed tomography (CT) images. Traditional methods for segmenting the pancreas, such as pixel based (Yao et al., 2020), morphology-based (Chiranjeevi and Jena, 2017), graph-based, statistics-based (Mallouli, 2019), sparse segmentation (Tang et al., 2022) and super-pixel (Farag et al., 2014) and region based segmentation, frequently used straightforward models or were influenced by a synthetic inductive bias. Additionally, traditional methods frequently lack the flexibility needed to adjust to pixel-level segmentation tasks, particularly when the segmentation aim varies significantly. However, because of the deep learning (DL) method's inherent benefits, a convolutional neural network (CNN)-based approach was suggested as a promising approach for pancreatic segmentation. Then, using either pipelined or irregular in shape segmentation models, various research developed cutting-edge pancreatic segmentation methods based on CNNs. But prior pancreas segmentation studies (Yan and Zhang, 2021), which included 82 patients from the National Institutes of Health (NIH) clinical centre, were carried out on small research populations. For various and sizable datasets, clinical evaluation of the pancreas segmentation performance is required since DL approaches are responsive to the properties of the data that are contained in the model. To the best of our knowledge, there are not many DL studies on sizable CT datasets that encompass different pancreatic sizes.

Figure 1 Visual complexity of pancreas organ (red colour) showing various in size, depth and override with other organs (see online version for colours)



2 Related work

Recent studies have found that machine learning and DL based semantic segmentation networks perform better on medical image segmentation tasks than more conventional methods. DL algorithms have been utilised in earlier work to classify and segment pancreas images and ResNet50 was used to detect malignancy of tumour. Through classification and segmentation using DL models such as Resnet50, UNet and U-Net++. UNet might collect additional features from pancreas tissue with skip connections and acquire more precise information than FCN when combining features in the decoder section. Deep supervision can be integrated into the network to significantly enhance quality of the model and reduce the danger of divergence vanish during training (Yan and Zhang, 2021).

In general, we may categorise pancreatic segmentation techniques into two groups: 2D network-based techniques and 3D network-based techniques. The full CT volume is used as the input of 3D networks [such as 3D U-net (Karri and Jena, 2016) and V-net (Milletari et al., 2016)], which can collect the CT volume's 3D spatial information. The depth of the networks and the quantity of the feature maps, which are two important elements for performance enhancement, are yet constrained by the high demand for computer resources, particularly GPU memory (Song et al., 2017). The fact that there are so few CT images accessible for training causes another issue for 3D networks. The primary explanation is that hand labelling takes a considerable amount of money and time to complete. Slices are taken from the CT volume by making cuts along coronal, sagittal, and axial axes as in Figure 1. To create a 3D pancreatic volume, each slice from three views is analysed independently in the model, and the forecast outcomes from each perspective are combined via majority voting (Cai et al., 2018). The pancreatic only takes up just under 2% of the total CT volume, and the border between it and the organs and tissues around it is unclear, thus the model can be misled by the background region. A two-stage structure is considered in order to address this issue. A coarse-to-fine approach was suggested by Zhou et al. (2017) to lessen the contamination of the background region. First, the slices are coarsely segmented using a 2D FCN. The pancreatic area is then detected using the outcomes of the rough segmentation, and the CT volume is trimmed to create a tiny block that contains the pancreatic region. This block is then fed into a second 2D FCN for detailed segmentation. An end-to-end layered CNN-RNN classification model was presented in Cai et al. (2019). The segmentation outcomes are fed into a recurrent neural network after first training a 2D CNN to separate multi-layer neighbouring pancreatic areas (RNN). By combining the data from its parallel layers, RNN further produces high accuracy. The pancreas was segmented using two cascaded 2D FCN (Yu et al., 2018). Utilising an FCN, a rough segmentation probability map is first acquired. After that, a foreground translation module converts the probability map into a spatial weight map, which is then compounded by the input slice to produce a slice that contains the fine segmentation data. The full CT volume is clipped into a smaller volume that contains the pancreatic region in accordance with the fine segmentation probability map. Finally, the segmentation's loss function for both coarse and fine slicing is optimised.

In Li et al. (2022) developed a thin 3D voxel by synthesising three nearby CT scans, and suggested a comparable label mapping technique by using 2D CNN to segment lightweight 3D voxels along with 2.5D segmentation method was created called

multi-attention dual context network (MADC-Net) features. For the pancreatic tumour detection modified U-Net called MobileNet-V2 (MBU-Net) was proposed (Huang and Wu, 2022) and investigated the effect of data augmentation on NIH dataset. They also proved that MBU-Net model requires less parameters (6.30 M) during the training phase. By integrating a novel object identification method (FCN-guided region proposal network) and U-Net, authors (Deng et al., 2023) suggest to locate the pancreatitis regions to avoid the imbalance in dataset. The detector creates a fixed feature map of the regions affected by pancreatitis by using FCN. To improve the salient aspects of pancreas segmentation, Long et al. (2022) used an attention mechanism and to ensure complete fusion of surrounding contextual and geographical information and to increase segmentation accuracy, they subsequently combine the two modules. Dai et al. (2023) initially create candidate pancreatic regions using coarse segmentation using 2D UNet. They suggested integrating flexible convolution during fine segmentation step to address the pancreas' deformation issue. Additionally, they suggested combining local and global features using the scale inter-active fusion (SIF) module. Chen and Wan (2022) suggested a brand-new network called CTUNet that integrates Transformer and 3D U-Net and can automatically segment the pancreas with great accuracy. In order to coordinate global explicit features and direct the network learning, they implemented the transformer on skip connections and attention mechanism. The absolute pancreas position could be recorded via the residual transformer block (Qiu et al., 2023). Pooling-related erroneous pancreatic anatomy is addressed with a dual down-sampling block. The network concentrates on the pancreas boundaries due to the Hausdorff distance limitation. For the segmentation of the pancreatic, the residual transformer UNet is actually advised.

Fei and Luo (2022) suggest the DTUNet, which builds on the UNet platform to introduce the dense ASPP modules and transformer and sequentially mounts the two. Transformer creates a global receptive field by connecting each pixel of the input feature maps, capturing the global context, and accomplishing the development of long-range dependency. As part of a semi-supervised learning framework based on iterative uncertainty-guided pseudo-label refinement (Liu et al., 2022), they proposed a novel graph-enhanced pancreatic segmentation network (GEPS-Net). In order to concentrate on the spatial relationship information, it plugs a graph improvement module on top of the CNN-based U-Net. The aberrant pancreas is initially segmented based on a dual branch coding network (DB-Net) and one branch of the encoder component extracts the semantic aspects of the pancreas and its surrounds, and the other branch uses wide-channel convolution and minimal down sampling to capture the complicated pancreas (Zhou et al., 2023). A U-Net decoder is utilised to combine the various feature maps acquired by the two branches. The exterior contours of the pancreas with lesion occupancy are precisely delineated by M3Net (Qu et al., 2022), which can segment both normal and diseased pancreases. To investigate location and channel non-Euclidean relationships between phases, cross-phase non-local attention is encouraged. For pancreas segmentation, we suggest a cascaded multi-scale feature calibration UNet (CMFCUNet) where the multi-scale features in the core of each scaled segmentation are calibrated vertically in a pixel-wise manner (Qiu et al., 2022). Additionally, a dual enhancement module is used to connect the coarse-scaled segmentation with the fine-scaled segmentation (DEM). To maintain and spread the pancreas' form traits, low-level and high-level features are gradually joined with FPF-Net (Chen et al., 2022). To address the issues brought on by the diversity in form and small size of the pancreas, attentional

feature fusion (AFF) is utilised in place of context-unaware addition or concatenation. To utilise long-range dependencies and multi-scale spatial features they used coordinate and multi-scale spatial attention (CMSA). Cui et al. (2022) suggested a SCU-Net++, an enhanced channel attention mechanism and Laplacian sharpening filter-based semantic segmentation model: Sharpening filters are utilised to construct dense skip connections in order to close any semantic gaps, and channel attention modules are employed to train the model to focus more on feature maps that are relevant to our pixel-level classification objective.

Below is a list of our key contributions:

- A modification to UNet architecture with Atrous and depth wise convolution.
- We present a brand-new loss function that is based on metric recall, focal and cross entropy loss. Recall loss compares the instantaneous training recall performance of each class with its standard cross entropy and focal loss.
- The suggested hybrid loss develops a better semantic segmentation model that offers enhanced and balanced accuracy and IOU performance.
- The suggested loss enhances feature learning for image segmentation along with the proposed modification in UNet architecture.

The rest of the paper is organised as followed: In Section 3, detailed architecture of proposed semantic segmentation model followed with results and discussion in Section 4. Finally, conclusions at end of the Section 5.

3 Methods and materials

In this paper, we present depthwise convolution, atrous and hybrid loss (DAH-UNet), an upgraded version of the UNet-based backbone network that combines residual blocks, a rebuilt atrous spatial pyramid pooling (ASPP), and depth-wise convolutions. Also, a hybrid loss function which is explicitly aware of the boundary is another thing we suggest.

3.1 Model architecture

Our model utilises a U-shaped decoder and encoder layout, as seen in Figure 2, which enhances the fundamental UNet layout in a number of ways.

- First, we use group convolution to substitute the ordinary convolutions in both decoder and encoder section of layout, with a notable exception of the first most layer, such that the intra and inter channel correlation information is individually extracted throughout each level's of encoding process (Zhou et al., 2018). In order to separately capture the specific changes among neighbouring slices of images, which is beneficial for more precise segmentation, we use the overlay of neighbouring slices holding the foreground as the input of our model, which is based on this structure. In principle, the channels should be handled individually; it is preferable to avoid mapping them together.

- Second, we have included a residual architecture (Hu et al., 2020) between neighbouring convolution blocks, which helps to reduce the loss incurred during down-sampling process.
- Third, to minimise risk of data loss while down-sampling, we established a bottleneck layer utilising ASPP (Chen et al., 2017), which is crucial for retrieving multi-level relevant information. Convolution procedures are carried out simultaneously on the feature maps created during the encoding stage with various dilation rates in order to gather the environment of the image at several scales and produce more precise foreground location data (Huang et al., 2020).

Existing approaches might not be capable of recognising the presence of the pancreas when given a hard input since the occupancy of pancreas in CT images is very small and is override with the neighbour organs like liver, stomach, etc. and also size of pancreas is flexible and variable with respect to patients. For such a small and complex structured organs, extraction of multi-level and single level contextual information is crucial. In the decoding stage of model, we applied group deconvolution (Roth et al., 2018) at every stage of each layers to get back original input image size and 1 by 1 convolution was applied to limit the number of feature maps to two.

3.2 Depth-wise separable convolution

At, encoder section of model, we substitute the ordinary convolution with a particular group convolution called depth-wise separable convolution. Ordinary convolution is a simultaneous mapping of spatial and correlation information of an individual channels (Alhichri et al., 2021). Although these two types of input are connected, depth-wise convolution in inception allows the correlations between the two to be broken (Roth et al., 2015). According to the Inception hypothesis, the two correlations can really be mapped individually to produce superior results since; they are independent of one another (Naidu et al., 2018). As pancreas images have many slices, an individual mapping of information across channels makes more sense. In our approach, we employ the Exception which is extreme case of Inception method, in which; the number of groups in the group convolution is equal to the number of input channels. With this process, totally decoupling of spatial correlation between channels and the inter-channel correlation occur (Yang et al., 2020). The input feature map undergoes channel-by-channel linear transformation using 1 by 1 convolution, and the resultant feature map is fed through a series of 3 by 3 convolutions. Because there are exactly as many groups in our grouped convolution as there are input channels, each filter in this convolution process has a convolution kernel of 3 by 3, meaning that each channel of the input feature map is only convolved by one kernel with a size of 3 by 3 by 1. The output feature map is created by stacking the outputs of these filters. Because there are exactly as many groups in grouped convolution as there are input channels, each filter in this convolution process has a convolution kernel of 3 by 3, meaning that each channel of the input feature map is only convolved by one kernel with a size of 3 by 3 by 1. The output feature map is created by stacking the outputs of these filters.

Figure 2 Proposed model architecture (see online version for colours)

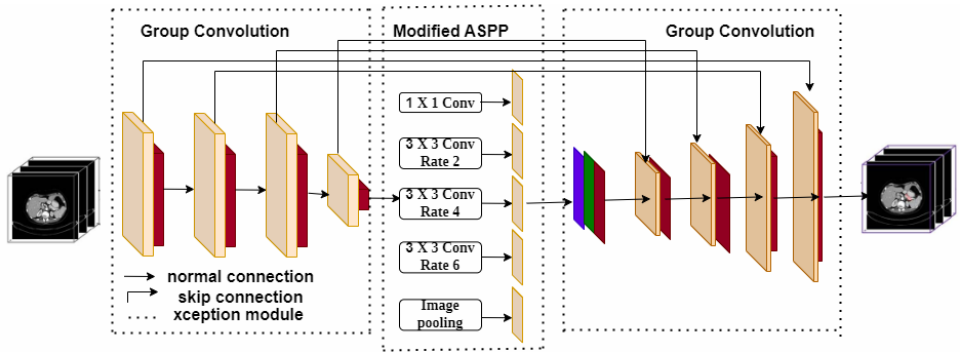
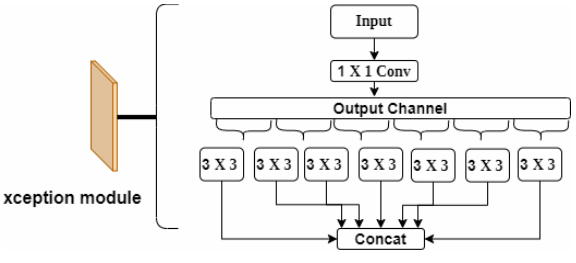


Figure 3 Xception module (see online version for colours)



In calculation of model parameters, lets assume P is the feature maps at input layer of model and Q feature maps at output layer with kernel size of convolution layer is 3, then number of parameters of model is $3 \times 3 \times P \times Q$. In case of depth-wise separable convolution, the number of parameters are addition of parameters of depth-wise and point wise ($Pg = P_{depth–wise} + P_{point–wise} = (3 \times 3 \times Q) + (1 \times 1 \times Q \times P)$. From these observations, our model parameters are less as compared to ordinary UNet and it enhances the feature maps which help for good classification of tumour and non-tumour regions of pancreas. In addition to this, our model has two level down-sampling with replacement of ordinary convolution with depth-wise separable structure Xception as in Figure 3. To achieve information decoupling, convolution kernels with the same amount of input channels are employed in each down-sampling phase, followed by an ordinary convolution to increase the amount of feature maps produced. If ordinary convolution is being used entirely, a whole down-sampling approach involves 1,261,532 3×3 convolution kernels, whereas our enhanced structure requires only 679,329 3×3 convolution kernels.

3.3 ASPP module

Pancreas images typically have slightly out of focus boundaries and are easily confused with neighbouring soft tissue structures, especially because it holds a comparatively small area in a CT image with such a complex background and less than 1.5% in a 2D image. This makes determining whether or not the pancreas occurs in the image much more difficult. Just above all, the existing models are incapable of extracting sufficient information about the position of the pancreas, which is heavily dependent on the image's

overall context. To enhance the capability of extracting features in our architecture, we use an ASPP plugin with atrous convolution. The ASPP plugin, uses numerous concurrent atrous convolution layers with various sampling rates and takes its design cues from the spatial pyramid. At various scales, the feature map simultaneously captures the context. We presume that the deep and superficial features of medical images are significant in the case where the medical image does not already consist noise, any unwanted information or complex background, and that the fusion of various levels of features can result in improved judgment in this situation. The ASPP module that we employ mostly consists of the components shown in Figure 3:

- This architecture has one ordinary and three atrous convolution layers with size 1 × 1 and 3 × 3 respectively. Filters will deteriorate into a straightforward 1 × 1 convolution with only the filter centre functioning when the dilated rate is almost equal to the feature map size. The dilated ratio of the original module is therefore scaled to (2, 4, 6).
- Features of the image are obtained by applying average pooling globally and these extracted features are passed through a convolution layer of size 1 × 1 which performs bi-linear interpolation to get back the original image of same size/resolution as like input image.
- In order to create a new extracted features with 256 layers, the four different types of extracted features from the previous two steps are combined in the channel dimension and thereafter routed through 1 × 1 convolution for fusion.
- The ASPP unit, which can collect multi-level location information more effectively and has higher classification and learning skills to recognise and localise the pancreas, partially corrects the standard UNet's deficiency in describing information. Additionally, if the dilate rate approaches or even exceeds the size of the input feature map, it will collapse into 1 × 1 convolution, and a dilate rate that is too high will not allow for pixel-level output. For these reasons, we adopt a reduced dilate rate of (2, 4, 6).

3.4 Loss function

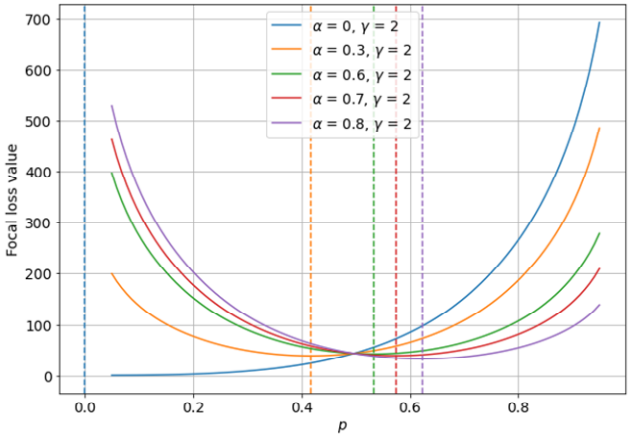
Loss functions determine the amount by which the forecasts differ from the actual. Lower loss values signify higher accuracy of model, whilst higher loss values indicate predictions of model are fairly correct. Always model should looks for minimisation of loss function as feasible, ideally it must be near to zero. Models takes the loss function to learn the trainable parameters, like biases and weights and these are updated in each iteration based on gradient descent process. Cross entropy is one loss function mostly used for binary classification but some times fails for multi-class classification problems. It defined from the concept of entropy in mathematics which is defined as:

$$Entropy = P_i \log(P_i) \tag{1}$$

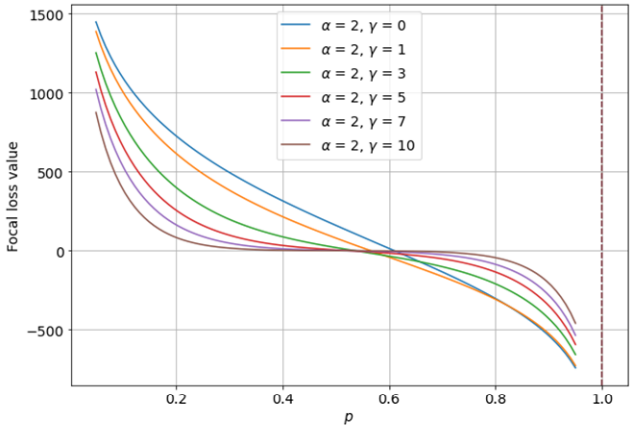
$$Cross\text{-}entropy = \sum_{i}^{n} Y_i \log(P_i) \tag{2}$$

where Y is the ground true or actual output and P is the model predicated output. Due to the fact that probabilities might range from 0 to 1, the logarithm of this produce negative value so in order make the entropy positive we included negative in front of equation. In case of class imbalanced problems, as loss function followed by gradient descent mostly focus on majority class, which will cause the weights to update in a way that increases the model's confidence in forecasting the majority class while decreasing its concentration on the minority classes. Solution to this issue is focal loss (Lin et al., 2017).

Figure 4 Focal loss variation with respect to (a) α, (b) γ (see online version for colours)



(a)



(b)

3.4.1 Focal loss

By concentrating on the instances where the model fails instead of the ones where it can reliably forecast, focal loss makes sure that forecasts on challenging examples get better over time instead of getting too confident with simple ones. This is accomplished by a process termed 'down weighting' in focal loss. By lessening the impact of simple

instances on the loss function, down weighting emphasizes the importance of difficult examples. By incorporating a modulating factor (μ) $\mu = (1 - P_i)^{\gamma}$ into the cross-entropy loss, this method can be put into practice.

$$Focal\ loss = \sum_{i}^{n} \alpha (1 - P_i)^{\gamma} \log(P_i) \tag{3}$$

where the cross validation tunable focusing parameter is called γ (gamma) and α is the weighing factor. The behaviour of focal loss for various values of γ and α is depicted in Figure 4. From figure, the following observations are made:

- Since the p_i is low for the sample that was incorrectly classified, the μ is near to or exactly 1 which leads loss function remains unaltered and became a cross-entropy loss.
- The μ will tend to zero as the model's confidence level rises, which is indicated by p_i is equal to 1, which will reduce the loss amount for correctly categorised cases. In order to lessen the impact of the simple instances on the loss function, γ re-scale the μ such that the easy instances are down-weighted more than the hard ones. To our dataset focal loss performance is better with γ value equal to 2.
- Focal loss is the same as cross entropy when γ is equal to 0.

3.4.2 Recall loss

To solve the imbalance dataset issue, we used a novel recall-based performance-balanced loss called recall loss. Based on the value of recall of model during training, the model weights are updated towards minimisation of loss function of that class. As opposed to the hard example mining method in the focal loss, it is an illustration of hard class mining. The recall loss, in contrast to focal loss and other losses, dynamically modifies its weights with training based on per-class recall value. At the cost of intersection over union (IOU), which takes false positives into account in semantic segmentation, the CB loss increases accuracy. The precision and recall of each class may be successfully balanced by our recall loss, which enhances accuracy while maintaining a competitive IOU.

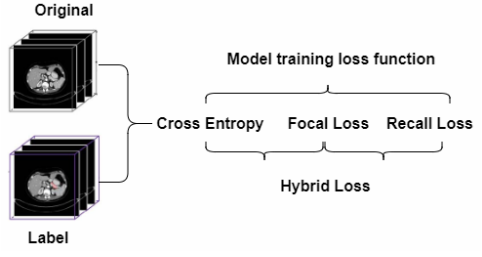
$$Recall\ loss = -\sum_{i=1}^{C} \frac{FN_C}{FN_C + TP_C} N_C \log(P^C) \tag{4}$$

where NC, FN_C and TP_C are number of samples, false positive and true positive rate of class C respectively. The same formula is applicable for other classes if applicable. It is possible to use the recall-loss as the standard cross entropy which is weighted by the class-wise FN_C. The second realisation is that minority classes, which have greater FN_C, are probably harder to categorise than large classes, which have lower FN_C. Gradients of minority classes will therefore be raised, and gradients of majority classes will be repressed, analogous to inverse frequency loss (Tian et al., 2020).

3.4.3 Hybrid loss function

In our work we took the advantages and disadvantages of above said three loss functions for better training of model and better performance of the model. In our case, the loss function is the sum of above three loss functions as in Figure 5.

Figure 5 Proposed training loss function



4 Experimental results

4.1 Dataset and settings

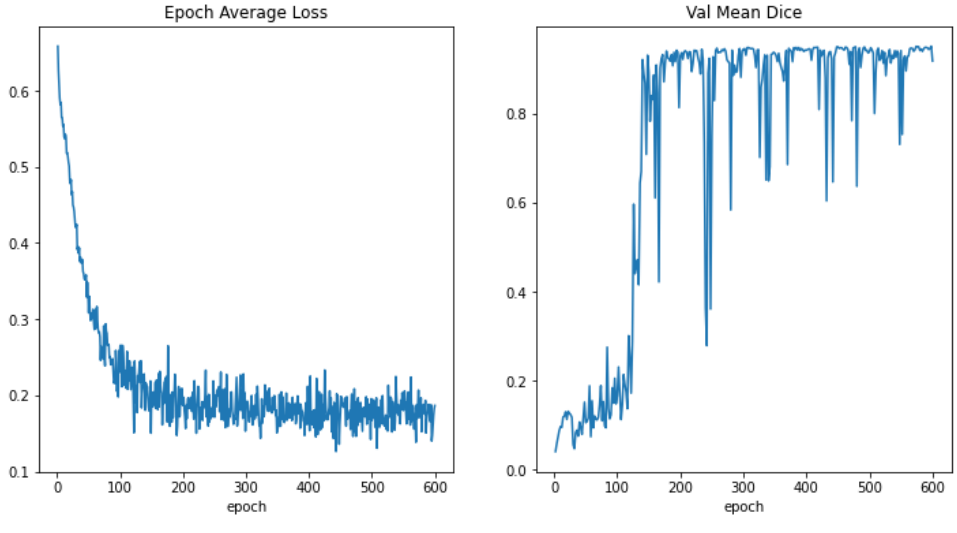
On the TCIA Pancreas-CT dataset (https://wiki.cancerimagingarchive.net/display/Public/Pancreas-CT) provided by NIH and abdominal contrast-enhanced CT dataset (https://chaos.grand-challenge.org/Data/), we contrasted the most recent methodologies. There are 82 CT images in the NIH collection, representing 53 men and 27 women between the ages of 18 and 76 respectively. Each volume of image has a size of 512 by 512 with variable length of volumes in each image with specific thickness between slices. Scanners made by Philips and Siemens are used to collect all quantities. Our studies employ two-fold-cross-validation and a random sample of 82 patients, 41 samples are used for test and remaining for validation. To speed up the training process and quick convergence of model, the proposed model followed a strategy of Naidu et al. (2018) in fixing the initial weights. The model starts with learning rate of 10^{-3} and is gradually reduced by 0.5 for every 50 epochs during the training phase using Adam (Lin et al., 2017) as an optimisation technique for gradient. All the images and respective labels are resized to 96 × 96 × 96 irrespective of their volume sizes during per-processing stage of dataset.

4.2 Implementation details

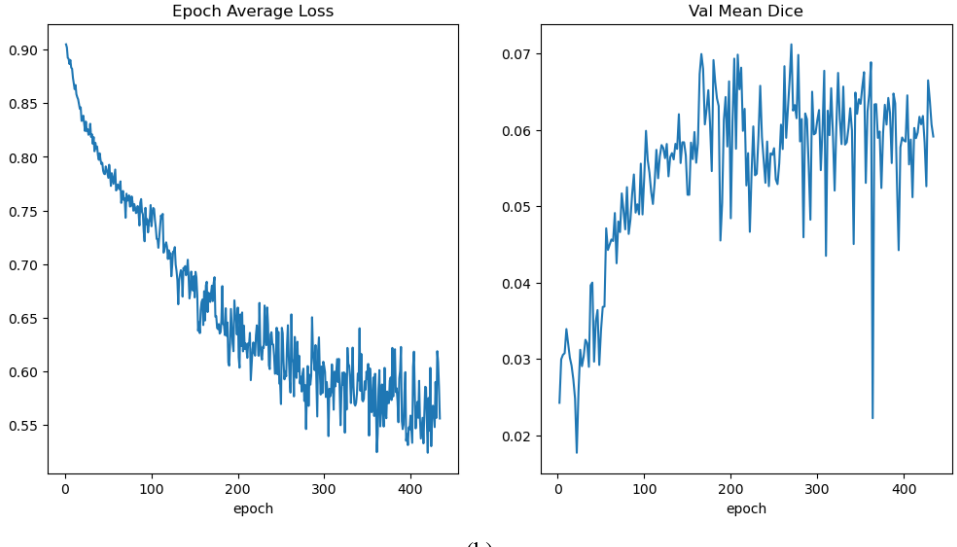
We use the PyTorch platform in windows 11 to execute our Python 3.7 investigations. On a single Nvidia RTX A3000 GPU (graphics card), all models have been trained in batches of eight. The models are trained over 600 and 460 iterations for dataset (https://wiki.cancerimagingarchive.net/display/Public/Pancreas-CT) and dataset (https://chaos.grand-challenge.org/Data/) respectively. In terms of the data augmentation technique, we employ scale-intensity-ranged with minimum and maximum values of –57 and 164 respectively, foreground of images are cropped with crop foreground transform, and composed cascaded transforms like orientation, spacing, rotation and crop form

MONAI transforms. The average loss and mean dice graphs from the model training and cross validation phases carried out on the training and validation image sets are shown in Figure 6.

Figure 6 Average loss and mean dice for (a) dataset (https://wiki.cancerimagingarchive.net/display/Public/Pancreas-CT), (b) dataset (https://chaos.grand-challenge.org/Data/) (see online version for colours)



(a)



(b)

4.3 Metrics of measures

- *Jaccard index or IoU:* it is basically a way to measure the amount of overlap amongst our prediction output and the target mask. This measure has a tight relationship to the dice score (DSC), which is frequently employed while training as a loss function.
- *Average Hausdorff distance:* between original image and predicated images is calculated by dividing the total minimum distances between all points in original and predicated by the total number of points in original. As a more reliable approximation of the maximum deviation, the 95% Hausdorff distance95 (Karri, 2021) has indeed been utilised. Though its meaning is less clear, the 95% Hausdorff distance is generally understood to be the 95th percentile among ordered distance metrics.
- *Average symmetric surface distance (ASSD):* the average of all the distances between predicted on the boundary of a region that was segmented correctly and the boundary of the original, and vice versa. Additional metrics are tabulated in Table 1.

Table 1 Metrics and formulas

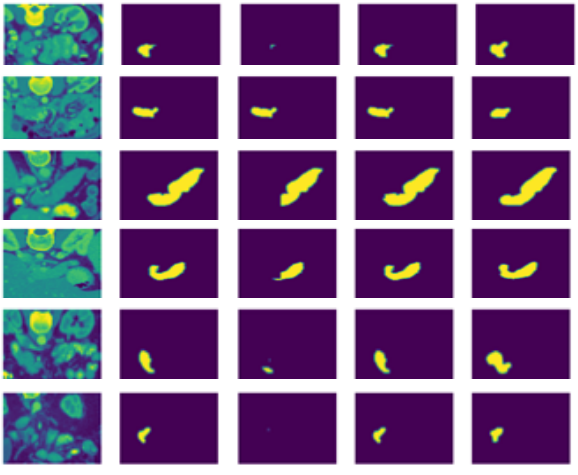
Metrics	*Description*	*Equation*
Accuracy (ACC)	The percentage of input images' weight that is correctly segmented is measured by accuracy.	$Accuracy = \frac{T_P + T_N}{T_P + F_P + T_N + F_N} \times 100$
Specificity (SP)	The segmentation process of input images proportion is how exactly performed for negative results are measured by the specificity.	$Specificity = \frac{T_N}{T_N + F_P} \times 100$
Sensitivity (SE)	The segmentation process of input images' proportion is how exactly performed for positive results are measured by the sensitivity.	$Sensitivity = \frac{T_P}{F_N + T_P}$
Positive prediction value (PPV)	The number of pixels obtained in the segmented images as the positive which are actually positive.	$PPV = \frac{TP}{FP + TP}$
Precision	The quality of a positive prediction made by the model.	$Precision = \frac{TP}{TP + FP}$
Recall	The ability of a model to identify only the relevant data points	$Recall = \frac{TP}{TP + FN}$
F1-score	It combines the precision and recall scores of a model	$F1\text{-}score = 2 * \frac{Precision * Recall}{Precision + Recall}$

4.4 Discussion

The disintegration and consumption of sugar levels and nutrients are significantly influenced by the pancreas, a vital metabolic organ located in the abdomen. For doctors, accurate pancreatic segmentation can offer useful data. We developed a specific strategy in developing the model in encoder and decoder section of ordinary UNet to accomplish

explicit localisation of the pancreas and suggested a hybrid loss to drive the model to concentrate its efforts to boundary pixels in order to overcome the inaccuracy of coarse-to-fine approaches and ambiguous boundaries in the pancreatic segmentation. As per the experiments performed on different datasets, the suggested methodology showed better performance than other state of art segmentation of pancreas methods on the two dataset with cross-validation rate of 2 without the aid of clear and specific pancreas localisation. This modification in encoder and decoder section can implicitly localise and highlight the pancreas regions, and thus improve the depiction of pancreatic features. Additionally, Tables 2 and 3 demonstrates that the proposed algorithm outperforms the existing pancreatic segmentation techniques in terms of average DSC, Jaccard coefficient, Hausdorff Distance95 (HD95) and ASSD. The DSC and Jaccard index measures were used to compare the segmentation results quantitatively to the ground truth. The two metrics' values fell into the range of 0% to 100%, with 0% signifying no overlap between the two segmented zones and 100% signifying an exact match between the two segmentation. Table 2 and Table 3 contains the results of analyses of segmented images made using DSC. Tables indicates that nearly all of the segmented images overlapped with the ground truth images, the dice scores for the hybrid procedure were close to 90%. For both scenarios, the average dice score for the proposed model was 86%–89%. The proposed model average Jaccard coefficient for the two scenarios was 76 and 81% respectively, lower HD95 and lower ASSD.

Figure 7 Outcomes of different models on dataset (https://wiki.cancerimagingarchive.net/display/Public/Pancreas-CT): from left to right image, ground truth, predicted with UNet and DAH-UNet (see online version for colours)



Additionally, utilising the assessment metrics precision, accuracy, sensitivity, specificity, F1-score, PPV, recall, and, which are presented in Table 4 and Table 5, the outcomes of our segmentation model were quantitatively assessed. The estimated values have nearly approached 91% accuracy with only minor numerical variations, demonstrating the general viability of the network for segmentation. It is demonstrated that the model results for the other indices all exceeded 70%, demonstrating the viability and efficiency of the suggested approach. The end results for the synthesis of proposed models are better balanced in all aspects when compared to the outputs of the six trained UNets, even

though the significant improvement may outperform well on some assessment criteria. As demonstrated in Figure 7 and Figure 8, the efficiency of model segmentation is not as excellent as that of other segmentation models because pancreas tissues are more complex than other tissues like stomach, liver and other organs, which are distinguished by intricate shape, uneven boundary, and quasi pixel intensity.

Table 2 Various models metrics on dataset

Model	*DSC*	*Jaccard coefficient*	*Hausdorff Distance95*	*ASSD*
UNet++ (Zhou et al., 2018)	0.727	0.571	8.000	2.806
SCU-Net++ (Cui et al., 2022)	0.738	0.585	4.899	1.280
FPF-Net (Chen et al., 2022)	0.790	0.653	5.196	1.599
CMFCUNet (Qiu et al., 2022)	0.797	0.663	6.164	1.979
M3Net (Qu et al., 2022)	0.833	0.714	3.606	1.076
RTUNet (Qiu et al., 2023)	0.878	0.782	4.123	1.046
DAH-UNet	0.897	0.813	2.449	0.813

Source: https://wiki.cancerimagingarchive.net/display/Public/Pancreas-CT

Table 3 Various models metrics on dataset

Model	*DSC*	*Jaccard coefficient*	*Hausdorff Distance95*	*ASSD*
UNet++ (Zhou et al., 2018)	0.469	0.306	24.755	1.846
SCU-Net++ (Cui et al., 2022)	0.754	0.605	12.247	1.212
FPF-Net (Cui et al., 2022)	0.774	0.631	5.745	1.277
CMFCUNet (Qiu et al., 2022)	0.841	0.725	4.123	1.060
M3Net (Qu et al., 2022)	0.844	0.730	3.742	1.208
RTUNet (Qiu et al., 2023)	0.854	0.745	3.464	1.242
DAH-UNet	0.869	0.769	3.317	1.060

Source: https://chaos.grand-challenge.org/Data/

Table 4 Various models metrics on dataset

Model	*ACC*	*Precision*	*Recall*	*Sensitivity*	*Specificity*	*F1-score*	*PPV*
UNet++ (Zhou et al., 2018)	0.648	0.924	0.759	0.759	1.000	0.833	0.924
SCU-Net++ (Cui et al., 2022)	0.707	0.723	0.889	0.889	0.997	0.797	0.723
FPF-Net (Chen et al., 2022)	0.754	0.720	0.734	0.734	0.996	0.727	0.720
CMFCUNet (Qiu et al., 2022)	0.775	0.931	0.865	0.865	0.999	0.897	0.931
M3Net (Qu et al., 2022)	0.794	0.782	0.799	0.799	0.997	0.790	0.782
RTUNet (Qiu et al., 2023)	0.853	0.847	0.910	0.910	0.997	0.878	0.847
DAH-UNet	0.869	0.687	0.798	0.798	0.997	0.738	0.687

Source: https://wiki.cancerimagingarchive.net/display/Public/Pancreas-CT

Figure 8 Outcomes of different models on dataset (https://chaos.grand-challenge.org/Data/): from left to right image, ground truth, predicted with UNet and DAH-UNet (see online version for colours)

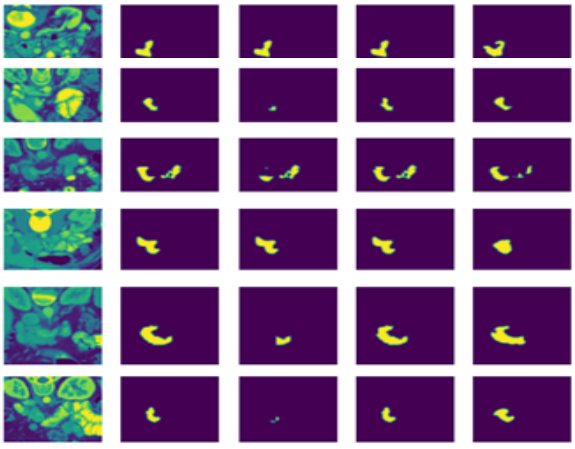


Table 5 Various models metrics on dataset

Model	*ACC*	*Precision*	*Recall*	*Sensitivity*	*Specificity*	*F1-score*	*PPV*
UNet++ (Zhou et al., 2018)	0.734	0.835	0.688	0.688	0.999	0.754	0.835
SCU-Net++ (Cui et al., 2022)	0.759	0.842	0.898	0.898	0.998	0.869	0.842
FPF-Net (Chen et al., 2022)	0.798	0.834	0.854	0.854	0.998	0.844	0.834
CMFCUNet (Qiu et al., 2022)	0.799	0.841	0.867	0.867	0.998	0.854	0.841
M3Net (Qu et al., 2022)	0.865	0.771	0.777	0.777	0.997	0.774	0.771
RTUNet (Qiu et al., 2023)	0.889	0.892	0.318	0.318	0.999	0.469	0.892
DAH-UNet	0.910	0.948	0.755	0.755	0.999	0.841	0.948

Source: https://chaos.grand-challenge.org/Data/.

Table 6 Effect of loss function on proposed model on dataset

Loss function	*DSC*	*Jaccard coefficient*	*Hausdorff Distance95*	*ASSD*
Cross entropy	0.754	0.547	11.045	2.961
Focal loss	0.775	0.632	8.602	1.629
Recall loss	0.794	0.658	5.099	1.280
Hybrid loss	0.853	0.744	3.162	0.977

Source: https://wiki.cancerimagingarchive.net/display/Public/Pancreas-CT

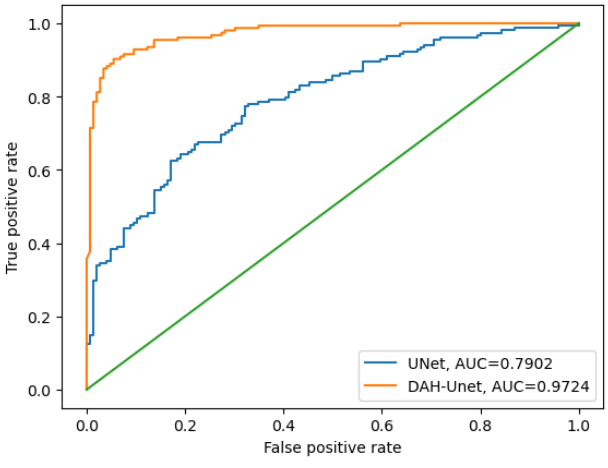
This finding suggests that the proposed model architecture instantaneously accumulate contextual information over feature points, use spatial information to encapsulate pancreas features, and enhance the model performance. Overall, the suggested algorithm

not only maintains a high level of pancreatic segmentation accuracy, but also increases the effectiveness of pancreas segmentation as in Table 4 and Table 5.

4.4.1 Importance of loss functions

Loss functions are crucial in determining how well a model performs. It is impossible to select a universal loss function for difficult tasks like segmentation. Most of the time, it is determined by the characteristics of the training dataset, such as its prevalence, deviation, constraints, etc. Any of the loss functions listed in table performs optimally in all use scenarios. However, we may assert that hybrid loss functions perform better with extremely unbalanced pancreas segmentation as in Table 6.

Figure 9 AUC curve for dataset (see online version for colours)



Source: https://wiki.cancerimagingarchive.net/display/Public/Pancreas-CT

As shown in Figure 9, the quantitative statistical result of the ROC curve shows that proposed model of segmentation outperforms the existing UNet based segmentation models. As seen in Figure 9, the proposed model exhibits improved performance as indicated by the area under the ROC curve. The effectiveness of the system improves as the curve gets progressively closer to the top left corner. The area under the curve (AUC) value, which equals 1 for an ideal system, is the most commonly used performance metric derived from the ROC curve.

5 Conclusions

On two publicly available datasets, with data augmentation on the training set, the proposed RTUNet, DAH-UNet and other models were assessed. The outcomes of the trial were contrasted with those of cutting-edge techniques. The findings showed that the suggested DAH-UNet exhibits greater dice, Jaccard, and Recall than other models with less computing parameters, and offers a potential strategy for semantic segmentation of tiny tissues. Visual examination of the overlapped maps reveals that the suggested

DAH-UNet can successfully fit the human segmentation, further supporting the efficacy of our approach. In comparison to state-of-the-art models, the suggested DAH-UNet model has high accuracy even pancreas differs greatly in size, position and shape. Our study does have certain limitations. Although the suggested DAH-UNet Net's parameters are adequate, not all of its outcomes are the best when considered to related research.

References

Alhichri, H., Alswayed, A.S., Bazi, Y., Ammour, N. and Alajlan, N.A. (2021) 'Classification of remote sensing images using EfficientNet-B3 CNN model with attention', *IEEE Access*, Vol. 9, pp.14078–14094, DOI: 10.1109/access.2021.3051085.

Cai, J., Lu, L., Xing, F. and Yang, L. (2018) *Pancreas Segmentation in CT and MRI Images via Domain Specific Network Designing and Recurrent Neural Contextual Learning*, arXiv 2018. arXiv preprint arXiv:1803.11303.

Cai, J., Lu, L., Xing, F. and Yang, L. (2019) 'Pancreas segmentation in CT and MRI via task-specific network design and recurrent neural contextual learning', in *Deep Learning and Convolutional Neural Networks for Medical Imaging and Clinical Informatics*, Springer, Cham, pp.3–21.

Chen, H., Liu, Y. and Shi, Z. (2022) 'FPF-Net: feature propagation and fusion based on attention mechanism for pancreas segmentation', *Multimedia Systems*, April, Vol. 29, No. 2, pp.525–538.

Chen, L. and Wan, L. (2022) 'CTUNet: automatic pancreas segmentation using a channel-wise transformer and 3D U-Net', *The Visual Computer*, Vol. 38, pp.1–15.

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K. and Yuille, A.L. (2017) 'Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFS', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 40, No. 4, pp.834–848.

Chiranjeevi, K. and Jena, U. (2017) 'Hybrid gravitational search and pattern search-based image thresholding by optimising Shannon and fuzzy entropy for image compression', *International Journal of Image and Data Fusion*, Vol. 8, No. 3, pp.236–269.

Cui, H., Pan, H. and Zhang, K. (2022) 'SCU-Net++: a nested U-Net based on sharpening filter and channel attention mechanism', *Wireless Communications and Mobile Computing*, Hindawi Wireless Communications and Mobile Computing, Vol. 2022, Article ID 2848365, 8pp, https://doi.org/10.1155/2022/2848365.

Dai, S., Zhu, Y., Jiang, X., Yu, F., Lin, J. and Yang, D. (2023) 'TD-Net: trans-deformer network for automatic pancreas segmentation', *Neurocomputing*, Vol. 517, pp.279–293, ISSN: 0925-2312, https://doi.org/10.1016/j.neucom.2022.10.060.

Deng, Y., Lan, L., You, L., Chen, K., Peng, L., Zhao, W. and Zhou, X. (2023) 'Automated CT pancreas segmentation for acute pancreatitis patients by combining a novel object detection approach and U-Net', *Biomedical Signal Processing and Control*, Vol. 81, p.104430, ISSN: 1746-8094, https://doi.org/10.1016/j.bspc.2022.104430.

Farag, A., Lu, L., Turkbey, E., Liu, J. and Summers, R.M. (2014) 'A bottom-up approach for automatic pancreas segmentation in abdominal CT scans', in *International MICCAI Workshop on Computational and Clinical Challenges in Abdominal Imaging*, September, Springer, Cham, pp.103–113.

Fei, C. and Luo, J. (2022) 'DTUnet: a transformer-based UNet combined with DenseASPP module for pancreas segmentation', in *2022 15th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, IEEE, November, pp.1–7.

Hu, P., Li, X., Tian, Y., Tang, T., Zhou, T., Bai, X. and Li, J. (2020) 'Automatic pancreas segmentation in CT images with distance-based saliency-aware DenseASPP network', *IEEE Journal of Biomedical and Health Informatics*, Vol. 25, No. 5, pp.1601–1611.

Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y. and Wu, J. (2020) 'Unet 3+: a full-scale connected unet for medical image segmentation', in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, May, pp.1055–1059.

Huang, M.L. and Wu, Y.Z. (2022) 'Semantic segmentation of pancreatic medical images by using convolutional neural network', *Biomedical Signal Processing and Control*, Vol. 73, p.103458, ISSN: 1746-8094, https://doi.org/10.1016/j.bspc.2021.103458.

Karri, C. (2021) 'Secure robot face recognition in cloud environments', *Multimedia Tools and Applications*, Vol. 80, No. 12, pp.18611–18626.

Karri, C. and Jena, U. (2016) 'Fast vector quantization using a bat algorithm for image compression', *Engineering Science and Technology, an International Journal*, Vol. 19, No. 2, pp.769–781.

Li, J., Liao, G., Sun, W., Sun, J., Sheng, T., Zhu, K. and Zhang, Y. (2022) 'A 2.5 D semantic segmentation of the pancreas using attention guided dual context embedded U-Net', *Neurocomputing*, Vol. 480, pp.14–26.

Lin, T.Y., Goyal, P., Girshick, R., He, K. and Dollár, P. (2017) 'Focal loss for dense object detection', in *Proceedings of the IEEE International Conference on Computer Vision*, pp.2980–2988.

Liu, S., Liang, S., Huang, X., Yuan, X., Zhong, T. and Zhang, Y. (2022) 'Graph-enhanced U-Net for semi-supervised segmentation of pancreas from abdomen CT scan', *Physics in Medicine & Biology*, Vol. 67, No. 15, p.155017.

Long, J., Song, X., An, Y., Li, T. and Zhu, J. (2022) 'Parallel multi-scale network with attention mechanism for pancreas segmentation', *IEEJ Transactions on Electrical and Electronic Engineering*, Vol. 17, No. 1, pp.110–119.

Mallouli, F. (2019) 'Robust em algorithm for iris segmentation based on mixture of Gaussian distribution', *Intelligent Automation & Soft Computing*, Vol. 25, No. 2, pp.243–248.

Milletari, F., Navab, N. and Ahmadi, S.A. (2016) 'V-net: fully convolutional neural networks for volumetric medical image segmentation', in *2016 Fourth International Conference on 3D vision (3DV)*, IEEE, October, pp.565–571.

Naidu, M.S.R., Kumar, P.R. and Chiranjeevi, K. (2018) 'Shannon and fuzzy entropy based evolutionary image thresholding for image segmentation', *Alexandria Engineering Journal*, Vol. 57, No. 3, pp.1643–1655.

Qiu, C., Liu, Z., Song, Y., Yin, J., Han, K., Zhu, Y. and Sheng, V.S. (2023) 'RTUNet: residual transformer UNet specifically for pancreas segmentation', *Biomedical Signal Processing and Control*, Vol. 79, Part 2, p.104173, ISSN: 1746-8094.

Qiu, C., Song, Y., Liu, Z., Yin, J., Han, K. and Liu, Y. (2022) 'CMFCUNet: cascaded multi-scale feature calibration UNet for pancreas segmentation', *Multimedia Systems*, Vol. 29, pp.871–886.

Qu, T., Wang, X., Fang, C., Mao, L., Li, J., Li, P. and Jin, Z. (2022) 'M3Net: a multi-scale multi-view framework for multi-phase pancreas segmentation based on cross-phase non-local attention', *Medical Image Analysis*, Vol. 75, p.102232, ISSN: 1361-8415, https://doi.org/10.1016/j.media.2021.102232.

Roth, H., Oda, M., Shimizu, N., Oda, H., Hayashi, Y., Kitasaka, T. and Mori, K. (2018) 'Towards dense volumetric pancreas segmentation in CT using 3D fully convolutional networks', in *Medical Imaging 2018: Image Processing*, SPIE, March, Vol. 10574, pp.59–64.

Roth, H.R., Lu, L., Farag, A., Shin, H.C., Liu, J., Turkbey, E.B. and Summers, R.M. (2015) 'Deeporgan: multi-level deep convolutional networks for automated pancreas segmentation', in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, October, pp.556–564.

Song, M., Hu, Y., Chen, H. and Li, T. (2017) 'Towards pervasive and user satisfactory CNN across gpu microarchitectures', in *2017 IEEE International Symposium on High Performance Computer Architecture (HPCA)*, IEEE, February, pp.1–12.

Tang, J., Wang, L., Huang, J., Shi, A. and Xu, L. (2022) 'Image semantic recognition and segmentation algorithm of colorimetric sensor array based on deep convolutional neural network', *Computational Intelligence and Neuroscience*, DOI: 10.1155/2022/2439371, PMID: 36210987; PMCID: PMC9546663.

Tian, J., Mithun, N.C., Seymour, Z., Chiu, H.P. and Kira, Z. (2020) *Recall Loss for Imbalanced Image Classification and Semantic Segmentation*.

Yan, Y. and Zhang, D. (2021) 'Multi-scale U-like network with attention mechanism for automatic pancreas segmentation', *Plos One*, Vol. 16, No. 5, p.e0252287.

Yang, Z., Peng, X. and Yin, Z. (2020) 'Deeplab_v3_plus-net for image semantic segmentation with channel compression', in *2020 IEEE 20th International Conference on Communication Technology (ICCT)* IEEE, October, pp.1320–1324.

Yao, X., Song, Y. and Liu, Z. (2020) 'Advances on pancreas segmentation: a review', *Multimedia Tools and Applications*, Vol. 79, No. 9, pp.6799–6821.

Yu, Q., Xie, L., Wang, Y., Zhou, Y., Fishman, E.K. and Yuille, A.L. (2018) 'Recurrent saliency transformation network: incorporating multi-stage visual cues for small organ segmentation', in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.8280–8289.

Zhou, Y., Xie, L., Shen, W., Wang, Y., Fishman, E.K. and Yuille, A.L. (2017) 'A fixed-point model for pancreas segmentation in abdominal CT scans', in *International conference on medical image computing and computer-assisted intervention*, Springer, Cham, September, pp.693–701.

Zhou, Z., Bian, Y., Pan, S., Meng, Q., Zhu, W., Shi, F. and Xiang, D. (2023) 'A dual branch and fine-grained enhancement network for pancreatic tumor segmentation in contrast enhanced CT images', *Biomedical Signal Processing and Control*, Vol. 82, p.104516.

Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N. and Liang, J. (2018) 'Unet++: a nested u-net architecture for medical image segmentation', in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer, Cham, pp.3–11.

Websites

https://chaos.grand-challenge.org/Data/.

https://wiki.cancerimagingarchive.net/display/Public/Pancreas-CT.